Multimodal MRI-based Stratification of Motor Phenotypes in Parkinson's Disease

Signe Hogstad

Master's Thesis in Medical Technology
Department of Physics and Technology
University of Bergen

June 1, 2023

Scientific Environment

This study was conducted at the Department of Physics and Technology at the University of Bergen in collaboration with the Mohn Medical Imaging and Visualization Centre (MMIV), the Department of Radiology at Haukeland University Hospital in Bergen, the Department of Radiology at St. Olavs University Hospital in Trondheim, and Neuro-SysMed.

Acknowledgements

I want to express my gratitude to all who have supported me throughout my work on this thesis. Firstly, I want to thank my supervisor, Frank Riemer, for providing me with this exciting project, guiding me in our meetings, and responding quickly to my messages on Slack. I am also thankful for the opportunity to present my work at the MMIV conference and the Neuro-SysMed seminar. Additionally, I thank my primary supervisor, Renate Grüner, for encouraging me throughout this project. Her solid knowledge in the field of MRI has been vital to me. Moreover, I appreciate her taking me to the 7th annual Nordic ISMRM Conference in Lund, where I held a poster presentation (travel costs covered by the Norwegian Cancer Society).

I am grateful to the internationally recognized MRI Professor Mark E. Haacke and his team at SpinTech MRI for providing the research environment at Haukeland University Hospital and for providing me with the STAGE image acquisition and image processing tool used in this thesis. In particular, I would like to thank Sagar Buch for helping me understand how STAGE works.

I want to extend my appreciation to Neuro-SysMed's Parkinson's group, especially to Charalampos Tzoulis and Simon Ulvenes Kverneng, for their feedback on my project. I am also thankful to Cecilie Brekke-Rygh, Vivian Skjeie, and Njål Brekke at the PET center at Haukeland University Hospital for providing me with MRI data and Kjersti Eline Stige from NTNU in Trondheim for providing me with clinical data. I am also grateful to MMIV and IFT for providing me with workspaces to write my thesis.

Thanks to all my friends and fellow students for the support and the good times we shared. In particular, I would like to thank my fellow MMIV master students Ingrid, Maria, Marius, and Thea, and my fellow IFT room 503 residents Julia, Linnea, and Sven. Lastly, I would like to thank my family for their constant support.

Signe Hogstad

Bergen, June 1, 2023

Abstract

Despite Parkinson's disease being described over 200 years ago, its clinical heterogeneity remains unknown. Current research aims to stratify Parkinson's disease by identifying subgroups of patients with distinct clinical characteristics. This thesis investigates the potential of multimodal magnetic resonance imaging in differentiating Parkinson's disease patients from healthy controls and distinguishing between patients with postural instability gait difficulty and tremor-dominant motor phenotype.

The study included 76 participants: 58 patients (27 with postural instability gait difficulty and 20 with tremor-dominant phenotype) and 18 healthy controls recruited from Haukeland University Hospital. Strategically acquired gradient echo imaging was used to generate various images and maps using two triple-echo acquisitions with different flip angles. Five classifiers, including XGBoost, were employed for the classification task, utilizing mean intensities of the substantia nigra pars compacta and the locus coeruleus as features.

The XGBoost classifier effectively differentiated Parkinson's disease patients from healthy controls (F1 score: 0.84, precision: 0.86, recall: 0.85). It also demonstrated high discriminatory power in distinguishing patients with postural instability gait difficulty from tremor-dominant patients (F1 score: 0.69, precision: 0.71, recall: 0.70).

The use of multimodal magnetic resonance imaging in differentiating Parkinson's disease patients from healthy controls and stratifying patients based on their motor phenotype was successfully demonstrated. The locus coeruleus in the simGRE image shows promise for distinguishing patients with Parkinson's disease from healthy controls, especially when combined with the Brief Smell Identification Test score. Potential biomarkers for discriminating patients with postural instability gait difficulty from tremor-dominant patients include the substantia nigra pars compacta and the locus coeruleus in the simDIR GM image, the substantia nigra pars compacta in the simDIR WM image, and the Montreal Cognitive Assessment score.

Sammendrag

Selv om Parkinsons sykdom ble beskrevet for over 200 år siden, er dens kliniske heterogenitet fortsatt ukjent. Forskning søker nå å stratifisere sykdommen ved å identifisere pasientgrupper med ulike kliniske egenskaper. Denne oppgaven undersøker multimodal magnetresonanstomografis potensiale til å skille mellom pasienter med Parkinsons sykdom og friske kontroller, samt skille mellom pasienter med postural instabilitet og gangvansker, og tremor-dominant motorisk fenotype.

76 deltakere, inkludert 58 pasienter med Parkinsons sykdom (27 med postural instabilitet og gangvansker, og 20 tremor-dominante) og 18 friske kontroller, ble rekruttert fra Haukeland universitetssykehus. Strategically aquired gradient echo imaging genererte bilder og maps basert på to trippel-echo sekvenser med ulike flippvinkler. Fem maskinlæringsmodeller, inkludert XGBoost, ble brukt til klassifiseringen med gjennomsnittsintensiteter av substantia nigra pars compacta og locus coeruleus som features.

XGBoost viste god evne til å skille mellom pasienter med Parkinsons sykdom og friske kontroller (F1 score: 0.84, precision: 0.86, recall: 0.85). Videre viste XGBoost høy diskriminerende kraft til å skille mellom pasienter med postural instabilitet og gangvansker, og tremor-dominante pasienter (F1 score: 0.69, precision: 0.71, recall: 0.70).

Denne oppgaven demonstrerte at multimodal magnetisk resonanstomografi kan brukes til å skille pasienter med Parkinsons sykdom fra friske kontroller, samt skille mellom pasienter med postural instabilitet og gangvansker, og tremor-dominante pasienter. Locus coeruleus i simGRE-bildet fremtrer lovende for å skille pasienter med Parkinsons sykdom fra friske kontroller, særlig i kombinasjon med Brief Smell Identification Test score. Potensielle biomarkører for å skille postural instabilitet og gangvansker fra tremor-dominant inkluderer substantia nigra pars compacta og locus coeruleus i simDIR GM-bildet, substantia nigra pars compacta i simDIR WM-bildet og Montreal Cognitive Assessment score.

Abbreviations

ACC	Accuracy
AR	Akinetic-rigid
BG	Basal ganglia
B-SIT	Brief Smell Identification Test
CNR	Contrast-to-noise ratio
CROWN	Constrained reconstruction of white noise
CSF	Cerebrospinal fluid
DA	Dopamine
DAN	Dopaminergic neuron
DAT	Dopamine transporter
DIR	Double inversion recovery
dSWI	Diamagnetic susceptibility-weighted imaging
DT	Decision tree
DTI	Diffusion tensor imaging
FA	Flip angle
FLAIR	Fluid-attenuated inversion recovery
FLIRT	FMRIB's linear image registration tool
FN	False negative
FP	False positive
FSL	FMRIB's software library
GM	Gray matter
GPe	External segment of globus pallidus

GPi	Internal segment of globus pallidus
GRE	Gradient echo
HC	Healthy control
HPF	High-pass filtered phase image
HUH	Haukeland University Hospital
IR	Inversion recovery
LB	Lewy body
LC	Locus coeruleus
LI	Laterality index
LR	Logistic regression
MAF1	Macro average F1 score
MAP	Macro average precision
MAR	Macro average recall
MDS-NMF	Movement Disorder Society Non-Motor Rating Scale
MDS-UPDRS	Movement Disorder Society Unified Parkinson's Disease Rating Scale
ME	Multi-echo
meiSWIM	Multi-echo iterative susceptibility-weighted imaging mapping
ME-GRE	Multi-echo gradient echo
MIP	Maximum intensity projection
mIP	Minimum intensity projection
MNI	Montreal Neurological Institute
MoCA	Montreal Cognitive Assessment
mpSWIM	Multi-phase susceptibility-weighted imaging mapping
MP-RAGE	Magnetization-prepared rapid gradient echo
MRA	Magnetic resonance angiography
MRI	Magnetic resonance imaging
MTC	Magnetization transfer contrast
NAN	Noradrenergic neuron

NF1	Negative F1 score
NM	Neuromelanin
NMF	Non-motor fluctuation
NPV	Negative predictive value
nTD	Non-tremor-dominant
N1	Nigrosome 1
NM-MRI	Neuromelanin-sensitive magnetic resonance imaging
PCR	Precision-recall
PD	Parkinson's disease
PDW	Proton density-weighted
PET	Positron emission tomography
PF1	Positive F1 score
PIGD	Postural instability gait difficulty
PPV	Positive predictive value
PrD	Proton density
pSWIM	Phase susceptibility-weighted imaging mapping
QSM	Quantitative susceptibility mapping
RaF	Random forest
RF	Radio frequency
ROC	Receiver operating characteristics
ROI	Region of interest
simDIR	Simulated double inversion recovery
simFLAIR	Simulated fluid-attenuated inversion recovery
simGRE	Simulated gradient echo
SN	Substantia nigra
SNR	Signal-to-noise ratio
SNpc	Substantia nigra pars compacta
SNpr	Substantia nigra pars reticulata

SPECT	Single-photon emission computed tomography
SPM	Statistical Parametric Mapping
STAGE	Strategically acquired gradient echo
STN	Subthalamic nucleus
SVC	Support vector classification
SWI	Susceptibility-weighted imaging
TD	Tremor-dominant
TE	Echo time
TI	Inversion time
TN	True negative
TNR	True negative rate
TP	True positive
TPR	True positive rate
TR	Repetition time
tSWI	True susceptibility-weighted imaging
T1W	T1-weighted
T1WE	T1-weighted enhanced
T2W	T2-weighted
T2*W	T2*-weighted
VFA	Variable flip angle
WAF1	Weighted average F1 score
WAP	Weighted average precision
WAR	Weighted average recall
WM	White matter
XGBoost	Extreme gradient boosting
3D	Three-dimensional

Symbols

α	Flip angle
α_E	Ernst angle
$\overrightarrow{B}_0$	Main magnetic field
$\overrightarrow{B}_1$	Radio frequency field
B_{1r}	Radio frequency receive field
B_{1t}	Radio frequency transmit field
$bias$	Radio frequency receive field variation
$\overrightarrow{C}$	Torque
e	Euler's number
ΔE	Energy difference
$E_{\uparrow\downarrow}$	Energy of anti-parallel state
$E_{\uparrow\uparrow}$	Energy of parallel state
γ	Gyromagnetic ratio
h	Planck's constant
k_b	Boltzmann's constant
k	Radio frequency transmit field variation
M_0	Net magnetization
M_z	Longitudinal magnetization
M_{xy}	Transverse magnetization

$\overrightarrow{\mu}$	Magnetic momentum
N_{tot}	Total number of hydrogen protons
$N_{\uparrow\downarrow}$	Number of hydrogen protons in anti-parallel state
$N_{\uparrow\uparrow}$	Number of hydrogen protons in parallel state
ω	Larmor frequency
$\overrightarrow{P}$	Angular momentum
ϕ_B	Magnetic flux
ρ_0	Proton density
ρ_{eff}	Effective proton density
ρ_{app}	Apparent proton density
S	Acquired signal
σ	Shielding constant
σ_0	Image noise standard deviation
T	Thermodynamic temperature
$T1$	Longitudinal relaxation time
$T1_{app}$	Apparent longitudinal relaxation time
$T2$	Transverse relaxation time
$T2^+$	Local inhomogeneity relaxation time
$T2^*$	Combined transverse relaxation time
θ	Angle of precession
τ_{B_1}	Duration of RF pulse
V	Induced voltage

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder and the most common neurodegenerative movement disorder in the world [1]. $2-3\%$ of the worldwide population above 65 years of age is affected by PD [2]. Patients with PD show highly heterogeneous clinical characteristics, indicating that they both can and should be classified into subgroups to provide effective treatment [3]. Currently, available treatments offer reasonable control of motor symptoms but do not halt the progression of neurodegeneration, the evolution of the disease, or the increasing disability [2]. Limited knowledge about PD's vast clinical and biological heterogeneity is a major obstacle preventing the development of patient-tailored therapies [4].

Current research aims to stratify PD by identifying and characterizing subgroups of patients with distinct clinical and molecular characteristics. Genetic, pathological, and imaging markers, as well as motor and non-motor symptoms, might define subtypes of PD [3]. Classifying PD patients into subgroups based on motor symptoms can be an essential step toward personalized treatment. The Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is a widely used assessment tool for motor and non-motor PD symptoms. It can also be used to determine the motor phenotype of PD patients [5].

The motor phenotypes of PD include tremor-dominant (TD) and non-tremor-dominant (nTD), which further splits into postural instability gait difficulty (PIGD) and akinetic-rigid (AR) subtypes [6]. The tremor and PIGD subscales in part III of the MDS-UPDRS consist of eight and five items, respectively. TD patients are identified by a ratio of mean tremor score to mean PIGD score greater than or equal to 1.5, while PIGD patients have a ratio of less than or equal to 1.0 [7]. Patients with ratios between 1.0 and 1.5 are classified as indeterminate.

On a pathological level, PD is characterized by progressive degeneration of subcortical dopaminergic nigrostriatal systems, Lewy body (LB) aggregations, and dopamine (DA) depletion in the striatum [8]. Postmortem histology has identified a loss of dopaminergic neurons (DANs) in the substantia nigra pars compacta (SNpc) and noradrenergic neurons (NANs) in the locus coeruleus (LC) in PD patients [9][10]. The degeneration of DANs in the SNpc in PD has been extensively studied and is better comprehended than the degeneration of NANs in the LC. Nonetheless, the loss of LC neurons occurs before the loss of SNpc neurons in the PD brain and may serve as some of the earliest evidence for the onset of PD [11]. Consequently, there is a growing interest in investigating the role of the LC in PD.

In addition to the loss and degeneration of DANs in the SNpc and NANs in the LC, PD has been strongly linked to the increased iron content in the basal ganglia (BG) [12]. Although it remains uncertain whether the rise in iron content is a primary factor or a consequence of the disease, brain structures naturally rich in iron, such as the SNpc, appear more susceptible to iron accumulation in neurodegenerative disorders like PD [13]. Iron is necessary for DA production, but excessive iron buildup in the SNpc may contribute to the degeneration of DANs [14]. Thus, high iron content in the SNpc may serve as a potential biomarker for PD.

While the total concentration of iron in the SNpc increases with age in healthy individuals, the total iron concentration in the LC is lower and remains stable throughout life [15][16][17]. Consequently, iron may play a more significant role in neurodegeneration in the SNpc compared to the LC. However, further research is necessary to fully understand the role of iron in the pathogenesis of PD.

The SNpc and the LC are the two most pigmented regions in the human nervous system [18]. They contain neuromelanin (NM), a dark-colored, complex, insoluble pigment. In PD, NM-containing neurons in the SNpc and the LC consistently show degeneration, which exceeds the degree expected from normal aging [19]. Several histopathological studies of these regions have shown how neuronal cell loss correlates well with the estimated NM loss, especially in the SNpc [20][21]. Thus, low NM content in the SNpc and the LC may serve as potential biomarkers for PD.

Single-photon emission computed tomography (SPECT) and positron emission tomography (PET) are considered gold-standard imaging techniques for evaluating the function of the brain's dopaminergic system [22]. SPECT and PET can aid in differentiating PD from other disorders with similar symptoms and monitor disease progression [22].

However, SPECT and PET findings cannot fully explain motor impairments in PD, as complex comorbid deficits and the degeneration of other neuronal systems co-occur [23]. Additionally, SPECT and PET have limitations such as being expensive, involving exposure to ionizing radiation and having a limited spatial resolution, which can make it challenging to localize changes in the brain precisely. The last limitation is significant in prodromal PD, where subtle changes in the dopaminergic system occurs [24].

Magnetic resonance imaging (MRI) is a non-ionizing imaging technique that utilizes a strong magnetic field and radio waves to generate detailed brain images. MRI has several advantages compared to SPECT and PET, including higher spatial resolution, lower cost, no ionizing radiation, and better availability [25]. Moreover, MRI can provide various types of imaging data, such as structural images, diffusion-weighted images, and functional images, enabling a more comprehensive evaluation of changes in the brain related to PD. Therefore, MRI can be particularly crucial when identifying and characterizing PD patient subgroups.

MRI can indirectly detect iron accumulation in the brain by measuring magnetic susceptibility in tissue [26]. Magnetic susceptibility is a property of tissue that reflects its ability to become magnetized in a magnetic field. Due to its large magnetic moment, iron influence the MRI signal in susceptibility-weighted imaging (SWI). Iron and other paramagnetic substances increase the magnetic field, creating a positive phase shift compared to the surrounding tissues. There is a direct correlation between brain iron concentration and phase shift [12]. Other MRI techniques developed to measure magnetic susceptibility in the brain includes T2*-weighted (T2*W) imaging and quantitative susceptibility mapping (QSM).

Alterations of NM and its role in PD pathophysiology can be assessed in vivo through NM-sensitive MRI (NM-MRI). NM has a paramagnetic T1 reduction effect on MRI due to the presence of melanin-iron complexes [27]. Loss of SNpc hyperintensity in an NM-MRI image is associated with the loss of NM-containing neurons in PD [28]. The utilization of NM-MRI for quantifying the loss of NANs in the LC is less prevalent than for the SNpc. Nevertheless, it is reasonable to assume that the loss of NANs in the LC in PD can also be quantified using NM-MRI.

Utilizing MRI to differentiate motor phenotypes in PD can enhance the ability to track disease progression, identify at-risk populations, and optimize the targeting of non-invasive and invasive neuromodulation therapies [29][30].

A study using SWI showed lower values in the globus pallidus for the PIGD subtype compared to non-PIGD subtypes, with a similar trend in other BG nuclei such as the SNpc [12]. This indicates that the PIGD motor phenotype exhibits a higher iron content in the SNpc compared to non-PIGD motor phenotypes, such as TD [31]. Additionally, NM-MRI revealed more severe signal attenuation in the medial part of the SNpc for the PIGD subtype compared to TD, with the SNpc ipsilateral to the most clinically affected side being the most robust discriminator between the two subtypes [32]. This indicates that the PIGD motor phenotype demonstrates a lower NM content in the SNpc compared to the TD motor phenotype [28].

In a study conducted in 2019, Depierreux et al. discovered a significant reduction in proton density (PrD) in the left and right SNpc of PD patients compared to healthy controls (HCs) [33]. This signal intensity decline can potentially be attributed to the loss of DANs in the SNpc, leading to decreased neuronal density and SNpc brightness in PrD maps. Therefore, low PrD in the SNpc and possibly in the LC could serve as a potential biomarker for PD.

The objective of this Master of Science project is to investigate the possibility of utilizing multimodal MRI to differentiate PD patients from HCs and stratifying PD patients based on their motor phenotypes, specifically PIGD and TD. Various MRI techniques will be employed to quantify the loss of DANs in the SNpc and the loss of NANs in the LC. Through exploring the correlations between the quantified neuronal losses and clinical data, this thesis aims to enable the classification of participants into either PD patients or HCs and further stratify patients based on their motor phenotype.

Building upon the research presented in this chapter, the hypothesis regarding the classification of PD patients and HCs is as follows: PD patients demonstrate more pronounced degeneration of DANs in the SNpc and NANs in the LC than HCs. These neuronal losses can be observed through increased iron content, decreased NM content, and decreased PrD in both the SNpc and the LC.

Regarding the PIGD/TD classification, the hypothesis is that PIGD patients demonstrate more pronounced degeneration of DANs in the SNpc and NANs in the LC compared to TD patients. As for the PD/HC classification, these neuronal losses can be observed through increased iron content, decreased NM content, and decreased PrD in both the SNpc and the LC.

Chapter 2

Theory

2.1 MRI

MRI is a non-invasive and non-ionizing modality that provides a spatial map of hydrogen nuclei and their surroundings in different tissues [34]. The intensity of an MRI image depends on the number of protons at each location in space, as well as tissue properties such as viscosity and iron content. Because disease processes contribute to changes in water content and the local chemical environment, MRI is particularly suited to study disease-related changes in soft tissues such as the brain.

2.1.1 Principles of MRI

Hydrogen Protons in a Magnetic Field

The hydrogen proton is a charged particle that spins around its rotational axis with angular momentum $\overrightarrow{P}$ [34]. Hence, it has a magnetic moment $\overrightarrow{\mu}$, which is proportional to the angular momentum:

$$|\overrightarrow{\mu}| = \gamma|\overrightarrow{P}| \tag{2.1}$$

Here, γ is the gyromagnetic ratio. The magnitude of the magnetic moment is fixed for a given nucleus since the angular momentum of the proton is quantized and fixed. However, the direction of the magnetic moment is random. Therefore, the net magnetization is zero in the absence of an external magnetic field.

The fundamental components of an MRI scanner are a superconducting magnet, three magnetic field gradient coils, and an radio frequency (RF) transmitter and receiver. The magnet produces a static magnetic field $\overrightarrow{B}_0$ in the horizontal direction ($+z$). When placed in $\overrightarrow{B}_0$, the proton spin will precess at a fixed angle about the direction of $\overrightarrow{B}_0$.

Precession is the circular motion of a spinning particle's rotation axis about another fixed axis caused by the application of torque in the direction of the precession [35]. The torque, $\overrightarrow{C}$, which is caused by the $\overrightarrow{B}_0$ field trying to align the proton magnetic moment, is given by:

$$\overrightarrow{C} = \frac{d\overrightarrow{P}}{dt} = \overrightarrow{\mu} \times \overrightarrow{B_0} = |\overrightarrow{\mu}||\overrightarrow{B_0}| \sin\theta \tag{2.2}$$

Here, θ is the angle of precession. The direction of the torque is tangential to the direction of $\overrightarrow{\mu}$, which causes the precession [34]. In a short time dt, $\overrightarrow{\mu}$ precess through an angle $d\phi$ resulting in a change $d\overrightarrow{P}$ in the angular momentum. Trigonometry gives the relationship that:

$$\sin d\phi = \frac{d\overrightarrow{P}}{|\overrightarrow{P}|\sin\theta} = \frac{\overrightarrow{C}\,dt}{|\overrightarrow{P}|\sin\theta} \tag{2.3}$$

The approximation that $\sin d\phi = d\phi$ holds for small $d\phi$ [34]. Combination of Equation 2.2 and 2.3 gives the angular precession frequency, also called the Larmor frequency:

$$\omega_0 = \frac{d\phi}{dt} = \frac{\overrightarrow{C}}{|\overrightarrow{P}|\sin\theta} = \frac{\overrightarrow{\mu} \times \overrightarrow{B_0}}{|\overrightarrow{P}|\sin\theta} = \frac{\gamma\overrightarrow{P} \times \overrightarrow{B_0}}{|\overrightarrow{P}|\sin\theta} = \frac{\gamma|\overrightarrow{P}||\overrightarrow{B_0}|\sin\theta}{|\overrightarrow{P}|\sin\theta} = \gamma B_0 \tag{2.4}$$

Parallel configuration is when the z-component of the magnetic moment is aligned with $\overrightarrow{B_0}$. Anti-parallel configuration is when the z-component of the magnetic moment is aligned in the opposite direction of $\overrightarrow{B_0}$. The parallel state requires less energy than the anti-parallel state. The energy differences between the two states are:

$$\Delta E = E_{\uparrow\downarrow} - E_{\uparrow\uparrow} = \frac{\gamma h B_0}{4\pi} - \left(-\frac{\gamma h B_0}{4\pi}\right) = \frac{\gamma h B_0}{2\pi} \tag{2.5}$$

Here, $E_{\uparrow\uparrow}$ is the energy of the parallel state, $E_{\uparrow\downarrow}$ is the energy of the anti-parallel state, and h is Planck's constant. The two energy states are populated following Boltzmann statistics:

$$\frac{N_{\uparrow\downarrow}}{N_{\uparrow\uparrow}} = e^{-\frac{\Delta E}{k_b T}} = e^{-\frac{\gamma h B_0}{2\pi k_b T}} \tag{2.6}$$

Here, $N_{\uparrow\uparrow}$ is the number of hydrogen protons in the parallel state, $N_{\uparrow\downarrow}$ is the number of hydrogen protons in the anti-parallel state, k_b is Boltzmann's constant, and T is thermodynamic temperature. Assuming $\gamma h B_0 << 2\pi k_b T$:

$$\frac{N_{\uparrow\downarrow}}{N_{\uparrow\uparrow}} = 1 - \frac{\gamma h B_0}{2\pi k_b T} \implies N_{\uparrow\downarrow} = N_{\uparrow\uparrow} - N_{\uparrow\uparrow}\frac{\gamma h B_0}{2\pi k_b T} \tag{2.7}$$

Assuming $N_{tot} = 2N_{\uparrow\uparrow}$:

$$N_{\uparrow\uparrow} - N_{\uparrow\downarrow} = N_{tot}\frac{\gamma h B_0}{4\pi k_b T} \tag{2.8}$$

Thus, due to a difference in energy, the number of protons in the parallel state is slightly larger than the number of protons in the anti-parallel state, leading to a net magnetization in the positive z-direction [34]:

$$M_0 = M_z = \sum_{n=1}^{N_{tot}} \mu_{z,n} = \sum_{n=1}^{N_{\uparrow\uparrow}} \frac{\gamma h}{4\pi} - \sum_{n=1}^{N_{\uparrow\downarrow}} \frac{\gamma h}{4\pi} = \frac{\gamma h}{4\pi}(N_{\uparrow\uparrow} - N_{\uparrow\downarrow}) \tag{2.9}$$

The net magnetization has only a z-component since the vector sum of the components on the x- and the y-axis is zero.

Inserting Equation 2.8 into Equation 2.9:

$$M_0 = |\overrightarrow{M}| = \frac{\gamma^2 h^2 B_0 N_{tot}}{16\pi^2 k_b T} \tag{2.10}$$

Hence, the MRI signal is proportional to the field strength B_0 and inversely proportional to the temperature T.

Signal Generation

Resonance, the basis for signal generation in MRI, happens when a nucleus is exposed to an oscillating force with a frequency close to the Larmor frequency [34]. Energy is then transferred to the nucleus. Inside the MRI magnet, when a RF pulse with a frequency equal to the Larmor frequency for hydrogen is applied, resonance occurs and the hydrogen nuclei are excited. This leads to an increase in the number of protons in the anti-parallel state. The RF pulse is a short-lived electromagnetic field, $\overrightarrow{B_1}$, perpendicular to the $\overrightarrow{B_0}$ field.

The $\overrightarrow{B_1}$ field produces a torque that causes the net magnetization to rotate towards the *xy*-plane. For a rectangular RF pulse, the flip angle (FA), α, which is the angle through which the net magnetization is rotated, is given as:

$$\alpha = \gamma B_1 \tau_{B_1} \tag{2.11}$$

Here, B_1 is the RF field strength in Tesla and τ_{B_1} is the time at which the RF field is applied. As long as the RF pulse is on, all the moving magnetic moments are phase coherent. The net magnetic moment will then precess at Larmor frequency in a plane at an angle determined by the FA.

Relaxation

The system will return to equilibrium immediately after the pulse is switched off [34]. This process, called *relaxation*, includes dephasing the net magnetization to realign with $\overrightarrow{B_0}$ and consists of two independent processes. T1-relaxation, or spin-lattice relaxation, is when the *z*-component return to equilibrium due to individual magnetic moments releasing energy to the surrounding tissue.

The longitudinal magnetization, M_z, at time t is given by:

$$M_z(t) = M_0 \cos\alpha + (M_0 - M_0 \cos\alpha)\left(1 - e^{-\frac{t}{T1}}\right) \tag{2.12}$$

The relaxation time $T1$ is defined as the time when $t = T1$ in Equation 2.12:

$$M_z(T1) = M_0 \cos\alpha + (M_0 - M_0 \cos\alpha)\left(1 - e^{-1}\right) \approx 0.63M_0 + 0.37M_0 \cos\alpha \tag{2.13}$$

Equation 2.13 states that when the FA is 90°, $T1$ is how long it takes for 63% of the original net magnetization to be restored.

T2-relaxation, or spin-spin relaxation, is when the *x*- and *y*-components return to equilibrium due to a loss of phase coherence, also called dephasing, which happens because of molecular dynamics resulting in a small spread in precessional frequencies [34]. The transverse magnetization, M_{xy}, at time t is given by:

$$M_{xy}(t) = M_0 \sin\alpha \, e^{-\frac{t}{T2}} \tag{2.14}$$

The relaxation time $T2$ is defined as the time when $t = T2$ in Equation 2.14:

$$M_{xy}(T2) = M_0 \sin\alpha\, e^{-1} \approx 0.37 M_0 \sin\alpha \tag{2.15}$$

Equation 2.15 states that when the FA is 90°, $T2$ is how long it takes before only 37% of the transverse magnetization is left.

Spatial inhomogeneities in the $\overrightarrow{B_0}$ within the patient is another factor that contributes to T2-relaxation [34]. Local variations in the $\overrightarrow{B_0}$ field lead to precession frequency variations, which causes faster dephasing than in normal T2-relaxation. One source for this variation is that designing a magnet with a perfectly uniform field over the imaging volume is impossible. Another source is the different magnetic susceptibilities in the different parts of the body, including surgical implants, interfaces between tissues with very different properties, and depositions of magnetic substances such as iron.

The effects of local inhomogeneity in the $\overrightarrow{B_0}$ field are represented by a relaxation time $T2^{+}$. The combined relaxation time $T2^{*}$ is then given by:

$$\frac{1}{T2^{*}} = \frac{1}{T2^{+}} + \frac{1}{T2} \tag{2.16}$$

Both $T1$ and $T2$ depend on the type of tissue and field strength because the strength of the magnetic field experienced by a hydrogen proton is affected by the electron configuration in its surroundings, leading to different resonance frequencies for different tissues [34]. Taking this into account, the resonance frequency of the hydrogen proton is given by:

$$\omega_0 = \gamma B_0 (1 - \sigma) \tag{2.17}$$

Here, σ is a shielding constant related to the electronic environment surrounding the nucleus. Since water and lipids have different electron configurations, the protons in the two mediums resonate at different frequencies. The resonant frequency in water is higher than in lipids because oxygen is more electronegative than carbon, making protons in water less shielded than protons in lipids.

Signal Detection

The MRI signal is commonly detected in quadrature using receiver coils sensitive to magnetic flux in two orthogonal directions [34].

A voltage will be induced in the coils according to Faraday's law of induction when the net magnetic moment precesses:

$$V \propto -\frac{d\phi_B}{dt} \tag{2.18}$$

Here, V is the induced voltage in each loop, and ϕ_B is the magnetic flux. This voltage represents the recorded MRI signal in time.

2.1.2 Imaging Sequences

MRI pulse sequences have three components: Preparatory module, acquisition, and recovery [36]. The preparatory module includes all pulses applied prior to phase encoding. An example is inversion recovery (IR), where an inversion pulse is applied. The acquisition portion can be any pulses that generate a signal, such as spin echo or GRE. Recovery is the relative dead time when no signals are generated, and the system can return to equilibrium. After recovery, the cycle is repeated.

Inversion time (TI), echo time (TE), and repetition time (TR) are basic pulse sequence parameters. TI is the time between the middle of the inversion pulse β, and the middle of the subsequent excitation pulse α [37]. TE is the time between the middle of the excitation pulse α and the middle of the echo. Several echo times may be defined for sequences with multiple echoes between each pulse. TR is the time between a pulse sequence's beginning and the succeeding pulse sequence. This thesis uses and describes the following imaging sequences: Magnetization-prepared rapid GRE (MP-RAGE) and multi-echo GRE (ME-GRE).

The 3D MP-RAGE Sequence

The three-dimensional (3D) MP-RAGE sequence can acquire T1-weighted (T1W) and proton density-weighted (PDW) images with high resolution and high contrast in a relatively short amount of time [38]. A typical 3D MP-RAGE pulse sequence consists of a cycle with three steps [39]: 1) Magnetization preparation with a 180° inversion pulse β and a medium TI of 600-900 ms for contrast control. 2) Data acquisition with a spoiled GRE sequence with a short TE of 2-4 ms and a small FA α of 5-12°. 3) Magnetization recovery with a long TR of around 2000 ms for additional contrast control. A spoiled sequence is a sequence where the transverse coherences are purposely disrupted [40]. Figure 2.1 illustrates a pulse sequence diagram for the 3D MP-RAGE sequence. The image contrast is determined by the effective TI. The resulting MP-RAGE image is primarily T1W, but density and T2* effects are also present.

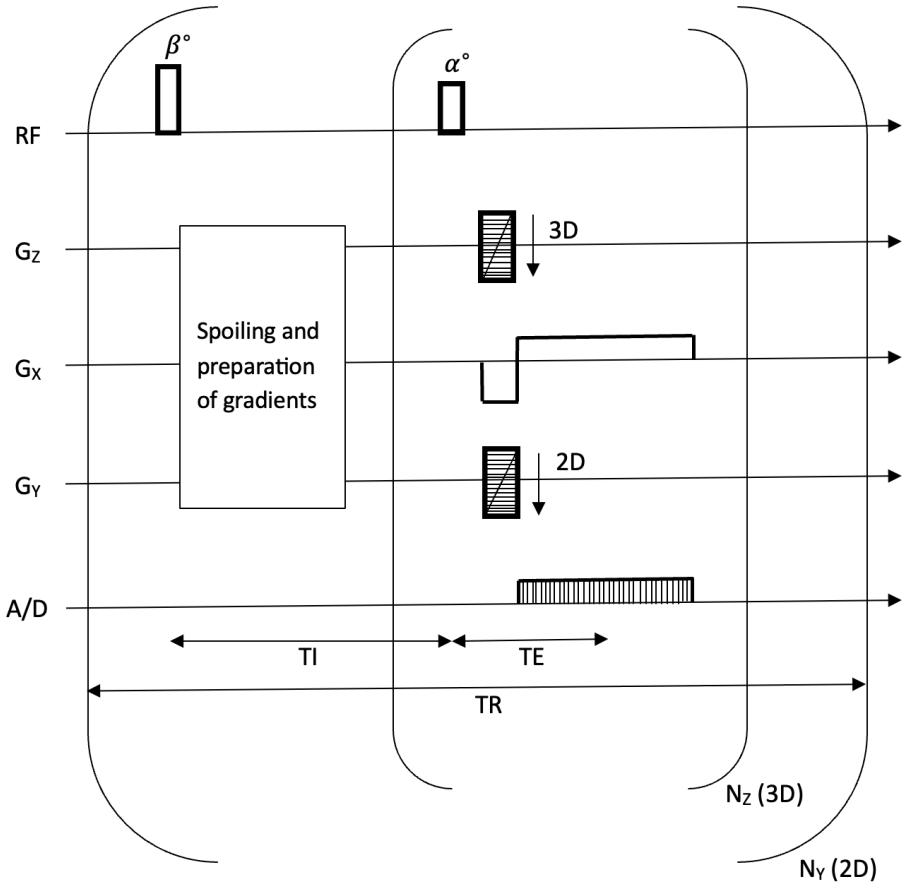

Figure 2.1: Pulse sequence diagram for the 3D MP-RAGE sequence. The magnetization period comprises a non-selective RF pulse β followed by a variable delay TI. During this delay, gradients are applied for spoiling. Data acquisition is performed with a GRE sequence with a excitation pulse α. There are N_Y phase encoding steps and N_Z slice encoding steps. All N_Z slice encoding steps are collected following the inversion pulse, then repeated for all N_Y phase encoding steps. RF = radio frequency, G_Z = slice encoding gradient, G_X = frequency encoding gradient, G_Y = phase encoding gradient, A/D = analog to digital converter, TI = inversion time, TE = echo time, TR = repetition time.

The 3D ME-GRE Sequence

Single GREs can be repeated to generate one or more additional GREs following a single RF pulse [41]. This technique is known as ME-GRE. Subsequent echoes progressively decrease in magnitude due to T2* relaxation. Figure 2.2 illustrates a pulse sequence diagram for the 3D ME-GRE sequence with three echoes and flow-compensation. Flow-compensation, which is employed to suppress flow artifacts and enhance image quality, involves the inclusion of additional gradient lobes before signal readout to compensate in advance for motion-induced dephasing at the time of the echo [42].

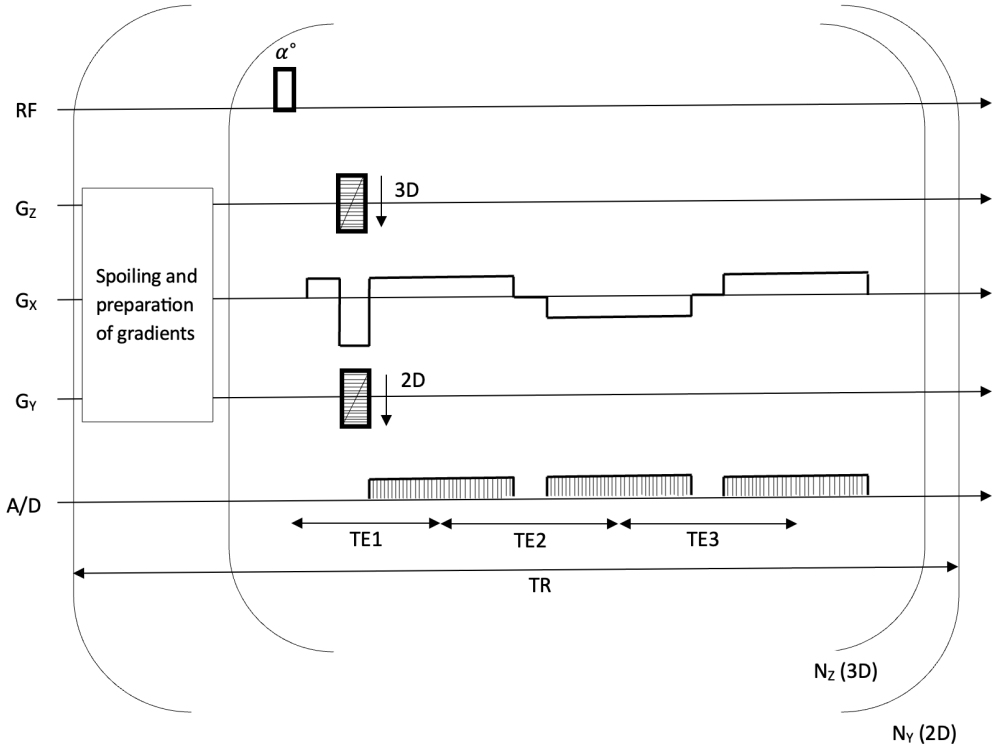

Figure 2.2: Pulse sequence diagram for the 3D ME-GRE sequence with three echoes. Spoiling and flow-compensation are performed prior to data acquisition. α is the excitation pulse. There are N_Y phase encoding steps and N_Z slice encoding steps. RF = radio frequency, G_Z = slice encoding gradient, G_X = frequency encoding gradient, G_Y = phase encoding gradient, A/D = analog to digital converter, TE = echo time, TR = repetition time.

Signal Intensity and FAs

The choice of FA is critical for determining the signal intensity and image contrast [34]. For an ideal steady-state RF spoiled GRE data acquisition, the acquired signal, S, as a function of FA, α, is given by the Ernst Equation:

$$S(\alpha) = \rho_0 \sin\alpha \frac{1-E_1}{1-\cos\alpha E_1} E_2 \tag{2.19}$$

Here, ρ_0 is the PrD, E_1 is $e^{-TR/T1}$, TR is the repetition time, T1 is the longitudinal relaxation time, E_2 is $e^{-TE/T2^*}$, TE is the echo time, and $T2^*$ is the transverse relaxation time [43]. Figure 2.3 show the acquired signal as a function of FA for gray matter (GM) and white matter (WM).

The FA maximizing signal from a tissue in a spoiled GRE sequence is called the Ernst angle and is given by:

$$\alpha_E = \arccos e^{-TR/T1} \tag{2.20}$$

An FA smaller than the Ernst angle of WM gives a PDW image, whereas an FA larger than the Ernst angle gives a T1W image [40].

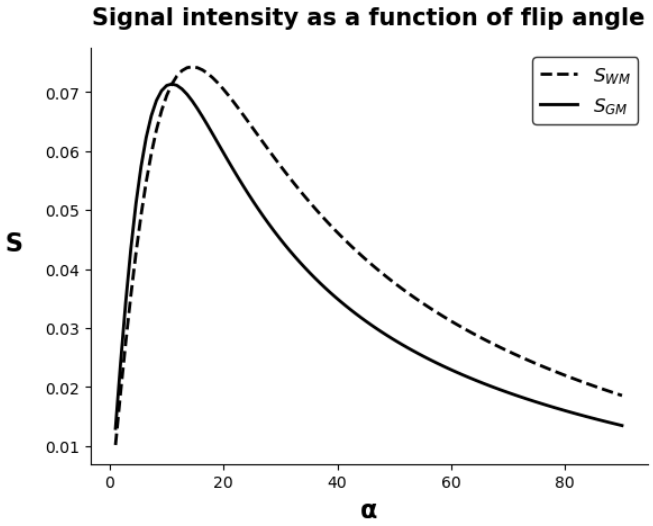

Figure 2.3: Signal intensity, S, as a function of FA, α. Generated using Equation 2.19 with TR = 29 ms, TE = 7.5 ms, $\rho_{0,GM}$ = 0.84, $\rho_{0,WM}$ = 0.68, $T1_{GM}$ = 1600 ms, $T1_{WM}$ = 900 ms, $T2^*_{GM}$ = 66 ms and $T2^*_{WM}$ = 50 ms. WM = white matter, GM = gray matter, TR = repetition time, TE = echo time, ρ_0 = proton density, $T1$ = longitudinal relaxation time, $T2^*$ = combined transverse relaxation time.

The WM/GM contrast-to-noise ratio (CNR) for a signal with a single FA, α, is given by:

$$CNR_\alpha = \frac{|S_{WM}(\alpha) - S_{GM}(\alpha)|}{\sigma_0} \tag{2.21}$$

Here, S_{WM} and S_{GM} are mean signal values for respectively WM and GM in a particular region, and σ_0 is image noise standard deviations for both WM and GM [43]. The T1 ratio of WM and GM restricts the contrast between WM and GM. However, the contrast can be improved by combining images with different FAs. One approach is to subtract a PDW image from a T1W image. The WM/GM CNR for a signal with different FAs, (α_1, α_2), is given by:

$$CNR_{(\alpha_1,\alpha_2)} = \frac{|(S_{WM}(\alpha_2) - S_{WM}(\alpha_1)) - (S_{GM}(\alpha_2) - S_{GM}(\alpha_1))|}{\sqrt{2}\sigma_0} \tag{2.22}$$

Including the effects of RF inhomogeneity, the acquired GRE signal is given as:

$$S'(\alpha) = \rho_0 \,\text{bias}\, \sin(k\alpha) \frac{1 - E_1}{1 - \cos(k\alpha) E_1} E_2 \tag{2.23}$$

Here, k is the B1 transmit field (B_{1t}) variation and bias is the B1 receive field (B_{1r}) variation [43]. Equation 2.23 states that a FA smaller than the Ernst angle of WM suffers more B_{1t} field variation than an FA large than the Ernst angle and, therefore, should be corrected. The T1W enhanced (T1WE) image is defined by the linear subtraction between the signal with the larger FA, α_2, and the signal with the smaller FA, α_1, as:

$$S_{T1WE} = S'(\alpha_2, TE_n) - \lambda\, S'(\alpha_1, TE_n) \tag{2.24}$$

Here, $\lambda = 1/k^a$, k is the extracted B_{1t} field variation, a is a constant and TE_n is the n-th echo time. For ME scans, the final T1WE image is given by the average of all T1WE images calculated from Equation 2.24 using all echoes.

Quantitative MRI

Everyday clinical practice increasingly relies on quantitative MRI, which includes methods such as quantitative PrD, T1, T2*, and QSM [44]. These techniques can aid in segmenting and classifying normal and abnormal tissue types by quantifying tissue properties. The variable FA (VFA) method has gained popularity due to its accessible data collection for estimating T1 and PrD. The VFA method is based on acquiring a few spoiled GREs with different FAs to quantify T1 and PrD.

Equation 2.23 can be rewritten as:

$$\frac{S}{\sin(k\,\alpha)} = E_1 \frac{S}{\tan(k\,\alpha)} + \rho_{eff}(1 - E_1) \tag{2.25}$$

Here, $\rho_{eff} = \rho_0 \,\mathrm{bias}\, E_2$ is the effective PrD [44]. To obtain T1 and ρ_{eff}, data for different FAs can be collected, transformed, and fitted to a line with slope E_1 and intercept $\rho_{eff}(1 - E_1)$. Two measurements with suitably optimized FAs are sufficient for reliable PrD and T1 mapping when one angle is below the Ernst angle in Equation 2.20 and the other is above it. However, due to RF inhomogeneity, the measured FA may not accurately reflect the actual value used to conduct the scan, resulting in significant errors when quantifying PrD and T1, particularly at high magnetic fields [44].

For small FAs where TR«T1, Equation 2.23 can be approximated as:

$$S = \rho_{eff} \frac{k\alpha}{1 + \frac{(k\alpha)^2}{\alpha_E^2}} \tag{2.26}$$

Estimations of the apparent PrD and T1 can be derived from Equation 2.26:

$$\rho_{app} = k\,\rho_{eff} = \rho_0\,k\,\text{bias}\,E_2 \tag{2.27}$$

$$T1_{app} = T1\,k^2 \tag{2.28}$$

Hence, if the actual FA is scaled by k, the apparent PrD is scaled by the product of bias and k, and the apparent T1 is scaled by k^2. Therefore, correcting the FAs spatially before fitting the data using Equation 2.25 is crucial.

The iron-sensitive MRI images and maps generated in this thesis are described in Table 2.1, while the non-iron-sensitive MRI images and maps are described in Table 2.2. The descriptions were obtained through direct communication with Sagar Buch from SpinTech MRI via email [45]. Some of the images are currently in an experimental phase and do not have any related publications.

Constrained reconstruction of white noise (CROWN) is a self-trained, algorithmic approach to improve image quality by increasing the signal-to-noise ration (SNR) of WM. The WM is chosen as a reference, and by using the WM mask, the image is smoothed to improve the SNR and increase the contrast between the WM and other tissues [45]. Maximum intensity projection (MIP) and minimum intensity projection (mIP) are projection images that are obtained from 3D datasets by selecting, respectively, the maximum and minimum intensity along lines that cut through the 3D volume [37].

Table 2.1: Description of iron-sensitive MRI images and maps generated in this thesis. The descriptions were obtained through direct communication with Sagar Buch from SpinTech MRI via Email [45]. An upward-pointing arrow indicates that the signal increases with an increase in iron content, while a downward-pointing arrow indicates that the signal decreases with an increase in iron content. FA = flip angle, CROWN = constrained reconstruction of white noise, TE = echo time, QSM = quantitative susceptibility mapping.

Image/map	Description	Signal
T2*	Combined transverse relaxation time. Generated using a mono-exponential approach for two sequences with different FAs.	↓ [46]
T2* MIP	Maximum intensity projection of the T2*.	↓ [46]
CROWN T2*	Same as T2*, but with CROWN to improve image quality.	↓ [46]
CROWN T2* A2	Another version of the CROWN T2*. Currently under investigation as an experiment.	↓ [46]
R2*	Inverted version of T2*, as R2* = 1/T2*.	↑ [46]
R2* MIP	Maximum intensity projection of the R2*.	↑ [46]
CROWN R2*	Same as R2*, but with CROWN to improve image quality.	↑ [46]
CROWN R2* A2	Another version of the CROWN R2*. Currently under investigation as an experiment.	↑ [46]
HPF	High-pass filtered phase. Used to enhance local tissues by removing unwanted global fields.	↓ [47]
pSWIM	Phase susceptibility-weighted imaging mapping. The maximum intensity projection of the HPF image.	↓ [47]
mpSWIM	Multi-phase susceptibility-weighted imaging mapping. Similar to pSWIM. Currently under investigation as an experiment.	↓ [47]
SWI	Paramagnetic susceptibility-weighted imaging. Generated by combining a long TE magnitude image and a high-pass filtered long TE phase image.	↓ [31]
SWI mIP	Minimum intensity projection of the SWI.	↓ [31]
meiSWIM	Multi-echo iterative susceptibility weighted imaging mapping. An iterative, geometry-constrained truncated k-space approach to generate QSM data. Obtained from multiple echos, as opposed to standard QSM, which improves the quality. Certain brain areas may be removed due to high phase gradients, creating 'holes'.	↑ [48]
meiSWIM filled	Same as meiSWIM, but with estimated information filled in for any 'holes' in the original meiSWIM data.	↑ [48]
meiSWIM filled MIP	Maximum intensity projection of the meiSWIM filled.	↑ [48]
meiSWIM HPF	High-pass filtered meiSWIM. Certain brain areas may be removed due to high phase gradients, creating 'holes'.	↑ [48]
meiSWIM HPF filled	Same as meiSWIM HPF, but with estimated information filled in for any 'holes' in the original meiSWIM HPF data.	↑ [48]
tSWI	True susceptibility-weighted imaging. Generated by combining the original magnitude data with the QSM data instead of combining the original magnitude data with the HPF as for SWI.	↓ [49]
tSWI mIP	Minimum intensity projection of the tSWI.	↓ [49]
tSWI HPF	High-pass filtered tSWI.	↓ [49]

Table 2.2: Description of non-iron-sensitive MRI images and maps generated in this thesis. The descriptions were obtained through direct communication with Sagar Buch from SpinTech MRI via Email [45]. An upward-pointing arrow indicates that the signal increases with degeneration or an increase in NM content, while a downward-pointing arrow indicates that the signal decreases with degeneration or an increase in NM content. DG = degeneration, NM = neuromelanin, VFA = variable flip angle, WM = white matter, CSF = cerebrospinal fluid, FA = flip angle, $T2^*$ = transverse relaxation time, GRE = gradient echo, SNR = signal-to-noise ratio, GM = gray matter, CSF = cerebrospinal fluid, SWI = susceptibility-weighted imaging, TE = echo time, B_1 = radiofrequency field.

Image/map	Description	Signal
PrD	Proton density. Generated using the VFA method.	↓ (DG) [33]
CROWN PrD	Same as PrD, but with CROWN to improve image quality.	↓ (DG) [33]
True PrD	True proton density. The WM and CSF PrD values are used as a reference to obtain values closer to the absolute PrD values across the tissues.	↓ (DG) [33]
CROWN true PrD	Same as true PrD, but with CROWN to improve image quality.	↓ (DG) [33]
CROWN true PrD A	Another version of the CROWN PrD. Currently under investigation as an experiment.	↓ (DG) [33]
T1	Longitudinal relaxation time. Generated using the VFA method.	↑ (NM) [28]
T1WE	T1-weighted enhanced image. Post-processing of large FA data is used to improve the T1 contrast between tissues.	↑ (NM) [28]
simGRE	Simulated gradient echo. Magnitude image for the first echo. Utilize the derived PrD, T1, T2*, and the known imaging parameters to simulate the acquired data using the 3D GRE signal equation. The simulated data have improved SNR.	
simGREa	Same as simGRE, but for the second echo.	
simFLAIR	Simulated fluid-attenuated inversion recovery image. Generated by utilizing the derived PrD, T1, T2*, and the known imaging parameters through the FLAIR signal equation.	
simDIR GM	Simulated double inversion recovery gray matter. Allows for GM segmentation.	
simDIR WM	Simulated double inversion recovery white matter. Allows for WM segmentation.	
simDIR CSF	Simulated double inversion recovery cerebrospinal fluid. Allows for CSF segmentation.	
dSWI	Diamagnetic susceptibility-weighted image. Show the opposite contrast to the SWI and enhance the edges of paramagnetic structures.	
dSWI mIP	Minimum intensity projection of the dSWI.	
MRA	Magnetic resonance angiogram. A map of the major arteries. Generated from a short TE large FA image.	
k	Map of B_1 transmit field variations.	

2.2 Parkinson's Disease

2.2.1 History

Dr. James Parkinson published his monograph 'An Essay on the Shaking Palsy' in 1817 [50]. This monograph is considered the first description of PD, named after James Parkinson himself. Parkinson described the disease as insidious with gradually developing disabilities. He identified three main symptoms: Resting tremor, flexed posture, and gait festination.

Tremor is a rhythmical, involuntary oscillatory movement of a body part [51]. Resting tremor occurs when the muscle is relaxed. A resting tremor often seen in PD patients is called 'pill-rolling' because of circular finger and hand movements which resemble the rolling of small objects or pills in hand [52]. A flexed posture is characterized by thoracic kyphosis, protrusion of the head, and, in severe cases, knee flexion [53]. Thoracic kyphosis refers to the excessive forward curvature of the spine in the upper back. Gait festination is a locomotion disturbance that involves progressive shortening of step length, accompanied by a compensatory increase in cadence [54].

A significant milestone in the history of PD was identifying LBs, intracytoplasmic inclusions, a pathological hallmark made by Frederick Lewy in 1912 [55]. LBs are abnormal clumps of specific substances [56]. Located within LBs is the misfolded protein α-synuclein, which affects chemicals in the brain and can lead to problems with thinking, movement, behavior, and mood [57].

In 1957, Arvid Carlsson and Oleh Hornykiewicz established a link between DA deficiency in the BG and PD. DA is a neurotransmitter that contributes to the regulation of movement and emotion. The BG refers to a group of interconnected subcortical nuclei responsible primarily for motor control and other tasks such as motor learning, executive functions and behaviors, and emotions [58]. When DA levels in the BG decrease, it causes atypical brain activity, leading to impaired movement and other symptoms of PD [56].

The first clinical trial in PD treatment was carried out in 1961 and included the administration of intravenous levodopa [55]. Levodopa is a precursor to DA and can be used as a DA replacement agent for the treatment of PD. It is most effectively used to control bradykinetic symptoms. Bradykinesia is often used synonymously with akinesia and hypokinesia.

Bradykinesia, the slowness of a performed movement, can make simple tasks such as walking and getting out of a chair difficult and time-consuming [59]. Akinesia refers to loss of automatic movements (i.e. decreased ability to perform unconscious movements) including blinking, smiling, or arm swing when walking [56]. Other manifestations of akinesia are freezing and prolonged time required to initiate a movement [59]. Hypokinesia refers to decreased movement amplitude, which might lead to changes in writing. It may become hard to write, and the writing may appear small [56].

In 1982, William Langston discovered that some patients using synthetic heroin developed Parkinsonian features. The cause of this drug-induced parkinsonism was later found to be MPTP, which is toxic to DANs [55]. DANs are the primary source of DA in the central nervous system. They are found in the SNpc, which is rich in DA and has a high iron content [60]. The substantia nigra (SN) is a dopaminergic nucleus located in the midbrain of the brainstem. It is considered the primary input into the BG circuitry and, thus, a critical element to the function of the BG.

In the late 1900s, several studies reported an increased iron deposition in the SNpc of PD patients [61]. These reports were supported by the discovery of a significant rise in the Fe^{3+}-binding protein, ferritin, and a shift in the Fe^{2+}/Fe^{3+} ratio in favor of Fe^{3+} in the SNpc. With these studies' emergence, researchers began exploring a potential connection between excessive iron and dopaminergic neurodegeneration in the SNpc in PD. However, whether or not the iron deposition represents the primary event in PD remains controversial.

2.2.2 Functional Neuroanatomy of the BG

The BG and related nuclei can be divided into input nuclei, output nuclei, and intrinsic nuclei [58]. Input nuclei receive information from cortical, thalamic, and nigral sources. The caudate nucleus, the putamen, and the accumbens nucleus, together called the striatum, are input nuclei. Output nuclei send information to the thalamus, which projects to the cerebral cortex. The internal segments of the globus pallidus (GPi) and the SN pars reticulata (SNpr) are output nuclei. Intrinsic nuclei send information from the input nuclei to the output nuclei. The external segment of the globus pallidus (GPe), the subthalamic nucleus (STN), and the SNpc are intrinsic nuclei.

Information from the input nuclei to the output nuclei can go through two pathways: The direct and indirect pathways [58]. In the direct pathway, striatal neurons project directly to GPi and SNpr.

In the indirect pathway, striatal neurons project to GPe, which again project to STN, GPi, and SNpr. DANs in the SNpc contain the DA precursor NM. NM is a pigment that can be seen as a black color in necropsies of the human brain. Its pigmentation increases from early childhood to young adulthood, plateaus during middle age, and decreases in the sixth decade [61]. Functions of NM include redox activities, free radical scavenging, binding of different biomolecules, and chelation of iron [62]. NM has been shown to yield neuroprotective properties but also to promote neurodegeneration [63].

NM strongly binds to metals, especially iron [61]. Apart from iron in hemoglobin, the two primary forms of non-heme iron include transferrin for transportation and ferritin for storage [26]. Iron binds to NM in the ferric Fe^{3+} form. Iron, DA metabolism, and NM balance are crucial for cell homeostasis and can be disrupted under certain conditions [63]. NM is visible to MRI due to the paramagnetic properties of the NM-iron complex [64].

2.2.3 Pathogenesis

In PD, DANs in the SNpc degenerate, leading to DA depletion. DA depletion shifts the balance in BG activity toward the indirect circuit, leading to excessive activity of the STN that overstimulates the GPi and the SNpr. Increased output from the GPi and the SNpr over-inhibits the thalamus, reducing cortical neuronal activation associated with movement initiation. Also, when DANs degenerate, neuronal NM breaks down and is released from the cells, resulting in depigmentation [65].

With the advancements in MRI technology, growing evidence shows increased iron deposition in the SNpc of PD patients [61]. Moreover, increased iron deposition has also been observed in other brain regions in PD patients, including the putamen, the red nucleus, and the globus pallidus. Increased tissue iron levels may saturate iron-chelating sites on NM, and the higher concentration of toxic free iron resulting from the weaker association between iron and NM may lead to increased production of free radical species, which contribute to the observed neuronal damage in PD [61].

Increased nigral iron accumulation in PD appears to be stratified according to disease motor severity and correlates with symptoms related to dopaminergic neurodegeneration [66]. Other vital molecular events in the SNpc in PD include α-synuclein misfolding and aggregation, mitochondrial dysfunction, impairment of protein clearance, neuroinflammation, and oxidative stress [55].

PD is also characterized by more widespread pathology in other brain regions [67]. LC is a nucleus of NANs located within pons in the brainstem [68]. It is a significant source of noradrenaline in the brain. Polymerization of noradrenaline produces NM, which makes LC blue in color. In addition to DANs in the SNpc, NANs in the LC are also degenerating, which decreases the noradrenaline content in the frontal cortices and the hypothalamus, causing motor symptoms such as frozen gait and autonomic nerve symptoms and non-motor symptoms such as depression and sleeping problems.

2.2.4 Treatment

Current treatments of PD are targeted at disease symptoms and are not disease-modifying [55]. Levodopa, a precursor for DA, is the most commonly used drug for the symptomatic treatment of PD. The drug increases DA levels in the brain, leading to motor symptom relief. Anticholinergics, antiglutamatergics, and dopamine antagonists are other medications available for treating PD-related motor symptoms.

Non-motor symptoms of PD are treated with a variety of drugs, depending on the presence and severity of the symptoms. Deep brain stimulation is a surgical treatment used on patients with moderate and advanced disease with poor quality of life due to fluctuating response or levodopa-unresponsive symptoms.

An individualized therapeutic approach is required since the spectrum of motor and non-motor symptoms is broad. The development of disease-modifying treatments is challenging due to a lack of reliable biomarkers of progression and limited pathogenic understanding. In order to develop disease-modifying therapies, it is necessary to recognize the variable slopes of disease progression, reflecting the clinical heterogeneity of the disease. A potential neuroprotective or disease-modifying therapy uses α-synuclein monoclonal antibodies to minimize the accumulation of aggregated α-synuclein [69]. Another treatment in development is nicotinamide adenine dinucleotide replenishment therapy [70].

The PD diagnosis is based on motor symptoms of DA deficiency, including bradykinesia, rigidity, and tremor [24]. Motor symptoms generally appear when 50-60% of DANs in the SNpc are already lost, limiting the effectiveness of potential neuroprotective therapies. Prodromal PD, the latent phase of PD, represents an opportunity for early detection of PD. Early detection of PD could allow the initiation of neuroprotective therapies at a stage where a smaller percentage of DANs in the SNpc are lost, leading to more effective treatment.

The prodromal phase of PD can vary from 5 to more than 20 years. Non-motor clinical biomarkers for prodromal PD include olfactory loss, constipation, REM-sleep behavior disorder, depression, anxiety, and global cognitive deficit.

2.2.5 Motor Phenotypes

It is increasingly evident that PD is not a single entity but rather a heterogeneous neurodegenerative disorder [71]. Possible stratification uses include identifying biological subtypes that predict therapeutic response to symptomatic treatments, such as targeting DA deficiency, or disease-modifying treatments, such as targeting mitochondrial dysfunction [72]. PD patients can be divided into groups based on motor symptoms, cognitive features, age of onset, rate of progression, or a combination of these factors [73]. Most commonly, motor symptoms and age of onset are used to distinguish patients.

Already in 1990, Jankovic et al. provided support for the existence of motor phenotypes in PD by comparing TD patients (mean tremor score / mean PIGD score $\geq$ 1.5, N = 441) with PIGD patients (mean tremor score / mean PIGD score $\leq$, N = 233) [7]. The PIGD group reported significantly greater subjective intellectual, motor, and occupational impairment than the tremor group. The study also found that older age at onset combined with PIGD is associated with more functional disability than younger age at onset with TD.

nTD patients have shown to exhibit more severe loss of cells in the ventrolateral part of the SNpc than TD patients, which causes inhibition of the direct pathway [23]. In contrast, TD patients have shown to exhibit a more severe neuronal loss in the medial SNpc, which causes inhibition of the indirect pathway. Considering this evidence, it is hypothesized that the pathophysiology of nTD PD is mainly due to abnormal BG output [74]. In contrast, TD PD may involve additional compensatory mechanisms downstream.

2.2.6 Imaging Biomarkers

An ideal marker for prodromal PD would confirm its presence and provide information about the time until the onset of motor symptoms and the rate of progression [75]. Direct visualization of the SNpc and the LC with neuroimaging is an attractive diagnostic approach due to the specificity of degeneration in these brain areas.

Transcranial ultrasound has been used for many years to visualize structures in the midbrain. Increased echogenicity of the SNpc has been found in up to 90% of patients with PD [75]. However, hyperechogenicity does not appear to change with time or correlate with disease severity.

SPECT imaging can determine DA transporters (DAT) levels in the striatum, which are responsible for the uptake and clearance of DA from the synaptic cleft [22]. DAT-SPECT shows decreased striatal DAT uptake in patients with PD, indicating SNpc dopaminergic dysfunction [75]. PET imaging uses radiotracers such as [18F]-FDOPA to measure the levels of DAT and the activity of the aromatic L-amino acid decarboxylase enzyme in the brain, which estimates the levels of DA storage in the brain. Exposure to ionizing radiation is a significant disadvantage of both SPECT and PET, especially in longitudinal studies where the cumulative risk of repeated scans may be unacceptable.

MRI is commonly used in clinical practice to differentiate between Parkinsonian syndromes. However, reliable identification of the specific neurodegenerative signature of idiopathic PD has not been possible with the imaging sequences used in the clinical routine. Diffusion tensor imaging (DTI) has been used to assess the integrity of nigrostriatal fibers in PD. Although reduced fractional anisotropy in the SNpc in PD patients is reported in several studies, there is substantial variation in the results [75].

A promising finding with high field strength MRI with T2*-weighted sequences is that nigrosome 1 (N1) has a high signal intensity in HCs, but a signal loss in patients with PD [76]. N1, located within the dorsolateral SNpc, is the nigral area that sustains a maximal loss of DANs in PD. The high-intensity signal of N1 in HCs is called the 'swallowtail sign'. A follow-up study using 3T MRI with an SWI sequence replicated this finding [77]. Another study found that higher levels of iron in the SNpc, detected via filtered MRI phase images, were positively correlated with MDS-UPDRS III scores and bradykinesia-rigidity subscores, but not tremor subscores [66].

NM-sensitive modified T1W sequences, such as high-resolution turbo spin echo, display high signal intensity in NM-rich areas [64]. In 2006, Sasaki et al. found that in PD patients, the NM-MRI signal intensity in the LC and SNpc was greatly reduced, suggesting depletion of NM-containing neurons. Although the use of high-resolution turbo spin echo T1W images has been consistently applied for visualizing NM, the GRE sequence with a magnetization transfer contrast (MTC) has recently been demonstrated to achieve both sharper contrast and lower variability [78].

This thesis aims to investigate the feasibility of using multimodal MRI to differentiate PD patients from HCs and stratify PD patients based on their motor phenotype, specifically PIGD and TD. The loss of DANs in the SNpc and NANs in the LC is characterized by a decrease in NM content and an increase in iron content in these regions.

SWI, tSWI, and phase are qualitative iron-sensitive MRI techniques that have the potential to identify iron deposition related to the loss of DANs in the SNpc and NANs in the LC. Moreover, T2*, R2*, and meiSWIM are quantitative iron-sensitive MRI techniques that can not only identify but also quantify iron deposition related to the loss of DANs in the SNpc and NANs in the LC. As increased iron accumulation in PD appears to be positively correlated with disease motor severity, these iron-sensitive MRI techniques are promising imaging biomarkers for stratifying patients based on their motor phenotype.

Previous studies have shown that GRE sequences with MTC, which are NM-sensitive, can differentiate between PD patients and HCs [78]. However, in this thesis, GRE sequences without MTC will be used, making it unlikely that NM can be identified. However, if NM can be identified, it will show up as high intensity in the T1 map and the T1WE image.

Chapter 3

Methods

3.1 Participants

A total of 81 participants (60 PD patients and 21 HCs) were included through NeuroSysMed's STRAT-PARK study [4]. The study was performed in accordance with the Declaration of Helsinki, and all data were collected following written informed consent. Ethical approval for the study was obtained from the Regional Committees for Medical and Health Research Ethics (REK) in Norway (ID 74985).

NeuroSysMed is a research center for clinical treatment in neurology headed by Haukeland University Hospital (HUH) and the University of Bergen. STRAT-PARK is a population-based cohort study where a total of 1500-2000 PD patients and HC will be recruited from three clinical centers: HUH (Bergen, Norway), St. Olavs University Hospital (Trondheim, Norway), and the London Movement Disorders Centre (Ontario, Canada) [4]. Participants in the STRAT-PARK study are followed at yearly visits consisting of clinical investigation, neuroimaging, blood and CSF sampling, and muscle biopsies.

This thesis analyzed MRI images and clinical data from the participants' first visit. Initially, data from 81 participants were collected, 76 from HUH and five from St. Olavs University Hospital. However, the participant group from St. Olavs University Hospital was excluded from this thesis due to being few compared to the participant group from HUH. The gender and age of both PD patients and HCs are presented in Table 3.1, together with the Montreal Cognitive Assessment (MoCA) score and the Brief Smell Identification Test (B-SIT) score. The MoCA is a rapid screening tool for mild cognitive dysfunction, while the B-SIT is a 5-minute olfactory function screening test.

Table 3.1: Characteristics of PD patients and HCs. All clinical features except gender are presented as mean±standard deviation (min-max). PD = Parkinson's disease, HC = healthy control, SD = standard deviation, MoCA = Montreal Cognitive Assessment, B-SIT = Brief Smell Identification Test.

Status		PD (n=58)	HC (n=18)
Gender	Female	22	12
	Male	36	6
Age [years]	Mean ± SD	69±9	65±10
	Min-max	32-84	45-80
MoCA score	Mean ± SD	24±4	26±2
	Min-max	15-31	22-28
B-SIT score	Mean ± SD	6±2	10±2
	Min - max	0-11	6-11

3.2 Clinical Data

Table 3.2 describes patients' clinical examinations. The severity of PD was assessed using the MDS-UPDRS. This scale comprises four parts: part I (non-motor experiences of daily living), part II (motor experiences of daily living), part III (motor examination), and part IV (motor complications). The total MDS-UPDRS score includes scores from parts I-IV.

Table 3.2: Characteristics of patients with the PIGD, TD, and indeterminate motor phenotype. All clinical features except gender are presented as mean±standard deviation (min-max). PIGD = postural instability gait difficulty, TD = tremor-dominant, MDS-UPDRS = Movement Disorder Society Unified Parkinson's Disease Rating Scale, MDS-NMS = Movement Disorder Society Non-Motor Rating Scale, NMF = non-motor fluctuation, MoCA = Montreal Cognitive Assessment, B-SIT = Brief Smell Identification Test.

Motor phenotype	PIGD (n=27)	TD (n=20)	Indeterminate (n=11)
Gender (female, male)	(12, 15)	(6, 14)	(4, 7)
Age [years]	70±10 (32-84)	68±8 (56-81)	70±7 (57-79)
Age of diagnosis [years]	64±11 (30-78)	64±8 (55-79)	65±11 (57-73)
Age of symptoms [years]	62±10 (29-77)	62±8 (50-77)	62±7 (54-72)
Disease duration [months]	84±50 (14-228)	66±30 (35-132)	58±22 (30-84)
MDS-UPDRS I score	10±5 (4-23)	8±4 (2-17)	8±6 (3-18)
MDS-UPDRS II score	10±6 (1-24)	7±5 (1-24)	6±3 (3-10)
MDS-UPDRS III score	26±12 (4-57)	27±10 (10-48)	26±11 (12-47)
MDS-UPDRS IV score	3±3 (0-9)	1±2 (0-7)	1±3 (0-7)
Total MDS-UPDRS score	48±18 (12-90)	43±15 (17-68)	35±11 (18-45)
Hoen & Yahr score	2.1±0.5 (1-3)	1.9±0.4 (1-2)	1.9±0.7 (1-3)
Total MDS-NMS score	60±32 (16-155)	44±24 (19-98)	51±66 (7-181)
Total MDS-NMS NMF score	1±2 (0-6)	0.2±0.7 (0-3)	0±0 (0-0)
MoCA score	24±3 (18-30)	26±4 (19-31)	21±4 (15-27)
B-SIT score	6±2 (2-10)	6±2 (3-11)	6±3 (0-10)

The Hoehn and Yahr scale was used to describe the progression of motor symptoms, while the MDS Non-Motor Rating Scale (MDS-NMS) was used to assess non-motor symptoms. The MDS-NMS Non-Motor Fluctuations (NMF) subscale was used to determine whether and to what extent non-motor symptoms fluctuated.

A clinician calculated the time of diagnosis, the time at which motor symptoms first appeared, and the number of years since the onset of motor symptoms based on patient records and converted these to the patient's biological age. Additionally, a clinician determined the motor phenotype of each patient based on MDS-UPDRS part III, with the options being TD, PIGD, and indeterminate. The classification of motor phenotype excluded patients with an indeterminate motor phenotype since they did not represent a distinct patient group but rather a combination of PIGD and TD phenotypes.

3.3 MRI Acquisition

All imaging was conducted using a 3T Siemens Biograph mMR scanner, a simultaneous MRI-PET scanner manufactured by Siemens Healthineers in Erlangen, Germany. Seventy-six participants were scanned at Haukeland University Hospital in Bergen, Norway. A 16-channel head and neck coil was utilized for signal reception.

The imaging protocol included three sequences of particular interest to this study: A 3D MP-RAGE sequence and two triple-echo GRE sequences with different FAs, as shown in Table 3.3. The 3D MP-RAGE sequence was chosen for its excellent T1W contrast, rendering it suitable for normalization to a T1W template. The selection of GRE sequences enabled the use of Strategically Acquired GRE (STAGE) imaging.

Table 3.3: Specifications of the MRI protocol. MP-RAGE = magnetization-prepared rapid gradient echo, ME-GRE = multi-echo gradient echo, PDW = proton density weighted, T1W = T1-weighted, FOV = field of view, TR = repetition time, TE = echo time, FA = flip angle, BW = bandwidth, TA = acquisition time.

Sequence	3D MP-RAGE	3D ME-GRE (PDW)	3D ME-GRE (T1W)
FOV [mm] x Phase FOV	256 x 100%	384 x 75%	
Scanning matrix	256 x 256	384 x 288	
Slice thickness [mm]	1	1.34	
Voxel size [mm^3]	1 x 1 x 1	0.67 x 0.67 x 1.34	
TR [ms]	2400	29	
TE [ms]	2.26	7.5/15/22.5	
FA [degrees]	8	6	27
Pixel BW [Hz/pixel]	200	210/160/160	
TA [min:sec]	6:06	4:59	

3.4 Generating Images and Maps

STAGE Research (version 2.8.1) from SpinTech MRI (Detroit, MI, USA) was used to create the images and maps in Table 2.1 and Table 2.2 from the two ME-GRE sequences based on the framework proposed by Haacke et al. [43][44][79]. SPIN Research (version 1.5.12) from SpinTech MRI (Detroit, MI, USA) was used to sort the images from the MRI acquisition before using STAGE.

Figure 3.1 shows the data processing workflow for quantifying T1 and PrD. To calculate $T1_{app}$, two magnitude images with FAs 6° and 27° were used together with Equation 2.25, assuming k was equal to unity everywhere.

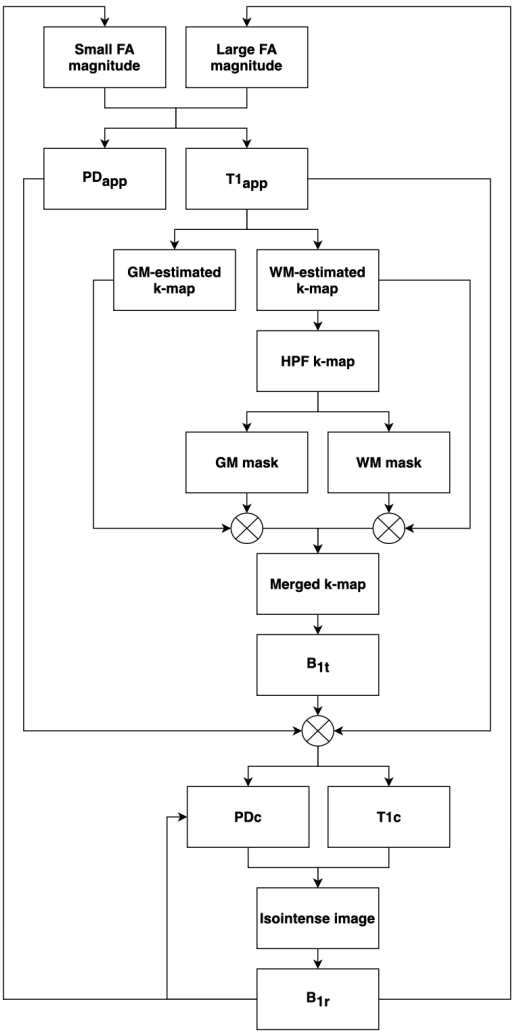

Figure 3.1: Data processing workflow for quantifying T1 and PrD. FA = flip angle, PD_{app} = apparent proton density, $T1_{app}$ = apparent longitudinal relaxation time, GM = gray matter, WM = white matter, k = radio frequency transmit field variation, HPF = high-pass filtered, B_{1t} = radio frequency transmit field, PD_c = corrected proton density, $T1_c$ = corrected longitudinal relaxation time, B_{1r} = radio frequency receive field

A mask representing pixels containing only WM was generated based on the $T1_{app}$ map by selecting all apparent values in the 800-1200 ms range as WM. The T1 value for WM was set to 900 ms everywhere, and a WM-estimated k map was calculated using Equation 2.28. The resulting apparent k map was high-pass filtered to remove low-frequency changes and allowing the tissue types to be separated by thresholding the image. The resulting high-pass filtered image is referred to as the k map. Masks were created for WM, GM, and CSF.

The next step was to obtain the correct k values for GM. The T1 of GM was globally fixed to be 1600 ms to obtain a GM-estimated k map using Equation 2.28. The WM and GM k maps were created by multiplying their respective masks with the estimated k maps and setting values outside the range of 0.5 to 1.2 to zero to prevent the GM from being included as WM and vice versa. These two k maps were then merged by adding them together. To remove noise spikes, a small sliding window calculated the mean and standard deviation, and points outside the standard deviation were set to zero. Finally, a second-order local polynomial fit was used to estimate the global B_{1t} field.

Once the B_{1t} was determined, it was utilized in Equation 2.25 to obtain T1c and PDc maps, that is, corrected T1 and PrD. The T1c and PDc maps were used to generate an image with a specific FA that made WM and GM isointense, so that the image's amplitude variation would represent the B_{1r} field. A sliding window was employed to eliminate bright vessels and noise spikes in the WM/GM isointense image. A third-order local polynomial fit was used to account for the relatively strong non-uniformities of the receive coil sensitivity, particularly near the coil elements at the edge of the brain. Once the B_{1r} field was known, the original two FA images and PrD maps was corrected by dividing them by the B_{1r} field.

The rest of the STAGE data processing workflow is described in Figure 3.2. The PrD image was subtracted from the T1W image for each of the three echoes, and the three T1WE images were averaged according to Equation 2.24. A constant was added to the T1WE image to avoid negative values. The SWI and tSWI images were generated from the third echo of the 6° scan. The $R2^*$ map were calculated using the three echoes for each scan, then averaged.

A brain mask was created using a brain extraction tool and adjusted using a phase image quality map to eliminate areas with rapid phase changes [80][81]. The resulting masked phase images at each echo were processed using sophisticated harmonic artifact reduction for phase data to remove the background field.

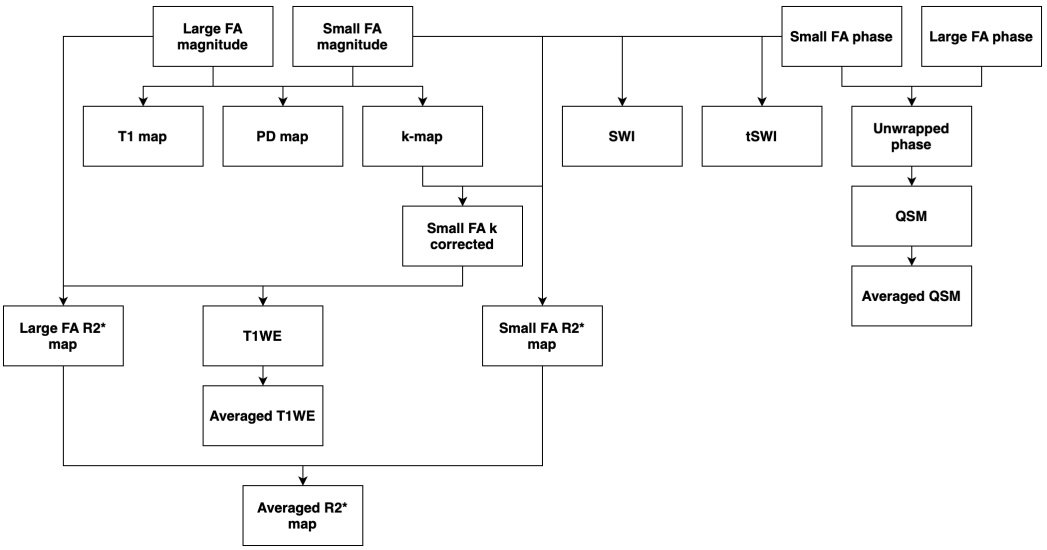

Figure 3.2: STAGE data processing workflow. R2 is the reciprocal of T2*. FA = flip angle, T1 = longitudinal relaxation time, PD = proton density, k = radio frequency transmit field variation, T2* = transverse relaxation time, T1WE = T1-weighted enhanced image, SWI = susceptibility-weighted image, tSWI = true susceptibility-weighted image, QSM = quantitative susceptibility map.*

QSM images were generated for each echo using a truncated k-space division algorithm [82][83]. Additionally, the R2* map was utilized as a geometry mask for deep GM and veins during the iterative QSM calculation. Finally, a combined QSM image was generated by weighted averaging of all single echo QSM images to improve the SNR of the final QSM image.

3.5 Image Preprocessing

The images and maps from STAGE were preprocessed through co-registration and normalization. Co-registration aligned the functional and structural images, while normalization moved all images to a standardized space using a template.

3.5.1 Co-registration

Co-registration refers to the process of aligning functional and anatomical images. Typically, registration algorithms begin by aligning the outlines of the images [84]. The algorithm then adjusts the relative positions of the images and assesses the correspondence of the voxels in one image with those in the other for each orientation. The registration algorithm also considers the potential differences in contrast weighting between the anatomical and functional images. Once the optimal alignment has been determined, the transformation is applied and saved as a matrix for future use.

Linear registration tools employ affine transformations to warp images to a template. These transformations are referred to as linear transformations because a transformation applied in one direction along an axis is counteracted by an equal transformation in the opposite direction. The four types of affine transformations are translation, rotation, zoom, and shear, each with three degrees of freedom that can be applied along the x-, y-, or z-axis. Consequently, affine transformations have twelve degrees of freedom.

FMRIB's Linear Image Registration Tool (FLIRT) from the FMRIB Software Library (FSL) was utilized to co-register the STAGE images and maps with the MP-RAGE image [85][86][87]. The MP-RAGE image was the reference image, while the T1WE image from STAGE was the input image. As a result, the T1WE image was aligned with the MP-RAGE image. The linear transformation that registered the T1WE image to the MP-RAGE image was saved as a 4x4 affine matrix. FSL FLIRT then applied the saved transformation to all the other images and maps in Table 2.1 and Table 2.2. The T1WE image was selected to make the transformation matrix because it resembled the MP-RAGE image the most.

3.5.2 Normalization

Normalization is the process of transforming each brain to have the same size, shape, and dimensions as a template. A template is a brain with standard dimensions and coordinates. This thesis employed the commonly used Montreal Neurological Institute (MNI152) template, an average of 152 T1W images of healthy adult brains. The MP-RAGE image was normalized to the MNI152 T1 1mm brain template using Statistical Parametric Mapping (SPM12), and the transformation was saved [88]. The same transformation was then applied to the co-registered images and maps, placing them in the standardized MNI152 1mm space. As a result, the SNpc and LC of each participant were now in alignment with one another.

The standard normalization procedure in SPM12 was inadequate for three of the participants. For two of them, normalization in FSL FLIRT with the MNI152 T1 1mm brain template as a reference and the MP-RAGE image as input was performed instead of normalization in SPM12. For the last participant, the origin and angle of the MP-RAGE image were set manually before normalization in SPM12 for the results to be acceptable.

3.6 Region of Interest Analysis

One way to create a region for a region of interest (ROI) analysis is to use an atlas that partitions the brain into anatomically distinct regions. The SNpc atlas used in this thesis was a probabilistic atlas generated using a 3T T1W fast spin-echo NM-MRI sequence [89][90]. The atlas was constructed by manually delineating SNpc ROIs in 27 HCs (age 39$\pm$11 years) and registering them to the MNI152 1mm template using Advanced Normalization Tools software. The `fslmaths` function in FSL was used to divide the SNpc mask into left and right hemispheres [91].

The LC atlas used in this thesis was as a population-based probabilistic atlas generated from magnetization transfer images of 53 healthy volunteers (aged 52 to 84 years) using 7T MRI [92][93]. Individual LC binary masks were averaged to obtain an atlas with a resolution of 0.4x0.4x0.5 mm. The `fslmaths` function in FSL was used to convert the atlas into a binary mask image with a zero threshold to prepare the atlas for further analysis. The image was then resampled to a resolution of 1x1x1 mm using the `flirt` function in FSL. Finally, a new mask image with a zero threshold was created and divided into left and right hemispheres using the `fslmaths` function in FSL.

Once masks of the ROIs were created, MRI data was extracted for each participant. The `fslmaths` function is FSL was used to apply masks of six structures, namely the left SNpc, the right SNpc, the SNpc, the left LC, the right LC, and the LC, to all the 38 normalized images and maps, and compute the mean intensity of all the voxels in the masks. This resulted in a total of 228 MRI features.

3.7 Laterality

The ROI analysis yielded mean intensities for both the left and right sides of the SNpc and LC, as well as the total structures. However, if the left and right structures display the same mean intensity, including both sides in the subsequent analysis is unnecessary. Therefore, the significance of laterality in the SNpc and the LC across all STAGE images was determined with statistical hypothesis testing.

Different statistical methods employed in data analysis rely on assumptions about normality, such as correlation, regression, *t*-tests, and analysis of variance [94]. Normality refers to a specific statistical distribution called the Gaussian distribution, also known as the normal distribution or bell-shaped curve [95]. The Gaussian probability density function for a random variable x is defined as follows:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{3.1}$$

Here, μ and σ^2 are the mean and variance of x, respectively, given by:

$$\mu = E[x] = \int_{-\infty}^{+\infty} xp(x)dx \tag{3.2}$$

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)dx \tag{3.3}$$

Here, $E[\cdot]$ denotes the random variable's mean or expected value. The normality assumption is met if the data distribution is similar to the Gaussian probability density function, meaning the values fall into a bell shape. If the two populations satisfy the normality assumption, parametric tests can be used to compare the means of the populations [94]. However, if the data is not normally distributed, the mean does not adequately represent the data. In that case, non-parametric methods are employed, comparing medians between groups.

The *central limit theorem* states that if a random variable is the outcome of a summation of several independent random variables, its probability density function approaches the Gaussian function as the number of summands tends to infinity [95]. In practice, the central limit theorem states that when a population has $N = 100$ or more observations, violation of normality is not a major issue [94]. However, since all populations in this thesis had $N < 100$ observations, their normality was tested.

Two commonly utilized tests for assessing normality are the Kolmogorov-Smirnov and the Shapiro-Wilk tests [94]. The Shapiro-Wilk test is more suitable for smaller sample sizes ($N < 50$), whereas the Kolmogorov-Smirnov test is employed for larger sample sizes ($N \geq 50$).

With $N = 76$ participants, the Kolmogorov-Smirnov test was employed to examine the normality of the lateral MRI features. The Kolmogorov-Smirnov test uses N independent and identically distributed observations of the variable being studied [96]. The empirical cumulative distribution function of the ordered observations X_i is defined as follows:

$$F_N(x) = \frac{\text{number of (elements in the sample} \leq x)}{N} = \frac{1}{N}\sum_{i=1}^{N} 1_{(-\infty,x]}(X_i) \tag{3.4}$$

Here, the indicator function $1_{(-\infty,x]}(X_i)$ takes the value 1 if $X_i < x$ and 0 otherwise. The Kolmogorov-Smirnov statistic is calculated as follows:

$$D_N = \sup_x |F_N(x) - F_E(x)| \tag{3.5}$$

In this equation, $\sup_x$ represents the supremum of the set of distances and $F_E(x)$ is the cumulative distribution function associated with the expected probability distribution function of the variable of interest. The null hypothesis, denoted as $H_0 : F_N(x) = F_E(x)$ for all x, is rejected if $D_N(x)$ exceeds a critical value D_α at a predetermined significance level α.

To employ the Kolmogorov-Smirnov test on the lateral MRI features, the `kstest` function from the `scipy.stats` library in Python was utilized. The `cdf` parameter was set to `norm`, enabling the normal distribution as the expected distribution. A significance level of 95% was selected, meaning that the null hypothesis was rejected in favor of the alternative hypothesis if the p-value was below 0.05. The obtained p-values were less than 0.05 for all the lateral MRI features, indicating that none followed a normal distribution. Therefore, a non-parametric method was employed when determining the laterality.

The Wilcoxon signed-rank test is a non-parametric paired test [97]. It tests the null hypothesis that the medians of two paired datasets are identical, implying that the paired samples originate from the same distribution. Specifically, it evaluates whether the distribution of differences between the values in the two classes is symmetric around zero. To determine the laterality of the SNpc and the LC in the MRI images, the `wilcoxon` function from the Python library `scipy.stats` was utilized. Three significance levels were chosen: 95% ($p < 0.05$) to indicate images with significant laterality, 99% ($p < 0.01$) to denote images with highly significant laterality, and 99.9% ($p < 0.001$) to represent images with strongly significant laterality.

3.8 Preprocessing Features

Preprocessing includes the processes performed on the features prior to their utilization. Normalizing a dataset is a typical prerequisite for numerous machine learning estimators. The `StandardScaler` function from the `sklearn.preprocessing` library and the `make_pipeline` function from the `sklearn.pipeline` module in Python was used to standardize the features by eliminating the mean and scaling them to have unit variance [98].

The categorical features were converted into numerical variables using encoding, where a unique numerical value is assigned to each category in categorical features. The `LabelEncoder` function from the `sklearn.preprocessing` library in Python was used to encode PD/HC status, PIGD/TD motor phenotype, and gender [98]. Once the categorical features were encoded, they could be included with any numerical features in the classification.

3.9 Feature Selection

Working with high-dimensional data in machine learning can give rise to various problems, collectively known as the *curse of dimensionality* [95]. One of these problems is that using features with high mutual information can increase complexity without providing many benefits.

Another factor is the required generalization properties of the classifier. Generalization refers to the classifier's ability to work effectively with data outside the training set. In general, the higher the ratio of the number of training observations to the number of features, the better the generalization properties of the resulting classifier. Therefore, for the limited number of training observations in this thesis, it was desirable to keep the number of features to a minimum to design classifiers with good generalization capabilities.

Feature selection is the process of selecting the essential features to reduce their number and, at the same time, retain as much of their class-discriminatory information as possible. In this thesis, feature selection was done based on statistical hypothesis testing and pairwise correlation. The goal was to identify features that resulted in a large between-class distance and small within-class variance in the feature vector space.

3.9.1 Feature Selection Based on Statistical Testing

During the feature selection process, the initial step often involves assessing the discriminatory potential of each feature individually for the specific problem at hand [95]. A common approach for this is statistical hypothesis testing. As discussed in Section 3.7, it is crucial to evaluate the normality of the data before choosing a test. The normality of features within the PD group ($N = 58$) was evaluated using the Kolmogorov-Smirnov test. The test was conducted utilizing the `kstest` function from the Python library `scipy.stats`, with the `cdf` parameter set to `norm` [98].

Conversely, the normality of features within the HC group ($N = 18$), PIGD group ($N = 27$), and TD group ($N = 20$) was assessed using the Shapiro-Wilk test. Like the Kolmogorov-Smirnov test, the Shapiro-Wilk test examines the null hypothesis that the data is derived from a normal distribution. It was implemented using the `shapiro` function from the Python library `scipy.stats` [98].

For features that displayed normality ($p \geq 0.05$) in both the PD and HC groups, the discriminatory capacity in the PD/HC classification was evaluated using the Student's unpaired t-test. The Student's unpaired t-test assesses whether the means estimated from two independent samples differ significantly [97]. It is a parametric method employed through the `ttest_ind` function from the `scipy.stats` library in Python [98].

Conversely, for features that did not exhibit normality ($p < 0.05$) in either the PD or HC group, the Mann-Whitney U test was used to evaluate their discriminatory ability in the PD/HC classification. The Mann-Whitney U test tests the null hypothesis that the medians of two independent datasets are equal. It is the non-parametric counterpart of the Student's unpaired t-test and was implemented using the `mannwhitneyu` function from the Python library `scipy.stats` [98].

The same approach was applied for the comparison between the PIGD group and the TD group. Features with a p-value of less than 0.05 were deemed significant, features with a p-value of less than 0.01 were considered highly significant, and features with a p-value of less than 0.001 were regarded as strongly significant. These groups of features were subsequently employed for the classification tasks.

The distribution of different classes was visualized using the `boxplot` function from the `seaborn` library in Python for some of the features. The boxplot is a frequently used graphical technique for analyzing univariate data [99]. The boxplot is constructed in four steps. The first step is to draw a line at the height of the sample median Q_2. Then, draw a box from the first quartile Q_1 to the third quartile Q_3. Q_1 is the value under which 25% of data points are found when arranged in increasing order. Q_3 is the value under which 75% of data points are found when arranged in increasing order. The length of this box equals the interquartile range $IQR = Q_3 - Q_1$, which is a robust scale measure. Next, classify all points outside the fence given by

$$[Q_1 - 1.5\,IQR,\ \ Q_3 + 1.5\,IQR]$$

as potential outliers and mark them on the plot. Finally, draw the whiskers as the lines that go from the ends of the box to the most remote points within the fence.

3.9.2 Feature Selection Based on Pairwise Correlation

An alternative approach for feature selection involves removing features that share redundant information with other features, which can be achieved by computing the correlation coefficient between all the features. In cases where the population does not follow a normal distribution, it is necessary to use a correlation measure that does not rely on assumptions about the population parameters. To address this, the observations can be ranked based on their magnitudes and calculations can be performed using these ranks instead of the original values [100]. Spearman's rank correlation coefficient is a metric that captures this idea and is defined as:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{3.6}$$

In the Equation above, ρ represents the rank coefficient of correlation, d_i denotes the difference in ranks between paired items in two series, and n represents the number of observations. Spearman's rank correlation coefficient assesses the strength and direction of the relationship between two variables, ranging from -1, indicating a perfect negative correlation, to +1, indicating a perfect positive correlation. A value of 0 signifies no correlation.

The `corr` function from the `pandas.DataFrame` library in Python with the `method` parameter set to `spearman` was used to compute the pairwise Spearman's rank correlation coefficient for all 228 MRI features. Then, an iterative process was applied to these features. If the absolute correlation between two features exceeded a specific threshold, one of the features was discarded. This procedure was repeated until only features with an absolute correlation below the threshold remained. Nine different thresholds were utilized: 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, and 0.1. The resulting nine sets of selected features were subsequently employed for classification purposes.

3.10 Differentiating PD from HC and TD from PIGD

3.10.1 Datasets

Classifying participants into PD patients and HCs involved using 12 sets of MRI features and two sets of clinical features, resulting in 24 distinct datasets. The MRI feature sets utilized were as follows:

1. All MRI features.
2. Features selected based on statistical hypothesis testing.
   (a) Significant features.
   (b) Highly significant features.
3. Features selected based on pairwise correlation.
   (a) Features with $|\rho| < 0.9$.
   (b) Features with $|\rho| < 0.8$.
   (c) Features with $|\rho| < 0.7$.
   (d) Features with $|\rho| < 0.6$.
   (e) Features with $|\rho| < 0.5$.
   (f) Features with $|\rho| < 0.4$.
   (g) Features with $|\rho| < 0.3$.
   (h) Features with $|\rho| < 0.2$.
   (i) Features with $|\rho| < 0.1$.

Additionally, two sets of clinical features were considered:

1. No clinical features.
2. All clinical features (see Table 3.1).

Classifying patients into PIGD and TD motor phenotypes involved using 12 sets of MRI features and three sets of clinical features, resulting in 36 distinct datasets. The MRI feature sets used were the same as those employed for the PD/HC classification. However, the sets of clinical features were defined as follows:

1. No clinical features.
2. Selected clinical features (see Table 3.1).
3. All clinical features (see Table 3.2).

The data was split into training and test sets using the `train_test_split` function from the `sklearn.model_selection` library in Python [98]. The `test_size` parameter was set to 0.2 so that 80% of the data was used for training and 20% for testing. The `stratify` parameter was set to the target array `y` to split the data stratified. *Stratified sampling* is a technique that allows the splitting of the data by preserving the distribution percentage of each class on the training and test set. This technique benefits imbalanced datasets by ensuring that the learning algorithms are proportionally trained with all classes [101].

3.10.2 Classification

Participants were classified into PD patients and HCs, and patients were further classified into PIGD and TD, using five different supervised machine learning algorithms:

1. Logistic regression (LR) classification.
2. Decision tree (DT) classification.
3. Random forest (RaF) classification.
4. Support vector classification (SVC).
5. Extreme gradient boosting (XGBoost) classification.

These specific algorithms were selected based on their widespread recognition and ease of implementation. XGBoost was chosen specifically for its built-in capability to handle missing values by default. When dealing with datasets with clinical features and missing values, XGBoost was the sole model for the classification tasks.

Supervised machine learning problems involve using training data with multiple features denoted as X_i to predict a target variable y_i [102]. A mathematical model predicts y_i based on the input X_i. The model's parameters θ represent the unknown variables learned from the available data. Training the model involves determining the optimal values for the parameters, which yield the best fit to the training data X_i and their corresponding labels y_i. An objective function measures the quality of the model's fit to the training data. The objective function, comprising the training loss and a regularization term, can be expressed as:

$$obj(\theta) = L(\theta) + \Omega(\theta) \tag{3.7}$$

Here, L represents the training loss function, quantifying the model's predictive accuracy on the training data. Ω denotes the regularization term, which controls the complexity of the model. Managing model complexity is crucial to prevent overfitting. The most common supervised learning tasks are classification, which involves separating data into different categories, and regression, which focuses on fitting data to a continuous function.

LR is a statistical model based on probability commonly employed in machine learning to address classification problems [102]. In LR, the estimation of probabilities is accomplished using a logistic function, also known as the sigmoid function, which is mathematically defined as:

$$g(z) = \frac{1}{1+e^{-z}} \tag{3.8}$$

In this thesis, LR was conducted using the `LogisticRegression` function from the `sklearn.linear_model` library in Python [98]. The `class_weight` hyperparameter was set to `balanced` to address the class imbalance in the input data. This adjustment ensured that the weights were inversely proportional to the class frequencies, effectively handling the class imbalance issue. Apart from that, default hyperparameters were utilized. The DT, RaF, and SVC classifications used the same hyperparameter configuration.

DT is a non-parametric supervised learning method [102]. The DT learning process involves classifying instances by traversing the tree from the root to the appropriate leaf nodes. Each node corresponds to an attribute, and the instance is classified based on the attribute value determined by that node. The most commonly used criteria for splitting in DT is the Gini impurity which is defined as:

$$\text{Gini}(E) = 1 - \sum_{i=1}^{c} p_i^2 \tag{3.9}$$

In this thesis, DT was performed using the `DecisionTreeClassifier` function from the `sklearn.tree` library in Python [98].

RaF is an ensemble classification technique [102]. This method employs parallel ensembling by training multiple DT classifiers in parallel on different subsets of the dataset, and the final result is obtained through majority voting or averaging. In this thesis, RaF was implemented using the `RandomForestClassifier` function from the `sklearn.ensemble` library in Python [98].

SVC is a specific implementation of a support vector machine. In SVC, a hyperplane or a set of hyperplanes is constructed in high-dimensional space [102]. The idea is to find a hyperplane that maximizes the distance from the nearest training data points in each class, as a larger margin generally leads to a lower generalization error of the classifier. In this thesis, SVC was performed using the `SVC` function from the `sklearn.svm` library in Python [98].

XGBoost is a variant of gradient boosting, which, similar to RaF, is an ensemble learning algorithm that combines multiple individual models, typically DTs [102]. XGBoost calculates second-order gradients of the loss function to minimize loss and utilizes advanced regularization techniques such as L1 and L2 regularization to address overfitting, thereby enhancing model generalization and performance. In this thesis, XGBoost was implemented using the `XGBClassifier` function from the `xgb` library in Python. Default hyperparameters were utilized.

3.10.3 Evaluation

The Confusion Matrix

Evaluating a classifier's predictive performance is crucial to determine its usefulness and compare it with other methods [103]. The evaluation is conducted on test data that the machine learning model has not been exposed to before. In binary classification, the data is divided into two distinct classes: positives and negatives. The binary classifier then assigns each data instance as either positive or negative. This classification results in four possible outcomes: two correct classifications, namely true positives (TP) and true negatives (TN), and two incorrect classifications, namely false positives (FP) and false negatives (FN).

The `confusion_metrics` function in the `sklearn.metrics` library in Python was used to compute a *confusion matrix* on the test data for each classification [98]. A confusion matrix is a 2x2 matrix with elements $A(i, j)$, representing the number of data points whose actual class label is i and were classified as class j [95]. Figure 3.3 shows the layout of a binary confusion matrix.

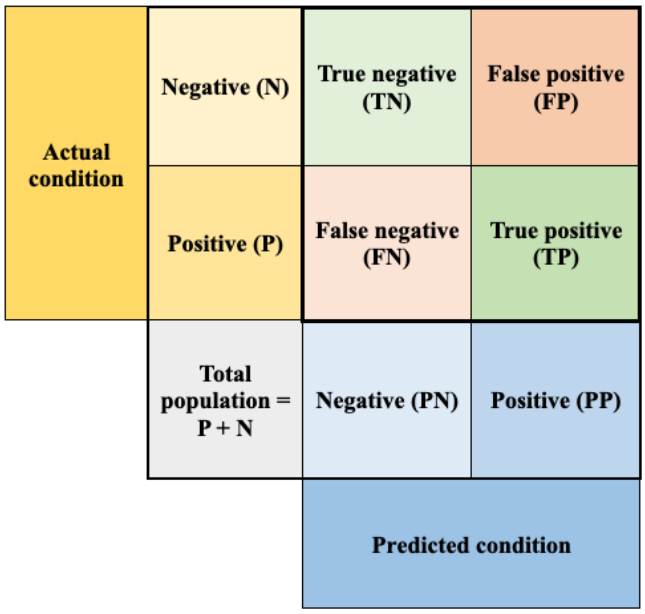

Figure 3.3: The layout of a binary confusion matrix. PP = predicted positive, PN = predicted negative.

In the PD/HC classification: P represents PD patients, N represents HCs, PP represents participants classified as PD patients, PN represents participants classified as HCs, TP represents PD patients correctly classified, FN represents PD patients wrongly classified, FP represents HCs wrongly classified, and TN represents HCs correctly classified.

In the PIGD/TD classification: P represents TD patients, N represents PIGD patients, PP represents patients classified as TD patients, PN represents patients classified as PIGD patients, TP represents TD patients correctly classified, FN represents TD patients wrongly classified, FP represents PIGD patients wrongly classified, and TN represents PIGD patients correctly classified.

Due to the small sample sizes, the confusion matrix was significantly influenced by choice of `random_state` hyperparameter in the train/test split. Therefore, one hundred train/test splits were conducted with `random_state` ranging from 0 to 99. The 100 resulting confusion matrices were then averaged.

The ideal binary classifier has a confusion matrix with TP=P, FN=0, FP=0, and TN=N. However, in reality, this is rarely the case. Therefore, the elements in the confusion matrix are often combined to make metrics that can evaluate binary classifiers.

Precision and Recall

Previous research has indicated that precision and recall are appropriate metrics for evaluating binary classifiers on imbalanced datasets [103]. Precision refers to the percentage of data points classified as class i, whose actual class label is i [95]. The precision of the negative class is referred to as the negative predictive value (NPV) given by:

$$NPV = \frac{TN}{PN} = \frac{TN}{TN+FN} \tag{3.10}$$

A model with low NPV may identify many negatives, but its selection process is prone to noise, resulting in numerous false negative predictions. On the other hand, a model with high NPV accurately identifies true negatives, even if it may miss some negatives. The precision of the positive class is referred to as the positive predictive value (PPV) given by:

$$PPV = \frac{TP}{PP} = \frac{TP}{TP+FP} \tag{3.11}$$

A model with low PPV may identify many positives, but its selection process is prone to noise, resulting in numerous false positive predictions. On the other hand, a model with a high PPV accurately identifies true positives, even if it may miss some positives. Macro average precision (MAP) is defined as the arithmetic average of the two classes' precision given by:

$$MAP = \frac{NPV+PPV}{2} \tag{3.12}$$

Weighted average precision (WAP) is defined as the weighted mean of the two classes' precision given by:

$$WAP = \frac{P}{P+N} \cdot PPV + \frac{N}{N+P} \cdot NPV = \frac{P \cdot PPV + N \cdot NPV}{P+N} \tag{3.13}$$

The recall is the percentage of data points with true class label i, correctly classified in that class [95]. The recall of the negative class is referred to as the true negative rate (TNR), also called the specificity, and is given by:

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} \tag{3.14}$$

A model with high TNR effectively detects negative cases in the data, even though it may also mistakenly identify some positive cases as negative. Conversely, a model with low TNR fails to find a significant portion of the negative cases in the data. The recall of the positive class is referred to as the true positive rate (TPR), also called the sensitivity, and is given by:

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} \tag{3.15}$$

A model with high TPR effectively detects positive cases in the data, even though it may also mistakenly identify some negative cases as positive. Conversely, a model with low TPR fails to find a significant portion of the positive cases in the data. Macro average recall (MAR) is defined as the arithmetic average of the two classes' recall given by:

$$MAR = \frac{TPR+TNR}{2} \tag{3.16}$$

Weighted average recall (WAR) is defined as the weighted mean of the two classes' recall given by:

$$WAR = \frac{P}{P+N} \cdot TPR + \frac{N}{N+P} \cdot TNR = \frac{P \cdot TPR + N \cdot TNR}{P+N} \tag{3.17}$$

In an ideal scenario, a model should identify all positive cases accurately while minimizing the number of false positives and identify all negative cases accurately while minimizing the number of false negatives. However, there is often a trade-off between precision and recall.

The F1 Score and the Classification Report

The F1 score, also known as the dice similarity coefficient when applied to binary data, is a metric that combines precision and recall by calculating their harmonic mean. The F1 score of the negative class is referred to as the negative F1 (NF1) score and is given by:

$$NF1 = \frac{2NPV \cdot TNR}{NPV + TNR} = \frac{2TN}{2TN + FN + FP} \tag{3.18}$$

The F1 score of the positive class is referred to as the positive F1 (PF1) score and is given by:

$$PF1 = \frac{2PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN} \tag{3.19}$$

The macro average F1 (MAF1) score is defined as the arithmetic average of the two classes' F1 score given by:

$$MAF1 = \frac{PF1 + NF1}{2} \tag{3.20}$$

The weighted average F1 (WAF1) score is defined as the weighted mean of the two classes' F1 score given by:

$$WAF1 = \frac{P}{P+N} \cdot PF1 + \frac{N}{N+P} \cdot NF1 = \frac{P \cdot PF1 + N \cdot NF1}{P+N} \tag{3.21}$$

When both precision and recall are high, the model will attain a high WAF1 score, whereas low values for both precision and recall will lead to a low WAF1 score. A model with either high precision or high recall, while the other metric is low, will achieve a moderate WAF1 score. The WAF1 score can be interpreted as a measure of the model's overall performance on a scale from 0 to 1, where 1 represents the best performance.

The WAF1 score is widely used in classification models as it yields reliable results for balanced and imbalanced datasets while considering the model's precision and recall capabilities [103]. The WAF1 score was therefore chosen as the primary evaluation metric in this thesis. The `classification_report` function in the `sklearn.metrics` library in Python was used to build a *classification report*, which is a text report showing the main classification metrics, on the test data for each classification [98]. Figure 3.4 shows the layout of a binary classification report. In addition to the metrics defined in the previous sections, it includes accuracy (ACC), which is defined as the percentage of correct classifications:

$$ACC = \frac{TP + TN}{P+N} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.22}$$

	Precision	Recall	F1 score	Support
Negative class	Negative predictive value (NPV)	True negative rate (TNR)	Negative F1 score (NF1)	Negative (N)
Positive class	Positive predictive value (PPV)	True positive rate (TPR)	Positive F1 score (PF1)	Positive (P)
Accuracy			Accuracy (ACC)	Total population = P + N
Macro average	Macro average precision (MAP)	Macro average recall (MAR)	Macro average F1 score (MAF1)	Total population = P + N
Weighted average	Weighted average precision (WAP)	Weighted average recall (WAR)	Weighted average F1 score (WAF1)	Total population = P + N

Figure 3.4: The layout of a binary classification report.

The WAF1 score and other evaluation metrics were significantly influenced by choice of `random_state` in the train/test split due to the small sample sizes. Therefore, one hundred train/test splits were conducted with `random_state` ranging from 0 to 99. The 100 resulting classification reports were then averaged. The combination of dataset and classifier that yielded the highest WAF1 score was considered the best.

Interpreting WAF1 scores lacks a standardized approach as it heavily relies on the specific case and dataset. However, comparing the score to the best achievable result without knowledge can provide insight into its quality. In the PD/HC classification scenario, the test set comprised 12 PD and 4 HC cases. Since PD represents the majority class, the best outcome is achieved by classifying all participants as PD without any prior information. Consequently, the WAF1 score without prior knowledge was determined to be 0.64 using Equation 3.21, Equation 3.19, and Equation 3.18. This score is referred to as the neutral WAF1 score for the PD/HC classification.

In the PIGD/TD classification, the test set comprised 6 PIGD and 4 TD cases. Without prior knowledge, the optimal outcome is achieved by classifying all patients as PIGD. Consequently, the WAF1 score without prior knowledge was determined to be 0.45 using Equation 3.21, Equation 3.19, and Equation 3.18. This score is referred to as the neutral WAF1 score for the PIGD/TD classification.

In general, if the WAF1 score for a classifier surpasses the neutral WAF1 score, it implies that the dataset provides additional information that enhances the performance of the classification. Moreover, the more significant the difference between a classifier's WAF1 score and the neutral WAF1 score, the better its performance.

Feature Evaluation

To identify biomarkers for differentiating PD from HC and PIGD from TD, understanding the interaction of features in the classifier's prediction is crucial. This involves determining the importance of features in the classification. One way to do this is using *permutation feature importance* [104]. The permutation importance of a feature is measured by calculating the increase in the model's prediction error after permuting the feature [105]. A feature is important if shuffling its values increases the model error, and unimportant if shuffling its values leaves the model error unchanged. The permutation feature importance technique benefits from being model agnostic [106].

The `permutation_importance` function in the `sklearn.inspection` library in Python was used to compute permutation importances for all features for each classification [98]. The permutation feature importance was computed on test data to highlight which features contribute the most to the generalization power of the inspected model [105]. The `scoring` hyperparameter was set to `f1_weighted` to use WAF1 as score.

The Scikit-learn permutation importance algorithm comprises the following steps [106]:

1. Compute the reference score s of the model m on the data D.
2. For each feature j in D:
   (a) For each repetition k in $1, \ldots, K$:
      i. Randomly shuffle column j of dataset D to generate a corrupted version of the data named $\tilde{D}_{k,j}$.
      ii. Compute the score $s_{k,j}$ of the model m on corrupted data $\tilde{D}_{k,j}$.
   (b) Compute importance i_j for feature j defined as:

$$i_j = s - \frac{1}{K} \sum_{k=1}^{K} s_{k,j} \tag{3.23}$$

Due to the small sample sizes, the permutation importances were significantly influenced by the choice of `random_state` hyperparameter in the train/test split. Therefore, one hundred train/test splits were conducted with `random_state` ranging from 0 to 99. The 100 resulting permutation importances for each feature were then averaged. The features with the highest permutation importance were considered the most important. Additionally, the closer the permutation importance of a feature were to the WAF1 score, the more important it was considered.

Chapter 4

Results

4.1 Generating Images and Maps

Figure 4.1 presents the images acquired from the two ME-GRE sequences described in Table 3.3, specifically for one of the TD patients. The corresponding iron-sensitive MRI images and maps generated for this patient are displayed in Figure 4.2, while the non-iron-sensitive MRI images and maps are shown in Figure 4.3. The quality of the images and maps was deemed adequate for further analysis.

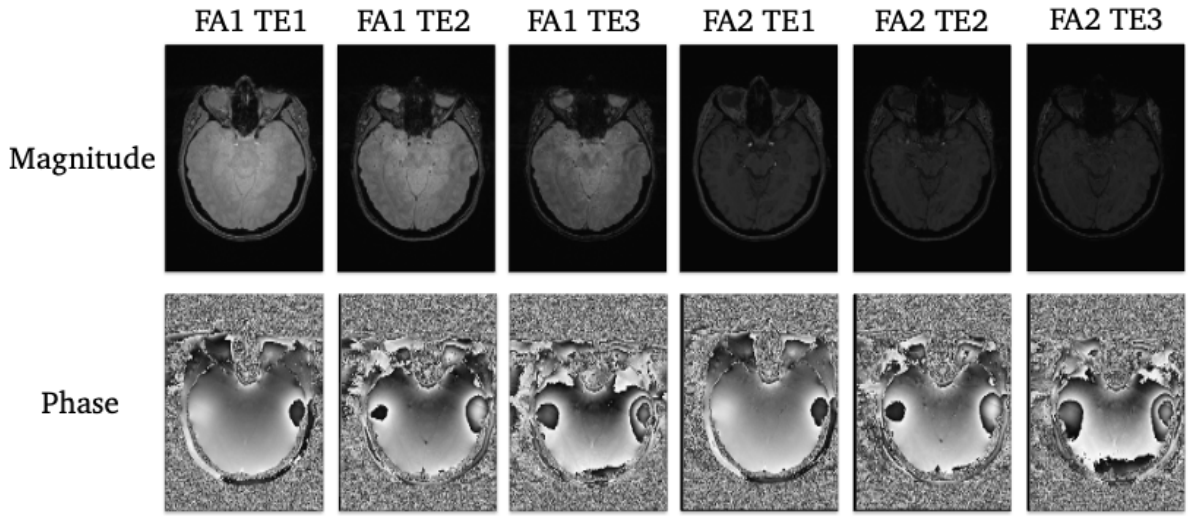

Figure 4.1: Images acquired from the two ME-GRE sequences for one of the TD patients. FA1 = 6°, FA2 = 27°, TE1 = 7.5 ms, TE2 = 15 ms, TE3 = 22.5 ms. FA = flip angle, TE = echo time, ME-GRE = multi-echo gradient echo, TD = tremor dominant.

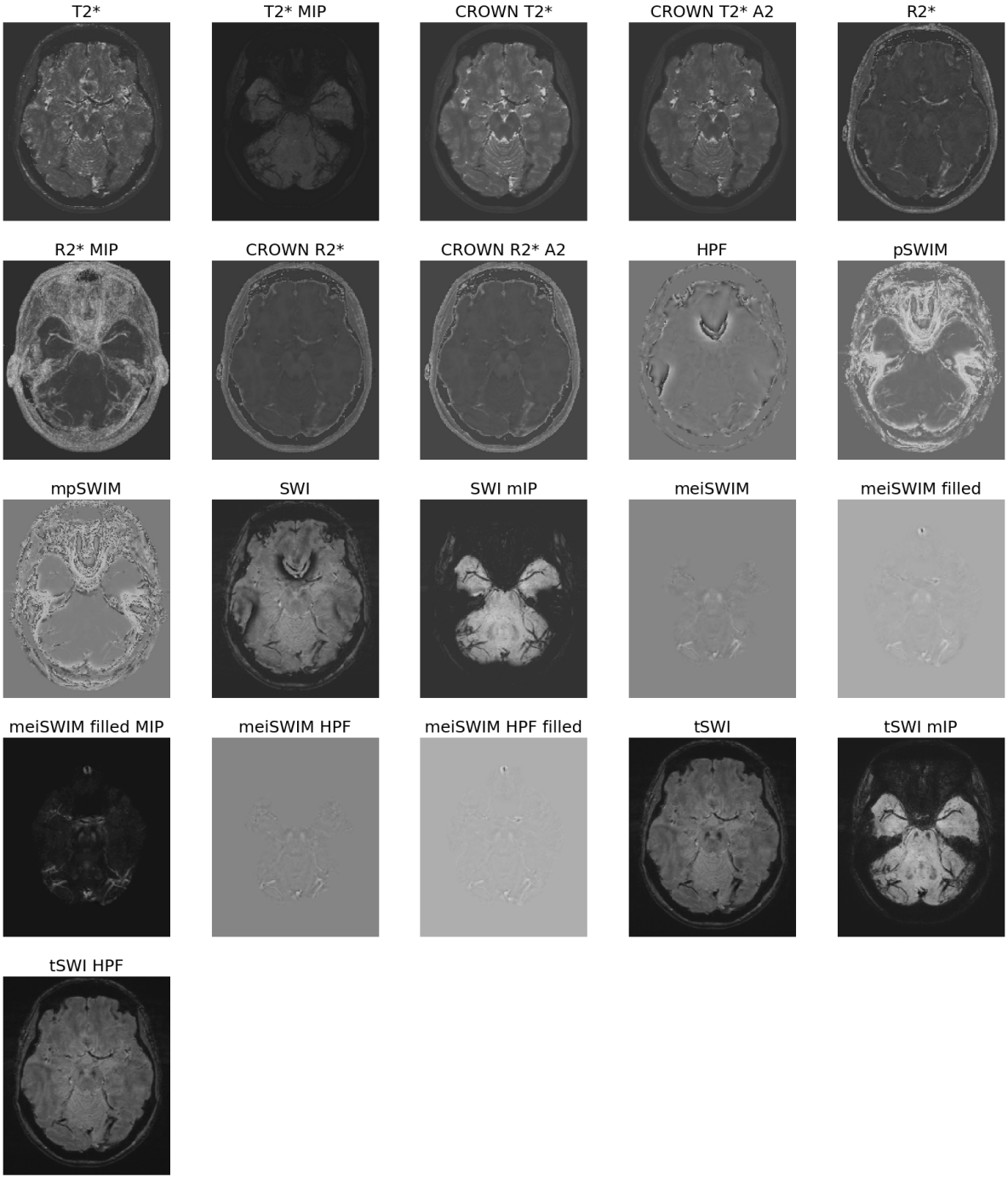

Figure 4.2: Iron-sensitive MRI images and maps for one of the TD patients. T2 = transverse relaxation time, MIP = maximum intensity projection, CROWN = constrained reconstruction of white noise, HPF = high-pass filtered phase, pSWIM = phase susceptibility-weighted imaging mapping, mpSWIM = multi-phase susceptibility-weighted imaging mapping, SWI = susceptibility-weighted image, mIP = mimimum intensity projection, meiSWIM = multi-echo iterative susceptibility-weighted imaging mapping, tSWI = true susceptibility-weighted image, MRI = magnetic resonance imaging, TD = tremor-dominant.*

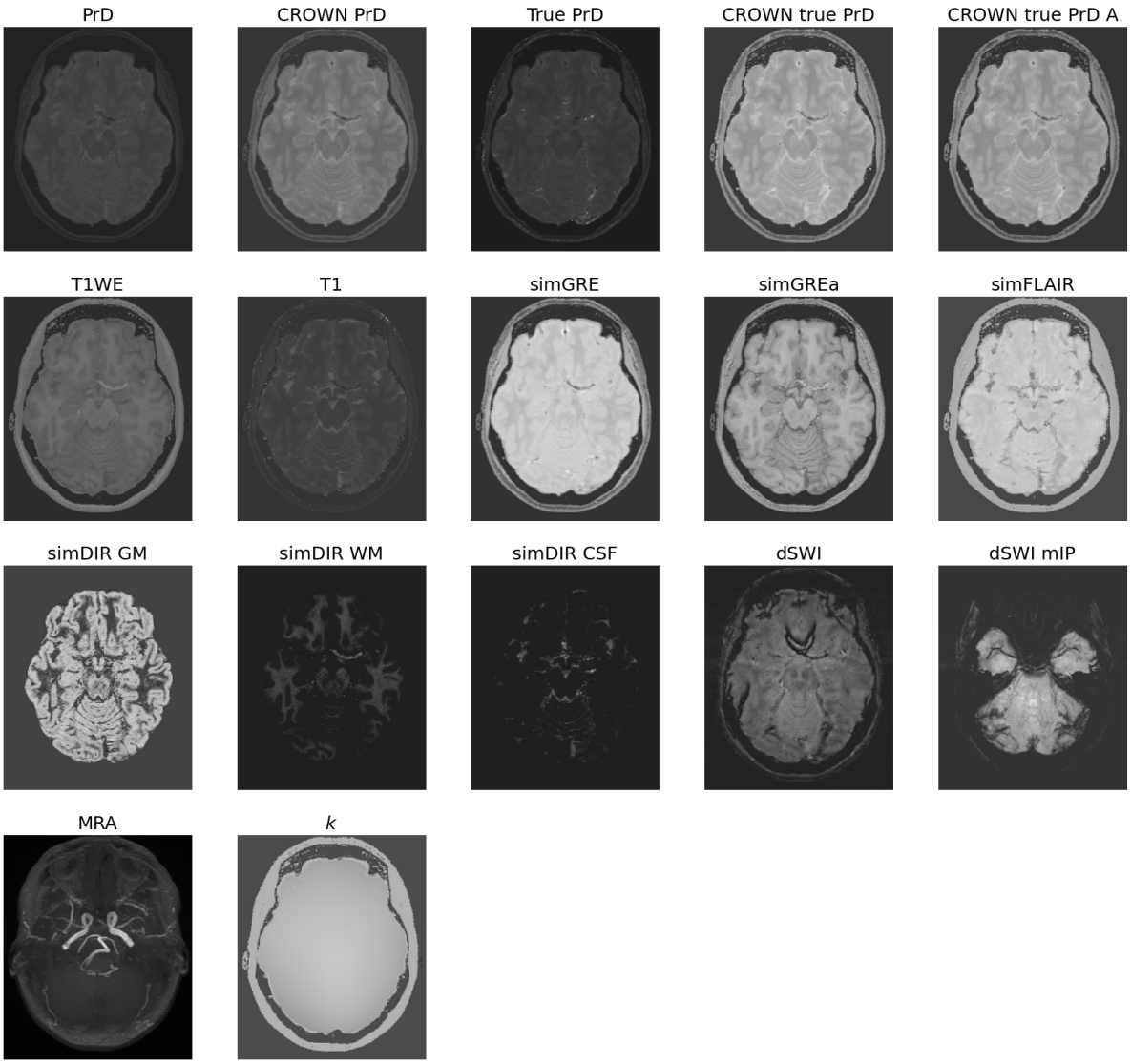

Figure 4.3: Non-iron-sensitive MRI images and maps for one of the TD patients. PrD = proton density, CROWN = constrained reconstruction of white noise, T1WE = T1-weighted enhanced image, simGRE = simulated gradient echo, simFLAIR = simulated fluid-attenuated inversion recovery, simDIR = simulated double inversion recovery, GM = gray matter, WM = white matter, CSF = cerebrospinal fluid, dSWI = diamagnetic susceptibility-weighted imaging, mIP = mimimum intensity projection, MRA = magnetic resonance angiogram, k = radio frequency transmit field variation, MRI = magnetic resonance imaging, TD = tremor-dominant.

4.2 Region of Interest Analysis

Visual inspection showed acceptable co-registration and normalization in the midbrain for all participants. Figure 4.4 shows the SNpc and LC mask outline in all participants' co-registered and normalized T1WE images. The SNpc mask is outlined in red, and the LC mask is outlined in blue. The quality of the normalization was deemed sufficient for further analysis.

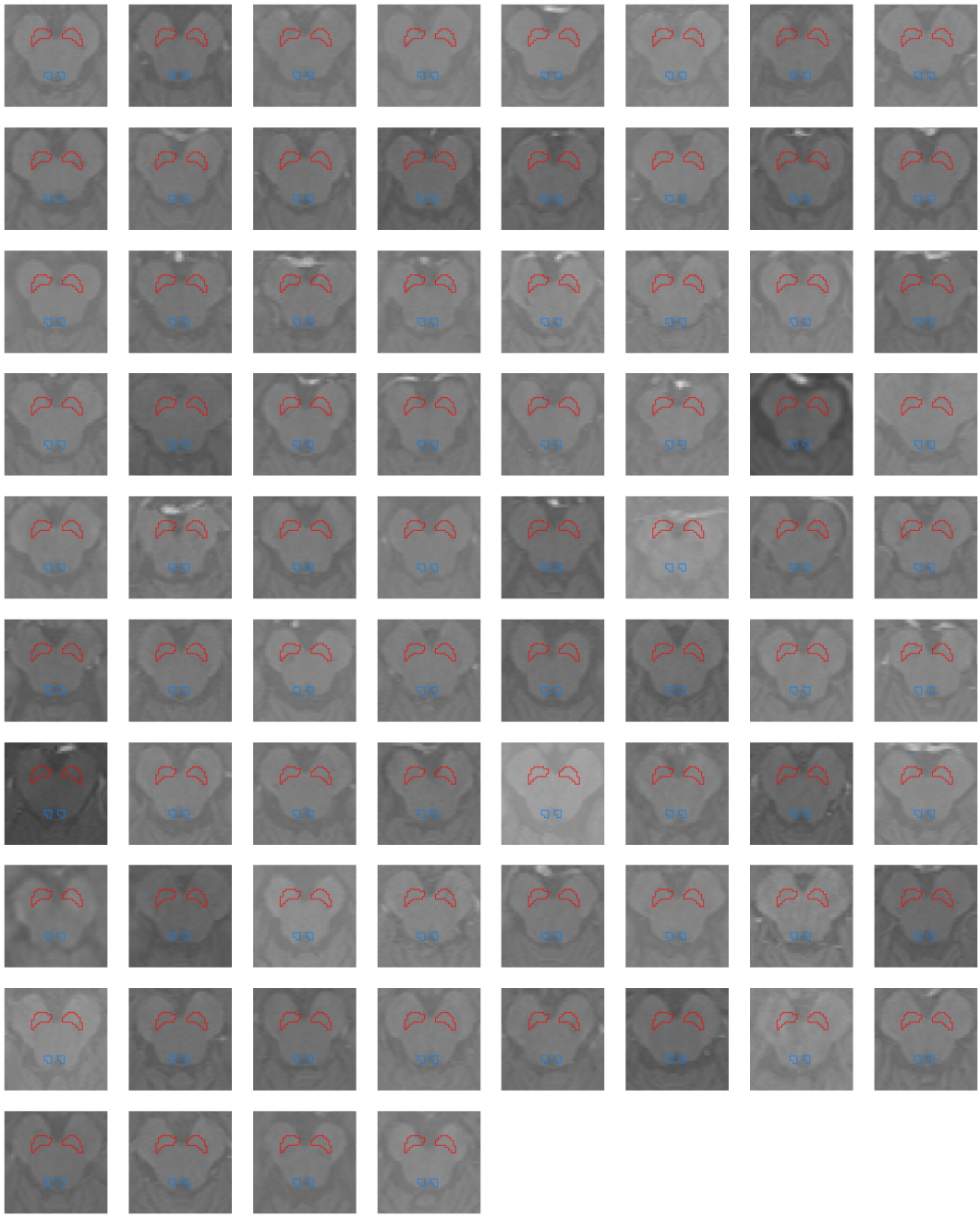

Figure 4.4: Mask outlines of the SNpc and the LC in the normalized T1WE image for all participants. The SNpc is outlined in red and the LC is outlined in blue. SNpc = substantia nigra pars compacta, LC = locus coeruleus, T1WE = T1-weighted enhanced.

4.3 Laterality

MRI images displaying significant SNpc and LC laterality are presented in Table 4.1. The *p*-values were calculated using the Wilcoxon signed-rank test. As participants demonstrated significant laterality across multiple images, it became necessary to consider the left and right components of the SNpc and LC as distinct features in the subsequent analysis.

Table 4.1: MRI features with statistically significant laterality. The p-values were calculated using the Wilcoxon signed-rank test. SNpc = substantia nigra pars compacta, LC = locus coeruleus, CROWN = constrained reconstruction of white noise, PrD = proton density, T2 = transverse relaxation time, HPF = high-pass filtered phase, k = radio frequency transmit field variation, SWI = susceptibility-weighted imaging, T1 = longitudinal relaxation time, T1WE = T1-weighted enhanced, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, GM = gray matter, WM = white matter, simFLAIR = simulated fluid-attenuated inversion recovery, simGRE = simulated gradient echo, tSWI = true susceptibility-weighted imaging, MRA = magnetic resonance angiogram, mIP = mimimum intensity projection, MRI = magnetic resonance imaging.*

Level of significance	Structure	Images/maps
Strongly significant ($p < 0.001$)	SNpc	CROWN PrD, CROWN T2*, CROWN T2* A2, CROWN true PrD, CROWN true PrD A, HPF, *k*, PrD, SWI, T1, T1WE, simDIR CSF, simDIR GM, simDIR WM
	LC	CROWN PrD, CROWN R2*, CROWN T2*, CROWN true PrD, CROWN true PrD A, *k*, PrD, R2*, T1, T2*, simDIR CSF, simFLAIR, simDIR WM, simGREa
Highly significant ($0.001 < p < 0.01$)	SNpc	CROWN R2*, true PrD, tSWI HPF
	LC	CROWN R2* A2, CROWN T2* A2, T1WE, true PrD
Significant ($0.01 < p < 0.05$)	SNpc	CROWN R2* A2, MRA, SWI mIP, T2*, tSWI
	LC	simGRE

4.4 Feature Selection Based on Statistical Testing

Table 4.2 displays the structures and MRI images demonstrating statistically significant differences between PD patients and HCs. The *p*-values were obtained using the Mann-Whitney U test. PD patients exhibited a significantly higher mean intensity in the simGRE image than HCs in all LC regions. Furthermore, the mean intensity in the simDIR CSF image was significantly higher for PD patients than for HCs in the SNpc. Figure 4.5 presents a boxplot illustrating the distribution of simGRE mean intensity across each class for every part of the LC.

Table 4.2: MRI features with statistically significant differences between PD patients and HCs. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface. Q_1 = first quartile, Q_3 = third quartile, PD = Parkinson's disease, HC = healthy control, LC = locus coeruleus, SNpc = substantia nigra pars compacta, simGRE = simulated gradient echo, simDIR CSF = simulated double inversion recovery cerebrospinal fluid, MRI = magnetic resonance imaging.

Image/map	**Structure**		**Median ($Q_1 - Q_3$)**		**p-value**
			PD	**HC**	
simGRE	LC	Total	**556 (520 - 565)**	531 (491 - 550)	0.02*
		Left	**554 (512 - 567)**	527 (474 - 550)	0.03*
		Right	**555 (538 - 566)**	541 (516 - 548)	0.008**
simDIR CSF	SNpc	Total	**16 (3 - 52)**	2 (0 - 13)	0.02*

** Significant difference between PD and HC ($p < 0.05$).*

*** Highly significant difference between PD and HC ($p < 0.01$).*

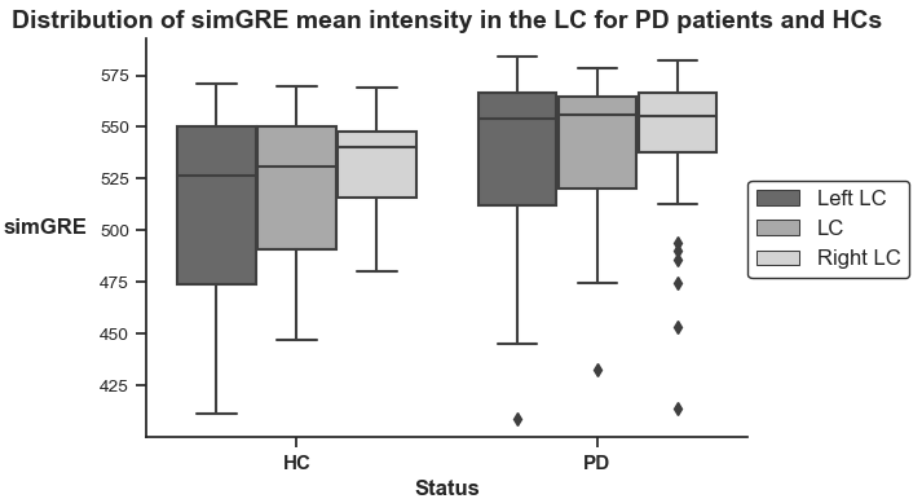

Figure 4.5: Distribution of mean intensity in the left LC ($p = 0.03$), the right LC ($p = 0.008$), and the LC ($p = 0.02$) in the simGRE image for PD patients and HCs. simGRE = simulated gradient echo, LC = locus coeruleus, PD = Parkinson's disease, HC = healthy control.

Appendix A includes the *p*-values associated with the discrimination of PD from HCs across various structures and MRI images. Non-significant differences between PD and HC can be referred to as trends. There was a trend indicating that PD patients had a higher mean intensity in the SNpc compared to HCs in the R2* maps and the meiSWIM maps. Conversely, there was a trend indicating that HCs had a higher mean intensity in the SNpc compared to PD patients in the T2* maps, SWI images, the tSWI images, and the PrD maps. Moreover, there was a trend indicating that PD patients had a higher mean intensity in the LC compared to HCs in the meiSWIM maps.

Table 4.3 displays the structures and MRI images demonstrating statistically significant differences between patients with PIGD and TD motor phenotypes. The *p*-values were calculated using the Student's unpaired *t*-test for normally distributed populations and the Mann-Whitney U test for non-normally distributed populations.

Table 4.3: MRI features with statistically significant differences between PIGD patients and TD patients. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface. Q_1 = first quartile, Q_3 = third quartile, SD = standard deviation, PIGD = postural instability gait difficulty, TD = tremor-dominant, SNpc = substantia nigra pars compacta, LC = locus coeruleus, simDIR = simulated double inversion recovery, GM = gray matter, CROWN = constrained reconstruction of white noise, PrD = proton density, simGRE = simulated gradient echo, WM = white matter, CSF = cerebrospinal fluid, k = radio frequency transmit field variation, MRI = magnetic resonance imaging.

Image/map	**Structure**		**Median ($Q_1 - Q_3$) / Mean±SD**		**p-value**
			PIGD	**TD**	
simDIR GM	SNpc	Total	66±21	**89±30**	0.004**
		Left	74±26	**95±33**	0.02*
		Right	58±25	**83±30**	0.003**
	LC	Total	105±29	**127±25**	0.008**
		Left	105±31	**126±26**	0.02*
		Right	105±29	**128±25**	0.006**
CROWN PrD	SNpc	Right	635±29	**654±32**	0.04*
CROWN true PrD A	SNpc	Right	760±30	**781±35**	0.03*
simGREa	SNpc	Total	553±22	**566±17**	0.03*
		Left	550±22	**563±17**	0.03*
simDIR WM	SNpc	Right	**274±65**	231±69	0.04*
simDIR CSF	SNpc	Total	**34 (5 - 85)**	8 (2 - 31)	0.03*
k	SNpc	Total	**1273±50**	1240±46	0.03*
		Left	**1270±50**	1238±46	0.03*
		Right	**1275±51**	1242±47	0.03*
	LC	Total	**1228±52**	1193±52	0.03*
		Left	**1234±52**	1199±51	0.03*
		Right	**1222±52**	1188±52	0.03*

** Significant difference between PIGD and TD ($p < 0.05$).*

*** Highly significant difference between PIGD and TD ($p < 0.01$).*

A significantly higher mean intensity was observed for TD patients compared to PIGD patients in all regions of the SNpc and the LC in the simDIR GM image. Furthermore, TD patients exhibited significantly higher mean intensity than PIGD patients in the following regions: the SNpc in the CROWN PrD map, the SNpc in the CROWN true PrD A map, and the left SNpc and the SNpc in the simGREa image.

Conversely, PIGD patients demonstrated significantly higher mean intensity than TD patients in the right SNpc in the simDIR WM image, the SNpc in the simDIR CSF image, and all SNpc and LC regions in the *k* map.

Figure 4.6 presents a boxplot showcasing the distribution of simDIR GM mean intensity for each part of the SNpc across each class. Likewise, Figure 4.7 displays a boxplot illustrating the distribution of simDIR GM mean intensity for each part of the LC across each class.

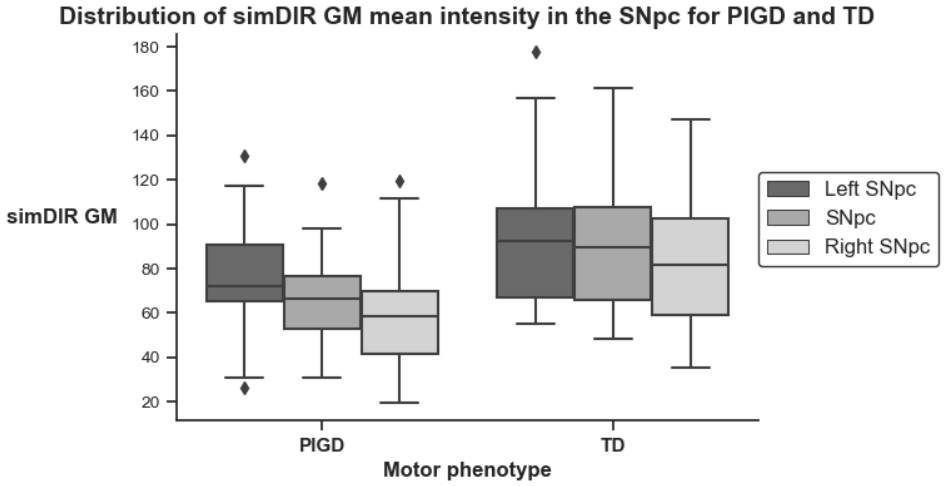

Figure 4.6: Distribution of mean intensity in the left SNpc ($p = 0.02$), the right SNpc ($p = 0.003$), and the SNpc ($p = 0.004$) in the simDIR GM image for PIGD patients and TD patients. simDIR GM = simulated double inversion recovery gray matter, SNpc = substantia nigra pars compacta, PIGD = postural instability gait difficulty, TD = tremor-dominant.

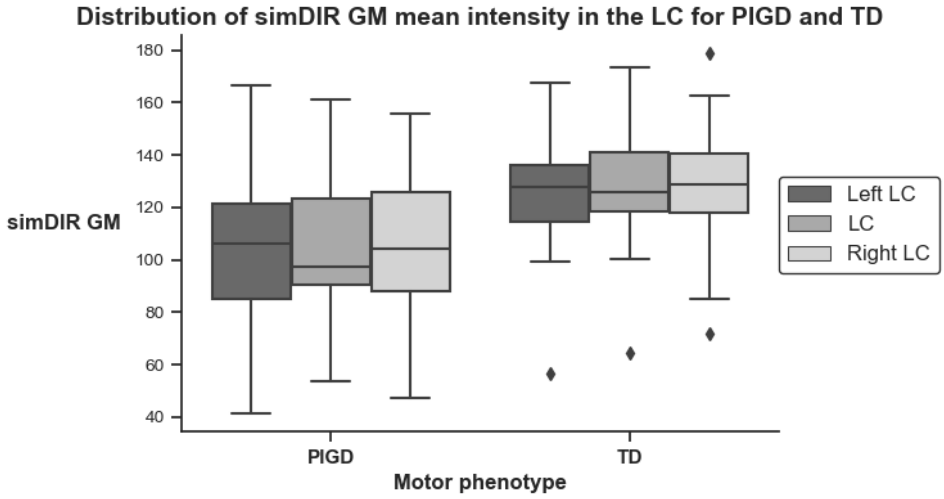

Figure 4.7: Distribution of mean intensity of the left LC ($p = 0.02$), the right LC ($p = 0.006$), and the LC ($p = 0.008$) in the simDIR GM image for PIGD patients and TD patients. simDIR GM = simulated double inversion recovery gray matter, LC = locus coeruleus, PIGD = postural instability gait difficulty, TD = tremor-dominant.

Appendix B includes the p-values related to the differentiation of PIGD from TD across all structures and MRI images. A trend indicated that PIGD had a higher mean intensity in the SNpc compared to TD in the R2* maps. Conversely, there was a trend indicating that TD had a higher mean intensity in the SNpc compared to PIGD in the PrD maps. Furthermore, there was a trend indicating that PIGD had a higher mean intensity in the LC compared to TD in the T2* maps and the meiSWIM maps.

4.5 Feature Selection Based on Pairwise Correlation

Table 4.4 presents the structures and MRI images with pairwise absolute Spearman's rank correlation coefficients below 0.7, while Table 4.5 showcases the structures and MRI images with pairwise absolute Spearman's rank correlation coefficients between 0.7 and 0.9. These nine sets of MRI features were further used to classify patients based on PIGD/TD motor phenotype.

Table 4.4: MRI features with pairwise absolute Spearman's rank correlation coefficients $|\rho|$ = below 0.7. SNpc = substantia nigra pars compacta, LC = locus coeruleus, CROWN = constrained reconstruction of white noise, PrD = proton density, HPF = high-pass filtered phase, k = radio frequency transmit field variation, MRA = magnetic resonance angiogram, T2 = transverse relaxation time, T1WE = T1-weighted enhanced, mpSWIM = multi-phase susceptibility-weighted imaging mapping, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, MIP = maximum intensity projection, simGRE = simulated gradient echo, WM = white matter, meiSWIM = multi-echo iterative susceptibility-weighted imaging mapping, simFLAIR = simulated fluid-attenuated inversion recovery, SWI = susceptibility-weighted imaging, GM = gray matter, dSWI = diamagnetic susceptibility-weighted imaging, mIP = mimumum intensity projection, tSWI = true susceptibility-weighted imaging, MRI = magnetic resonance imaging.*

Coefficient	Structure	Images/maps
$\|\rho\| < 0.1$	SNpc	CROWN PrD, CROWN R2*
$0.1 \leq \|\rho\| < 0.2$	SNpc	HPF
$0.2 \leq \|\rho\| < 0.3$	SNpc	*k*, MRA
	LC	HPF
$0.3 \leq \|\rho\| < 0.4$	SNpc	CROWN T2*, T1WE, mpSWIM, simDIR CSF
	LC	CROWN PrD, CROWN R2*, R2* MIP, simGRE
$0.4 \leq \|\rho\| < 0.5$	SNpc	simDIR WM, simGREa
	Right SNpc	simDIR CSF
	LC	meiSWIM, simFLAIR
$0.5 \leq \|\rho\| < 0.6$	SNpc	R2* MIP, SWI, simGRE
	LC	MRA, SWI, mpSWIM, simDIR GM, simDIR WM, simGREa
$0.6 \leq \|\rho\| < 0.7$	SNpc	T2* MIP, T2*, True PrD, dSWI mIP, meiSWIM, simFLAIR, tSWI
	LC	R2*, SWI mIP, T2* MIP, meiSWIM HPF

Table 4.5: MRI features with pairwise absolute Spearman's rank correlation coefficients $|\rho|$ = between 0.7 and 0.9. SNpc = substantia nigra pars compacta, LC = locus coeruleus, PrD = proton density, SWI = susceptibility-weighted imaging, mIP = mimumum intensity projection, T1 = longitudinal relaxation time, dSWI = diamagnetic susceptibility-weighted imaging, meiSWIM = multi-echo iterative susceptibility-weighted imaging mapping, MIP = maximum intensity projection, pSWIM = phase susceptibility-weighted imaging mapping, tSWI = true susceptibility-weighted imaging, MRA = magnetic resonance angiogram, CROWN = constrained reconstruction of white noise, HPF = high-pass filtered phase, simDIR = simulated double inversion recovery, GM = gray matter, T2 = transverse relaxation time, CSF = cerebrospinal fluid, WM = white matter, simGRE = simulated gradient echo, mpSWIM = multi-phase susceptibility-weighted imaging mapping, simFLAIR = simulated fluid-attenuated inversion recovery, MRI = magnetic resonance imaging.*

Coefficient	Structure	Images/maps
$0.7 \leq \|\rho\| < 0.8$	SNpc	PrD, R2*, SWI mIP, T1, dSWI, meiSWIM filled MIP, pSWIM, tSWI mIP
	Left SNpc	MRA, T1
	Right SNpc	MRA
	LC	meiSWIM filled MIP
	Right LC	CROWN PrD, meiSWIM HPF
$0.8 \leq \|\rho\| < 0.9$	SNpc	CROWN true PrD, meiSWIM HPF, simDIR GM, tSWI HPF
	Left SNpc	CROWN PrD, CROWN R2*, CROWN T2*, CROWN true PrD, HPF, R2*, SWI, SWI mIP, T2*, T2* MIP, True PrD, dSWI, simDIR CSF, simDIR WM, simGRE, simGREa, tSWI, tSWI HPF
	Right SNpc	CROWN T2*, HPF, R2* MIP, R2*, T1, T2* MIP, T2*, True PrD, meiSWIM HPF, mpSWIM, pSWIM, simGREa
	LC	T2*, dSWI mIP, mpSWIM, tSWI, tSWI mIP
	Left LC	CROWN PrD, CROWN R2*, HPF, R2*, T2*, T2* MIP, meiSWIM HPF, simFLAIR, simDIR WM
	Right LC	CROWN R2*, HPF, R2*, T2*, mpSWIM, simDIR CSF, simFLAIR, simDIR WM

4.6 Differentiating PD from HC

4.6.1 Classification without Clinical Features

Table 4.6 presents the WAF1 scores for the PD/HC classification of participants without clinical features. Four achieved an WAF1 score of 0.70 among the various combinations of datasets and classifiers. These combinations included: 1) RaF classification using significant features from Table 4.2, 2) SVC using features with $|\rho| < 0.9$ from Table 4.4 and Table 4.5, 3) SVC using features with $|\rho| < 0.6$ from Table 4.4, and 4) SVC using features with $|\rho| < 0.4$ from Table 4.4. These four combinations were therefore explored further.

Table 4.6: WAF1 scores for the PD/HC classification of participants without clinical features. The WAF1 score is an average of the WAF1 score of 100 train/test splits with random state 0-99. The highest scores are highlighted in bold. The test group included 12 PD patients and 4 HCs. MRI = magnetic resonance imaging, WAF1 = weighted average F1, LR = logistic regression, DT = decision tree, RaF = random forest, SVC = support vector classification, XGBoost = extreme gradient boosting, ρ = Spearman's rank correlation coefficient.

MRI features	WAF1 score				
	LR	DT	RaF	SVC	XGBoost
All	0.57	0.62	0.66	0.66	0.67
Significant [a]	0.64	0.64	**0.70**	0.64	0.68
Highly significant [a]	0.67	0.67	0.67	0.62	0.68
$\lvert\rho\rvert < 0.9$ [b]	0.58	0.61	0.66	**0.70**	0.66
$\lvert\rho\rvert < 0.8$ [b]	0.56	0.63	0.66	0.68	0.67
$\lvert\rho\rvert < 0.7$ [b]	0.55	0.62	0.66	0.68	0.63
$\lvert\rho\rvert < 0.6$ [b]	0.57	0.62	0.66	**0.70**	0.65
$\lvert\rho\rvert < 0.5$ [b]	0.59	0.64	0.67	0.68	0.65
$\lvert\rho\rvert < 0.4$ [b]	0.57	0.63	0.67	**0.70**	0.65
$\lvert\rho\rvert < 0.3$ [b]	0.52	0.61	0.65	0.62	0.62
$\lvert\rho\rvert < 0.2$ [b]	0.52	0.63	0.68	0.68	0.66
$\lvert\rho\rvert < 0.1$ [b]	0.52	0.62	0.65	0.68	0.64

[a] *Features with ($p < 0.05$) in Table 4.2.*

[b] *Features with ($p < 0.01$) in Table 4.2.*

Classification with Significant MRI Features

The RaF classifier obtained the highest WAF1 score of 0.70 for the PD/HC classification using only significant MRI features ($p < 0.05$), as depicted in Table 4.6. Figure 4.8 and Table 4.7 show the accompanied confusion matrix and classification report. Most patients were correctly classified as patients. However, the majority of HCs were also classified as patients. The precision, recall, and F1 score were notably higher for the patient group than for the control group. The PD group exhibited higher recall than precision, whereas the HC group demonstrated higher precision than recall.

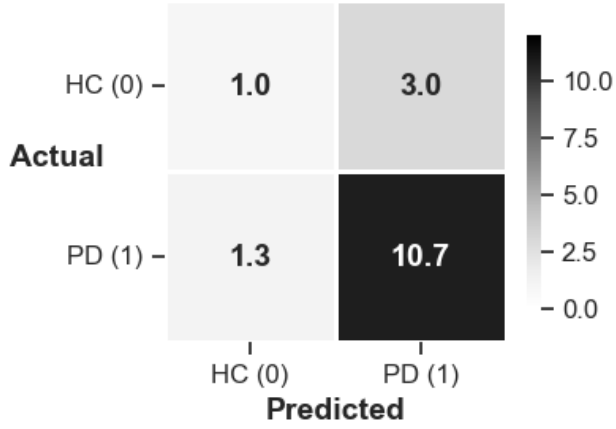

Figure 4.8: Confusion matrix for the RaF classification of PD/HC status with significant MRI features ($p < 0.05$). An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, RaF = random forest, MRI = magnetic resonance imaging.

Table 4.7: Classification report for the RaF classification of PD/HC status with significant MRI features ($p < 0.05$). An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, RaF = random forest, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
HC	0.43	0.24	0.29	4
PD	0.78	0.90	0.83	12
ACC			0.73	16
MA	0.61	0.57	0.56	16
WA	0.69	0.73	0.70	16

The permutation importances presented in Figure 4.9 reveal that the right LC in the simGRE image was an essential feature for the RaF classification of PD/HC status with significant MRI features ($p < 0.05$), followed by the SNpc in the simDIR CSF image and the LC in the simGRE image.

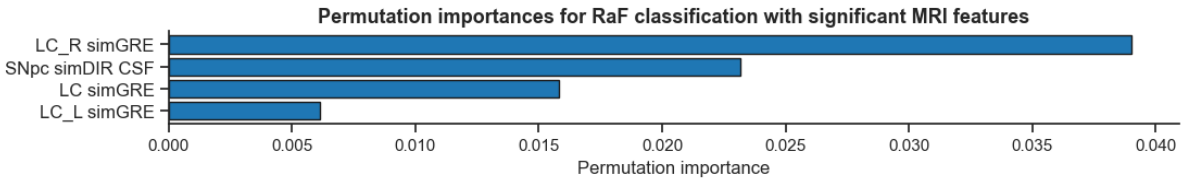

Figure 4.9: Permutation importances for the RaF classification of PD/HC status with significant MRI features ($p < 0.05$). An average of 100 train/test splits with random state 0-99. RaF = random forest, MRI = magnetic resonance imaging, LC_R = right locus coeruleus, simGRE = simulated gradient echo, SNpc = substantia nigra pars compacta, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, LC = locus coeruleus, LC_L = left locus coeruleus, PD = Parkinson's disease, HC = healthy control.

Classification with MRI Features with $|\rho| < 0.9$

The SVC attained the highest WAF1 score of 0.70 among all classifiers for the PD/HC classification using only MRI features with $|\rho| < 0.9$, as shown in Table 4.6. Figure 4.10 and Table 4.8 show the accompanied confusion matrix and classification report.

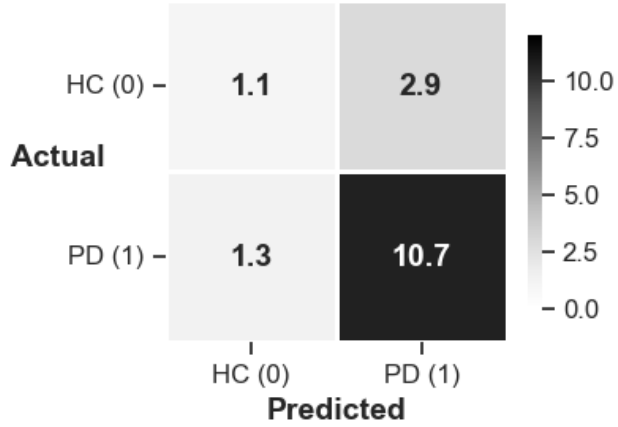

Figure 4.10: Confusion matrix for the SVC of PD/HC status with MRI features with $|\rho| < 0.9$. An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, SVC = support vector classification, MRI = magnetic resonance imaging.

Table 4.8: Classification report for the SVC of PD/HC status with MRI features with $|\rho| < 0.9$. An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, SVC = support vector classification, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
HC	0.41	0.26	0.30	4
PD	0.76	0.89	0.83	12
ACC			0.73	16
MA	0.60	0.58	0.57	16
WA	0.69	0.73	0.70	16

Most patients were correctly classified as patients, but many HCs were also classified as patients. However, the number of correctly classified HCs was slightly higher compared to the classification with significant MRI features. Despite this, the precision, recall, and F1 score were still higher for the PD group than for the HC group.

The permutation importances shown in Figure 4.11 indicate that the most crucial feature for SVC classification of PD/HC status with MRI features having $|\rho| < 0.9$ was the LC in the simGRE image, followed by the left LC in the simFLAIR image and the right SNpc in the simDIR CSF image.

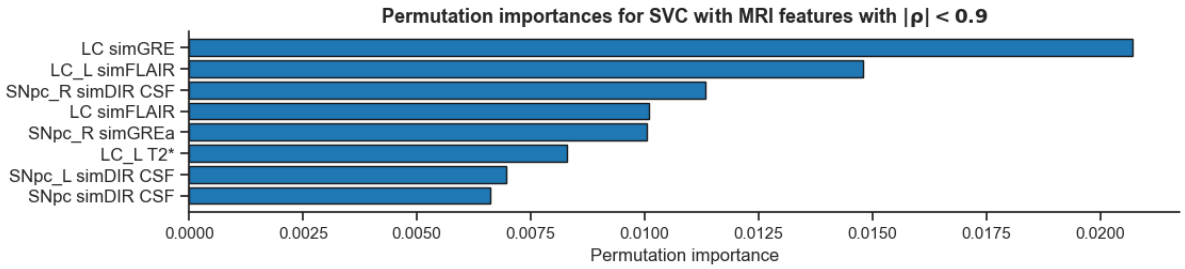

Figure 4.11: The eight highest permutation importances for the SVC of PD/HC status with MRI features with $|\rho| < 0.9$. An average of 100 train/test splits with random state 0-99. SVC = support vector classification, MRI = magnetic resonance imaging, ρ = Spearman's rank correlation coefficient, LC = locus coeruleus, simGRE = simulated gradient echo, LC_L = left locus coeruleus, simFLAIR = simulated fluid-attenuated inversion recovery, SNpc_R = right substantia nigra pars compacta, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, T2 = transverse relaxation time, SNpc_L = left substantia nigra pars compacta, SNpc = substantia nigra pars compacta, PD = Parkinson's disease, HC = healthy control.*

Classification with MRI Features with $|\rho| < 0.6$

SVC achieved the highest WAF1 score of 0.70 for the classification of PD/HC with MRI features having $|\rho| < 0.6$, as presented in Table 4.6. The corresponding confusion matrix is shown in Figure 4.12, and the classification report is provided in Table 4.9. Compared to the classifications with significant MRI features and with MRI features having $|\rho| < 0.9$, this classification correctly predicted more HCs but fewer PD patients.

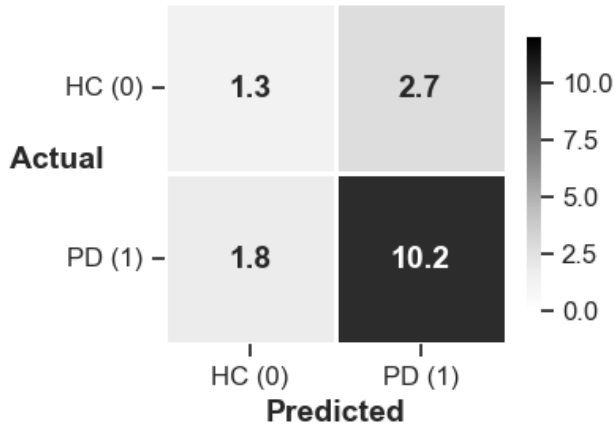

Figure 4.12: Confusion matrix for the SVC of PD/HC status with MRI features with $|\rho| < 0.6$. An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, SVC = support vector classification, MRI = magnetic resonance imaging.

Table 4.9: Classification report for the SVC of PD/HC status with MRI features with $|\rho| < 0.6$. An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, SVC = support vector classification, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
HC	0.43	0.32	0.34	4
PD	0.79	0.85	0.82	12
ACC			0.72	16
MA	0.61	0.58	0.58	16
WA	0.70	0.72	0.70	16

The permutation importances for the SVC of PD/HC status with MRI features with $|\rho| < 0.6$ are presented in Figure 4.13.

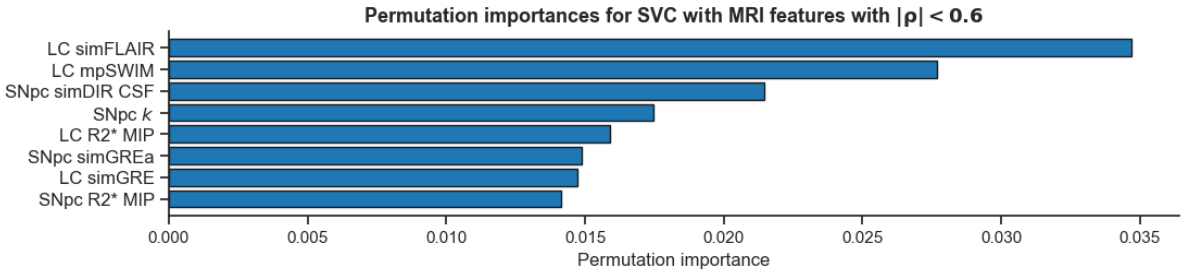

Figure 4.13: The eight highest permutation importances for the SVC of PD/HC status with MRI features with $|\rho| < 0.6$. An average of 100 train/test splits with random state 0-99. SVC = support vector classification, MRI = magnetic resonance imaging, ρ = Spearman's rank correlation coefficient, LC = locus coeruleus, simFLAIR = simulated fluid-attenuated inversion recovery, mpSWIM = multi-phase susceptibility-weighted imaging mapping, SNpc = substantia nigra pars compacta, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, k = radio frequency transmit field variation, MIP = maximum intensity projection, simGRE = simulated gradient echo, PD = Parkinson's disease, HC = healthy control.

The LC in the simFLAIR image was an essential feature for the SVC of PD/HC status with MRI features with $|\rho| < 0.6$, followed by the LC in the mpSWIM map and the SNpc in the simDIR CSF image.

Classification with MRI Features with $|\rho| < 0.4$

Among all classifiers, SVC achieved the highest WAF1 score of 0.70 for the classification of PD/HC using MRI features with $|\rho| < 0.4$, as presented in Table 4.6. The corresponding confusion matrix is shown in Figure 4.14, and the classification report is provided in Table 4.10. Compared to the other classifications without clinical features that also yielded an WAF1 score of 0.70, this classification correctly predicted more HCs but fewer PD patients. The precision and recall were nearly equal for both patients and HCs.

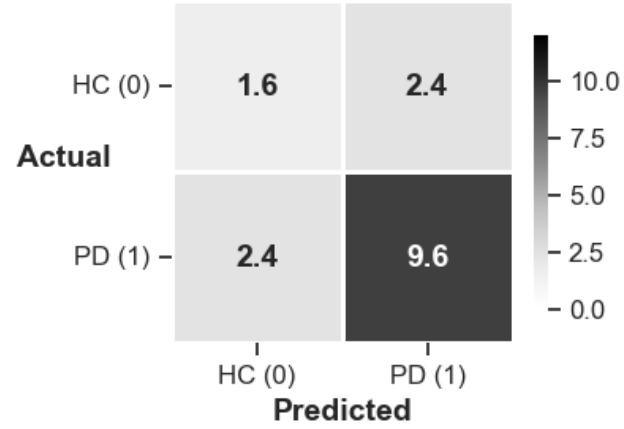

Figure 4.14: Confusion matrix for the SVC of PD/HC status with MRI features with $|\rho| < 0.4$. An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, SVC = support vector classification, MRI = magnetic resonance imaging.

Table 4.10: Classification report for the SVC of PD/HC status with MRI features with $|\rho| < 0.4$. An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, SVC = support vector classification, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
HC	0.40	0.41	0.39	4
PD	0.81	0.80	0.80	12
ACC			0.70	16
MA	0.60	0.61	0.59	16
WA	0.71	0.70	0.70	16

The permutation importances shown in Figure 4.15 indicate that the LC in the simGRE image was the most crucial feature for the SVC classification of PD/HC status using MRI features with $|\rho| < 0.4$, followed by the SNpc in the simDIR CSF image and the LC in the CROWN R2* map.

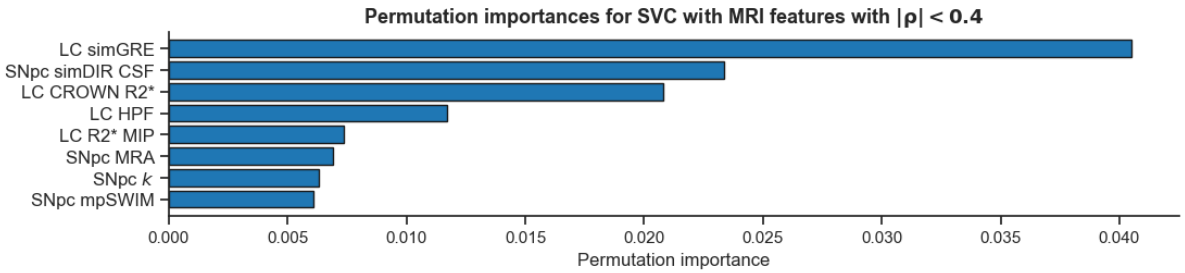

Figure 4.15: The eight highest permutation importances for the SVC of PD/HC status with MRI features with $|\rho| < 0.4$. An average of 100 train/test splits with random state 0-99. SVC = support vector classification, MRI = magnetic resonance imaging, ρ = Spearman's rank correlation coefficient, LC = locus coeruleus, simGRE = simulated gradient echo, SNpc = substantia nigra pars compacta, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, CROWN = constrained reconstruction of white noise, HPF = high-pass filtered phase, MIP = maximum intensity projection, MRA = magnetic resonance angiogram, k = radio frequency transmit field variation, mpSWIM = multi-phase susceptibility-weighted imaging mapping, PD = Parkinson's disease, HC = healthy control.

4.6.2 Classification with Clinical Features

The WAF1 scores for classifying PD/HC participants using clinical features are presented in Table 4.11. The clinical features considered include gender, age, and the MoCA and B-SIT scores. When combined with the clinical features and utilizing XGBoost, two MRI datasets achieved an WAF1 score of 0.84. These datasets were: 1) Significant features from Table 4.2, and 2) Highly significant features from Table 4.2. These two combinations were therefore explored further.

Classification with Significant MRI Features and Clinical Features

The PD/HC status classification using significant MRI features and clinical features with XGBoost yielded a WAF1 score of 0.84, as displayed in Table 4.11. The F1 score was higher for the PD group than the HC group, although the difference was smaller than in classifications without clinical data. The corresponding confusion matrix can be seen in Figure 4.16, and the classification report is provided in Table 4.12. The majority of both PD patients and HCs were accurately classified.

Table 4.11: WAF1 scores for the PD/HC classification of participants with clinical features. The WAF1 score is an average of the WAF1 score of 100 train/test splits with random state 0-99. The highest scores are highlighted in bold. The test group included 12 PD patients and 4 HCs. WAF1 = weighted average F1, MRI = magnetic resonance imaging, XGBoost = extreme gradient boosting, ρ = Spearman's rank correlation coefficient.

MRI features	**WAF1 score** XGBoost
All	0.80
Significant [a]	**0.84**
Highly significant [a]	**0.84**
$\|\rho\| < 0.9$ [b]	0.80
$\|\rho\| < 0.8$ [b]	0.81
$\|\rho\| < 0.7$ [c]	0.81
$\|\rho\| < 0.6$ [c]	0.80
$\|\rho\| < 0.5$ [c]	0.82
$\|\rho\| < 0.4$ [c]	0.82
$\|\rho\| < 0.3$ [c]	0.79
$\|\rho\| < 0.2$ [c]	0.80
$\|\rho\| < 0.1$ [c]	0.79

[a] *See Table 4.2.*

[b] *See Table 4.4 and Table 4.5.*

[c] *See Table 4.4.*

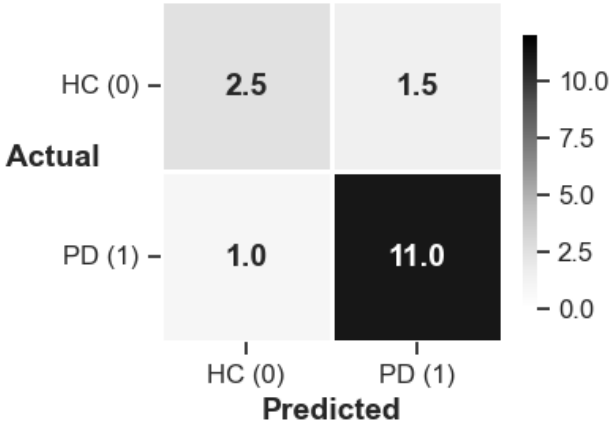

Figure 4.16: Confusion matrix for the XGBoost classification of PD/HC status with significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging.

Table 4.12: Classification report for the XGBoost classification of PD/HC status with significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
HC	0.77	0.62	0.66	4
PD	0.88	0.92	0.90	12
ACC			0.84	16
MA	0.83	0.77	0.78	16
WA	0.86	0.84	0.84	16

The permutation importances presented in Figure 4.17 highlight that the B-SIT score was the most noteworthy feature for the XGBoost classification of PD/HC status, utilizing significant MRI features and clinical features. Following the B-SIT score, the right LC in the simGRE image and the left LC in the simGRE image exhibited significant importance in the classification.

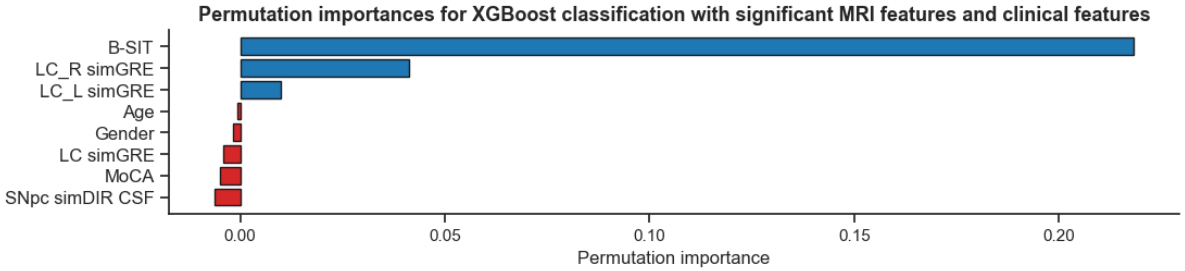

Figure 4.17: Permutation importances for the XGBoost classification of PD/HC status with significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging, B-SIT = Brief Smell Identification Test, LC_R = right locus coeruleus, simGRE = simulated gradient echo, LC_L = left locus coeruleus, LC = locus coeruleus, MoCA = Montreal Cognitive Assessment, SNpc = substantia nigra pars compacta, simDIR = simulated double inversion recovery, CSF = cerebrospinal fluid, PD = Parkinson's disease, HC = healthy control.

Classification with Highly Significant MRI Features and Clinical Features

The XGBoost classification of PD/HC status, incorporating highly significant MRI features and clinical features, achieved a WAF1 score of 0.84, as indicated in Table 4.11. The accompanying confusion matrix is displayed in Figure 4.18, and the detailed classification report is provided in Table 4.13. The results closely resemble those obtained from the classification with significant MRI and clinical features, showing similar weighted average precision and F1 score.

The permutation importances presented in Figure 4.19 highlight that the B-SIT score was an essential feature for the XGBoost classification of PD/HC status, utilizing highly significant MRI features and clinical features, followed by the right LC in the simGRE image.

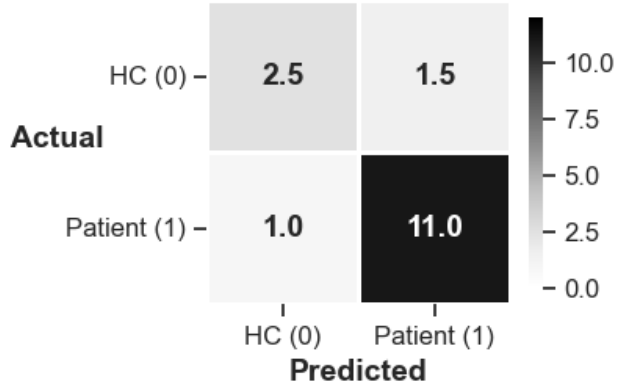

Figure 4.18: Confusion matrix for the XGBoost classification of PD/HC status with highly significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. HC = healthy control, PD = Parkinson's disease, XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging.

Table 4.13: Classification report for the XGBoost classification of PD/HC status with highly significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, XGBoost = extreme gradient boosting.

	PREC	REC	F1	SUP
HC	0.78	0.63	0.66	4
PD	0.89	0.92	0.90	12
ACC			0.85	16
MA	0.83	0.78	0.78	16
WA	0.86	0.85	0.84	16

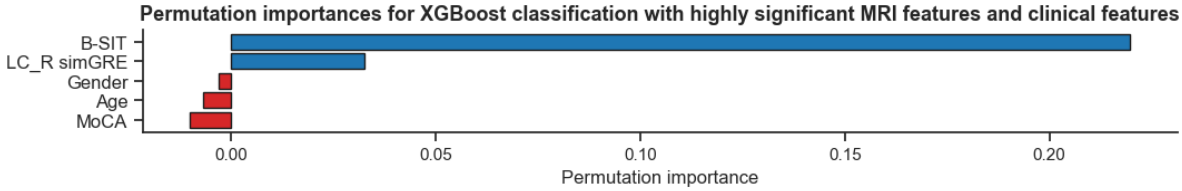

Figure 4.19: Permutation importances for the XGBoost classification of PD/HC status with highly significant MRI features and clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, MRI = magnetic resoance imaging, B-SIT = Brief Smell Identification Test, LC_R = right locus coeruleus, simGRE = simulated gradient echo, MoCA = Montreal Cognitive Assessment, PD = Parkinson's disease, HC = healthy control.

4.7 Differentiating PIGD from TD

4.7.1 Classification without Clinical Features

The WAF1 scores for the PIGD/TD classification of patients without clinical features are provided in Table 4.14. Among the different combinations of datasets and classifiers, the XGBoost classification with significant MRI features achieved the highest WAF1 score of 0.63. This combination was therefore explored further.

Table 4.14: WAF1 scores for the PIGD/TD classification of patients without clinical features. The WAF1 score is an average of the WAF1 score of 100 train/test splits with random state 0-99. The highest score is highlighted in bold. The test group included 6 PIGD and 4 TD. WAF1 = weighted average F1, MRI = magnetic resonance imaging, LR = logistic regression, DT = decision tree, RaF = random forest, SVC = support vector classification, XGBoost = extreme gradient boosting, ρ = Spearman's rank correlation coefficient.

MRI features	**WAF1 score**				
	LR	DT	RaF	SVC	XGBoost
All	0.52	0.52	0.50	0.51	0.53
Significant [a]	0.61	0.62	0.61	0.56	**0.63**
Highly significant [a]	0.62	0.59	0.57	0.59	0.59
$\|\rho\| < 0.9$ [b]	0.56	0.56	0.49	0.50	0.50
$\|\rho\| < 0.8$ [b]	0.59	0.56	0.51	0.56	0.48
$\|\rho\| < 0.7$ [c]	0.60	0.56	0.52	0.58	0.51
$\|\rho\| < 0.6$ [c]	0.61	0.58	0.55	0.59	0.51
$\|\rho\| < 0.5$ [c]	0.60	0.53	0.54	0.55	0.53
$\|\rho\| < 0.4$ [c]	0.52	0.51	0.54	0.56	0.58
$\|\rho\| < 0.3$ [c]	0.49	0.47	0.50	0.40	0.50
$\|\rho\| < 0.2$ [c]	0.55	0.50	0.51	0.50	0.50
$\|\rho\| < 0.1$ [c]	0.56	0.45	0.49	0.57	0.45

[a] *See Table 4.3.*

[b] *See Table 4.4 and Table 4.5.*

[c] *See Table 4.4.*

Classification with Significant MRI Features

Table 4.14 shows that the XGBoost classifier achieved the highest WAF1 score of 0.63 for classifying motor phenotype using only significant MRI features. Figure 4.20 and Table 4.15 show the accompanied confusion matrix and classification report. The majority of both PIGD and TD cases were classified correctly. Moreover, the precision, recall, and F1 score were higher for PIGD than for TD.

The permutation importance presented in Figure 4.21 highlights that the right SNpc in the simDIR WM image was the most crucial feature for the XGBoost classification of PIGD/TD motor phenotype using significant MRI features, followed by the LC in the simDIR GM image and the right SNpc in the simDIR GM image.

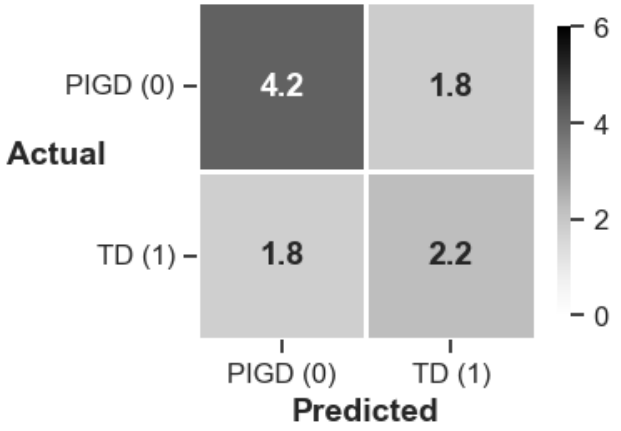

Figure 4.20: Confusion matrix for the XGBoost classification of PIGD/TD motor phenotype with significant MRI features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, HC = healthy control, PD = Parkinson's disease, MRI = magnetic resonance imaging.

Table 4.15: Classification report for the XGBoost classification of PIGD/TD motor phenotype with significant MRI features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
PIGD	0.71	0.69	0.69	6
TD	0.56	0.56	0.54	4
ACC			0.64	10
MA	0.64	0.63	0.61	10
WA	0.65	0.64	0.63	10

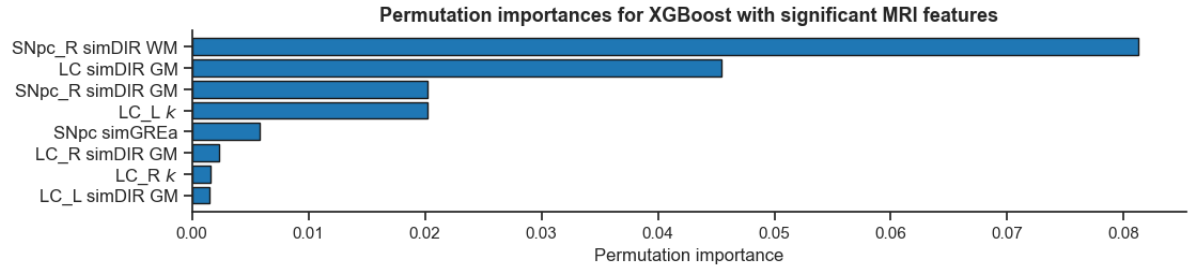

Figure 4.21: The eight highest permutation importances for the XGBoost classification of PIGD/TD motor phenotype with significant MRI features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging, SNpc_R = right substantia nigra pars compacta, simDIR = simulated double inversion recovery, WM = white matter, LC = locus coeruleus, GM = gray matter, LC_L = left locus coeruleus, k = radio frequency transmit field variation, SNpc = substantia nigra pars compacta, simGRE = simulated gradient echo, LC_R = right locus coeruleus, LC_R = right locus coeruleus, PIGD = postural instability gait difficulty, TD = tremor-dominant.

4.7.2 Classification with Selected Clinical Features

The WAF1 scores for the PIGD/TD classification of patients with selected clinical features are provided in Table 4.16. The clinical features considered in this analysis included gender, age, MoCA score, and B-SIT score. The XGBoost classification with highly significant MRI features and selected clinical features achieved the highest WAF1 score of 0.69. This combination was therefore explored further.

Table 4.16: WAF1 scores for the PIGD/TD classification of patients with selected clinical features. The WAF1 score is an average of the WAF1 score of 100 train/test splits with random state 0-99. The highest score is highlighted in bold. The test group included 6 PIGD and 4 TD. WAF1 = weighted average F1, MRI = magnetic resonance imaging, XGBoost = extreme gradient boosting, ρ = Spearman's rank correlation coefficient.

MRI features	**WAF1 score** XGBoost
All	0.52
Significant [a]	0.68
Highly significant [a]	**0.69**
$\|\rho\| < 0.9$ [b]	0.50
$\|\rho\| < 0.8$ [b]	0.49
$\|\rho\| < 0.7$ [c]	0.50
$\|\rho\| < 0.6$ [c]	0.52
$\|\rho\| < 0.5$ [c]	0.54
$\|\rho\| < 0.4$ [c]	0.58
$\|\rho\| < 0.3$ [c]	0.50
$\|\rho\| < 0.2$ [c]	0.52
$\|\rho\| < 0.1$ [c]	0.52

[a] *See Table 4.3.*

[b] *See Table 4.4 and Table 4.5.*

[c] *See Table 4.4.*

Classification with Highly Significant MRI Features and Selected Clinical Features

The classification of motor phenotype using XGBoost with highly significant MRI features and selected clinical features yielded an WAF1 score of 0.69, as shown in Table 4.16. Figure 4.22 and Table 4.17 show the accompanied confusion matrix and classification report. The majority of both PIGD and TD cases were classified correctly. Moreover, the precision, recall, and F1 score were higher for this classification than for the one without clinical features.

The permutation importances presented in Figure 4.21 indicate that the right SNpc in the simDIR GM image was an essential feature for the XGBoost classification of PIGD/TD motor phenotype using highly significant MRI features and selected clinical features, followed by the LC in the simDIR GM image and the MoCA score.

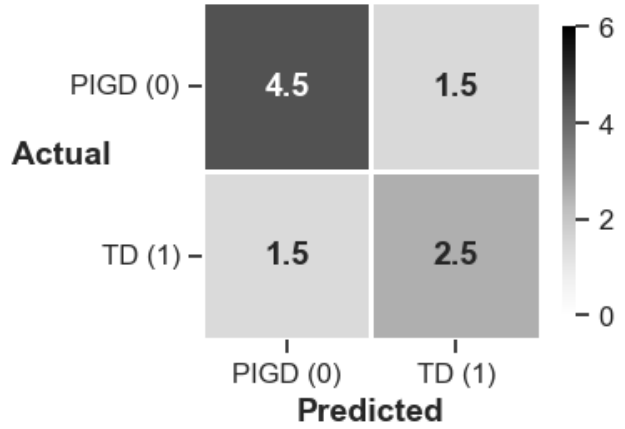

Figure 4.22: Confusion matrix for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and selected clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, HC = healthy control, PD = Parkinson's disease, MRI = magnetic resonance imaging.

Table 4.17: Classification report for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and selected clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
PIGD	0.77	0.75	0.75	6
TD	0.63	0.63	0.61	4
ACC			0.70	10
MA	0.70	0.69	0.68	10
WA	0.71	0.70	0.69	10

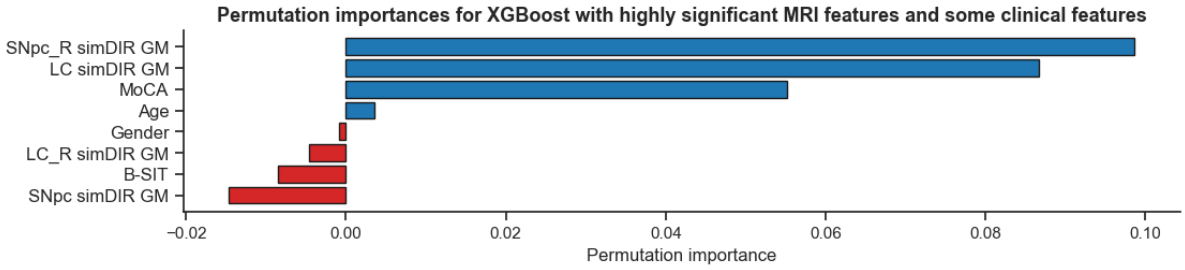

Figure 4.23: Permutation importances for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and selected clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging, SNpc_R = right substantia nigra pars compacta, simDIR = simulated double inversion recovery, GM = gray matter, LC = locus coeruleus, MoCA = Montral cognitice assessment, LC_R = right locus coeruleus, B-SIT = Brief Smell Identification Test, SNpc = substantia nigra pars compacta, PIGD = postural instability gait difficulty, TD = tremor-dominant.

4.7.3 Classification with All Clinical Features

The WAF1 scores for the PIGD/TD classification of patients with all clinical features are provided in Table 4.18. The XGBoost classification with highly significant MRI features and all clinical features achieved the highest WAF1 score of 0.66. This combination was therefore explored further.

Table 4.18: WAF1 scores for the PIGD/TD classification of patients with all clinical features. The WAF1 score is an average of the WAF1 score of 100 train/test splits with random state 0-99. The highest score is highlighted in bold. The test group included 6 PIGD and 4 TD. WAF1 = weighted average F1, MRI = magnetic resonance imaging, XGBoost = extreme gradient boosting, ρ = Spearman's rank correlation coefficient.

MRI features	**WAF1 score** XGBoost
All	0.50
Significant [a]	0.64
Highly significant [a]	**0.66**
$\|\rho\| < 0.9$ [b]	0.48
$\|\rho\| < 0.8$ [b]	0.51
$\|\rho\| < 0.7$ [c]	0.52
$\|\rho\| < 0.6$ [c]	0.53
$\|\rho\| < 0.5$ [c]	0.55
$\|\rho\| < 0.4$ [c]	0.54
$\|\rho\| < 0.3$ [c]	0.53
$\|\rho\| < 0.2$ [c]	0.55
$\|\rho\| < 0.1$ [c]	0.57

[a] *See Table 4.3.*

[b] *See Table 4.4 and Table 4.5.*

[c] *See Table 4.4.*

Classification with Highly Significant MRI Features and All Clinical Features

The classification of motor phenotype using XGBoost with highly significant MRI features and all clinical features resulted in a WAF1 score of 0.66, as shown in Table 4.18. Figure 4.24 and Table 4.19 respectively show the accompanied confusion matrix and classification report. The majority of both PIGD and TD cases were classified correctly. However, it is noteworthy that the F1 scores for PD and HC were lower than those for the classification that incorporated only selected clinical features.

The permutation importances presented in Figure 4.25 indicate that the right SNpc in the simDIR GM image was the most crucial feature for the XGBoost classification of PIGD/TD motor phenotype using highly significant MRI features and all clinical features, followed by the MoCA score and the LC in the simDIR GM image.

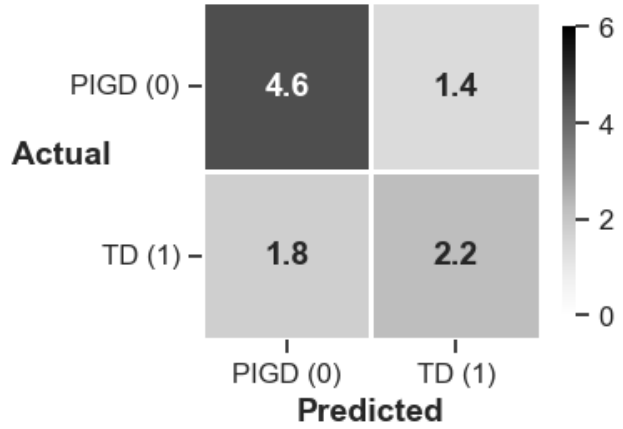

Figure 4.24: Confusion matrix for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and all clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, HC = healthy control, PD = Parkinson's disease, MRI = magnetic resonance imaging.

Table 4.19: Classification report for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and all clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, PREC = precision, REC = recall, F1 = F1 score, SUP = support, HC = healthy control, PD = Parkinson's disease, ACC = accuracy, MA = macro average, WA = weighted average, MRI = magnetic resonance imaging.

	PREC	REC	F1	SUP
PIGD	0.74	0.76	0.74	6
TD	0.60	0.56	0.55	4
ACC			0.68	10
MA	0.67	0.66	0.65	10
WA	0.68	0.68	0.66	10

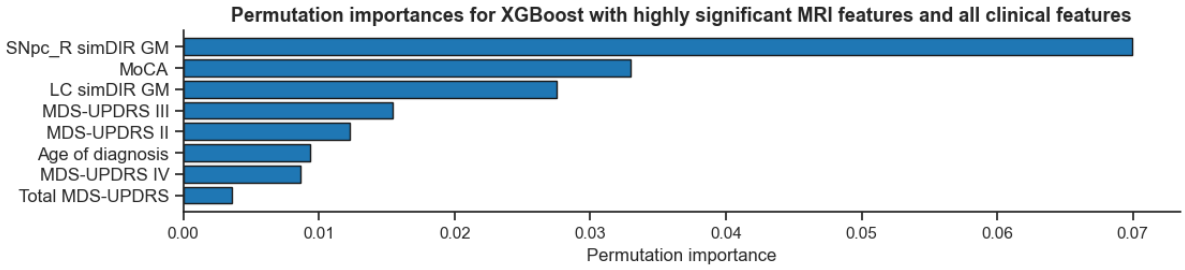

Figure 4.25: The eight highest permutation importances for the XGBoost classification of PIGD/TD motor phenotype with highly significant MRI features and all clinical features. An average of 100 train/test splits with random state 0-99. XGBoost = extreme gradient boosting, MRI = magnetic resonance imaging, SNpc_R = right substantia nigra pars compacta, simDIR = simulated double inversion recovery, GM = gray matter, MoCA = Montreal Cognitive Assessment, LC = locus coeruleus, MDS-UPDRS = Movement Disorder Society Unified Parkinson's Disease Rating Scale, PIGD = postural instability gait difficulty, TD = tremor-dominant.

Chapter 5

Discussion

This thesis aimed to investigate the possibility of utilizing MRI to classify study participants as patients or HCs and evaluate its capacity to stratify patients according to their motor phenotype. Both iron-sensitive and non-iron-sensitive MRI techniques were employed to measure the loss of DANs in the SNpc and NANs in the LC. This chapter presents the findings from Chapter 4 and discusses the methods' limitations.

5.1 Laterality

Participants showed significant laterality in both the SNpc and the LC in multiple images. Laterality in the SNpc indicates asymmetric DAN loss in the brain, while laterality in the LC indicates asymmetric NAN loss. This supports previous research showing PD's asymmetric nature, with only 16.4% of patients exhibiting symmetric motor symptoms [107]. Future studies should consider including motor function laterality as a clinical feature and see if it correlates with the laterality in the MRI images.

5.2 Feature Selection Based on Statistical Testing

The first hypothesis in this thesis was that individuals with PD exhibit more significant degeneration of DANs in the SNpc and NANs in the LC compared to HCs, and that this neuronal loss can be observed through increased iron and decreased NM content. Elevated iron levels lead to a decrease in T2* and an increase in R2* [46], a decrease in SWI [31], an increase in meiSWIM [48], and a decrease in phase [47]. Reduced NM content results in a decrease in T1 [28]. Moreover, PrD maps have indicated a decreased signal in the SNpc of PD patients compared to HCs, suggesting that degeneration leads to a reduction in the PrD signal [33].

Significantly higher mean intensity levels were observed in PD patients compared to HCs in all regions of the LC within the simGRE image. Moreover, within the simDIR CSF image, PD patients exhibited significantly higher mean intensity levels in the SNpc than HCs. Interestingly, these two images were the only ones to demonstrate significant differences between PD patients and HCs. This finding is interesting as it contrasts with the initial expectation that SWI-based techniques and relaxation maps would be the best methods for distinguishing between PD patients and HCs.

The PrD maps exhibited a trend wherein higher signal intensity was observed in HCs compared to PD patients, aligning with the hypothesis and the findings of Depierreux et al. [33]. In the SNpc, PD patients tended to have decreased T2*, increased R2*, decreased SWI, and increased meiSWIM. These trends indicate that PD patients have a greater iron accumulation in the SNpc compared to HCs, which aligns with the hypothesis.

A similar trend was observed in the LC for T2*, R2*, and meiSWIM, suggesting that PD patients have a higher iron deposition than HCs in the LC. It is important to note that these associations did not reach statistical significance. However, they might have been significant with larger sample sizes, particularly for HCs. No trends were observed for the T1 map and the T1WE image. This can be because previous studies assessing NM have employed sequences with MTC, which was not used in this thesis [78][28].

The second hypothesis addressed in this thesis proposed that individuals with PIGD motor phenotype exhibited more pronounced degeneration of DANs in the SNpc and NANs in the LC compared to TD patients. The mean intensity within the simDIR GM image was significantly higher in TD patients than in PIGD patients across all SNpc and LC regions. DIR images are sensitive to variations in T1, which is associated with NM content [108]. The higher intensity in the simDIR GM image suggests a higher NM content. Consequently, patients with the PIGD motor phenotype demonstrate lower NM content than patients with the TD motor phenotype in both the SNpc and the LC. This finding supports the hypothesis and is consistent with previous studies that have reported elevated iron content in the SNpc of PIGD patients compared to non-PIGD patients [12].

The anticipation was that SWI-based techniques and relaxation maps would best differentiate between PIGD and TD patients. However, similar to the PD/HC classification, the simulated images exhibited superior discriminatory capabilities.

Among the non-simulated images, the signal intensity in the SNpc within the CROWN PrD map and the CROWN true PrD A map was significantly higher in TD patients compared to PIGD patients. This finding aligns well with the hypothesis, suggesting that PIGD patients experience more pronounced SNpc degeneration than TD patients.

TD patients also exhibited significantly higher mean intensity than PIGD patients in the left SNpc and the SNpc in the simGREa image. In contrast, PIGD patients had significantly higher mean intensity than TD patients in the right SNpc within the simDIR WM image, the SNpc in the simDIR CSF image, and all regions of the SNpc and the LC in the *k* map.

The observed trend, where PIGD patients exhibited higher mean intensity in the SNpc compared to TD patients within the R2* maps, indicates greater iron deposition in the SNpc for PIGD patients, consistent with the hypothesis. However, the trends indicating higher mean intensity in the LC for PIGD compared to TD in both the T2* maps and meiSWIM maps contradict expectations since increased iron leads to decreased T2* and increased meiSWIM [46][48].

5.3 Feature Selection Based on Pairwise Correlation

Consistent with expectations, several MRI features exhibited strong correlations with each other. After eliminating features with a pairwise absolute Spearman's rank correlation coefficient $|\rho| > 0.9$, the number of MRI features decreased from 228 to 108. When features with $|\rho| > 0.4$ were removed, no lateral MRI features were no longer present. The two least correlated features were the CROWN PrD map and the CROWN R2* map in the SNpc. This observation is logical, as PDW sequences aim to minimize T2* contrast.

5.4 Differentiating PD from HC

The selection of the classifier and MRI features significantly affected the WAF1 score when classifying participants as either PD patients or HCs without clinical features, as depicted in Table 4.6. Among the combinations tested without clinical features, four achieved the highest WAF1 score of 0.70: 1) RaF classification using significant MRI features, 2) SVC with MRI features where $|\rho| < 0.9$, 3) SVC with MRI features where $|\rho| < 0.6$, and 4) SVC with MRI features where $|\rho| < 0.4$. Although these four combinations obtained the same WAF1 score, their performance varied for each class.

The RaF classification with significant MRI features and the SVC with MRI features where $|\rho| < 0.9$ accurately predicted the most PD patients. Conversely, the SVC with MRI features where $|\rho| < 0.4$ achieved the highest number of correct predictions for HCs.

The permutation importances for the top four combinations of classifiers and MRI features when clinical features were not considered, highlighted the importance of the following features in distinguishing between PD and HC: the LC in the simGRE image, the SNpc in the simDIR CSF image, the LC in the simFLAIR image, the LC in the mpSWIM map, and the LC in the CROWN R2* map. However, whether these features are genuinely crucial for this specific classification task is unclear. As discussed in Section 3.10.3, the neutral WAF1 score for the PD/HC classification is 0.64, representing the score obtained without any prior information. The achieved score, higher than 0.64, indicates that the model has been able to generalize to some extent based on the provided information. However, an increase of 0.06 in the WAF1 score is relatively small.

Integrating clinical features into the classification model alongside MRI features resulted in higher WAF1 scores compared to using MRI features alone. When clinical features were included, the XGBoost classifier achieved the highest WAF1 score of 0.84 when utilizing significant MRI features and highly significant MRI features. This score of 0.84 is significantly higher than the score obtained without clinical data, 0.70, and the neutral score of 0.64 obtained without any information. This finding suggests that gender, age, MoCA score, and B-SIT score provide valuable information that aids in determining the PD/HC status of the participants. The F1 score for PD patients was 0.90, while the F1 score for HCs was 0.66, suggesting that the model performed better in identifying PD patients than HCs. This discrepancy in performance could be attributed to class imbalance, with PD patients being the majority class.

The permutation importances for the top two combinations of classifiers and MRI features, with clinical features, indicated that the B-SIT score, in combination with the LC in the simGRE image, was the most crucial feature when distinguishing between PD and HC.

Using all MRI features, regardless of whether clinical data were included, reduced the WAF1 score. While adding more features to a classification model often improves its ability to distinguish between classes, there are instances where including more features can decrease the WAF1 score [95].

This phenomenon can be attributed to overfitting, where the model becomes overly complex and starts to fit the noise in the training data rather than capturing the underlying patterns. Consequently, although the model may perform well on the training data, it may struggle to generalize effectively to new data, resulting in a decline in the WAF1 score. Moreover, in cases where the classes are imbalanced, adding more features can cause the model to exhibit bias towards the majority class, leading to poor performance on the minority class and a subsequent decrease in the WAF1 score.

5.5 Differentiating PIGD from TD

In the absence of clinical features, the WAF1 score for the PIGD/TD classification was influenced by choice of classifier and MRI features, similar to the PD/HC classification. The XGBoost classifier achieved the highest WAF1 score of 0.63 for the motor phenotype classification, utilizing only significant MRI features. This score of 0.63 is significantly higher than the neutral WAF1 score of 0.45, representing the score obtained without any prior information. The permutation importances for the optimal combination of classifier and MRI features, without clinical features, indicate that the SNpc in the simDIR WM image and the LC in the simDIR GM image were essential when distinguishing between PIGD and TD.

By incorporating the clinical features of age, gender, MoCA score, and B-SIT score into the classification model alongside MRI features, higher F1 scores were achieved compared to when using MRI features alone, suggesting that these clinical features provide valuable information for identifying the motor phenotype of PD patients. When incorporating these clinical features, the XGBoost classifier attained the highest WAF1 score of 0.69 when utilizing highly significant MRI features. This score of 0.69 surpasses the score obtained without clinical data of 0.64 and the neutral score of 0.45 without any information. The F1 score for PIGD patients was 0.75, while for HC patients, it was 0.61, indicating that the model performs slightly better in identifying PIGD patients than TD patients, potentially due to class imbalance, with PIGD being the majority class.

The permutation importances for the best combination of classifier and MRI features, considering selected clinical features, indicated that the SNpc and the LC in the simDIR GM image and the MoCA score were the most significant when distinguishing between PIGD and TD.

When incorporating all clinical features, including MDS-UPDRS scores, the XGBoost classifier attained the highest WAF1 score of 0.66 when utilizing highly significant MRI features. The decreased score observed when adding all clinical features suggests that these features provide redundant information when classifying patients into PIGD and TD. The permutation importances for this classification highlighted that the SNpc and the LC in the simDIR GM image, in conjunction with the MoCA score, were the most influential features when distinguishing between PIGD and TD.

Like the classification of PD/HC status, the classification of motor phenotype also exhibited a decline in the WAF1 score when utilizing all MRI features instead of solely significant MRI features. Several factors can contribute to this, including overfitting, irrelevant or redundant features, class imbalance, or insufficient training samples [95].

5.6 Methodological Considerations

5.6.1 Participants

This thesis included 58 PD patients and 18 HCs, resulting in imbalanced classes. Imbalanced classes can affect machine learning algorithms, leading to biased predictions and poorer performance for the minority class [95]. Although excluding patients would have addressed the issue, it would have resulted in small sample sizes for both PD patients and HCs. Therefore, the class imbalance was tolerated. Additionally, the control group is often considered more homogeneous and requires a smaller sample size. The ongoing STRAT-PARK study will include 1500-200 PD patients and HCs, providing more data for future studies. It is recommended that future studies prioritize balanced classes to ensure accurate predictions for all classes and prevent bias towards the majority class.

The participant group had balanced age distributions: PD patients (69$\pm$9 years) and HCs (65$\pm$10 years). The patient group comprised 22 females and 36 males, while the HC group comprised 12 females and six males. The significant difference in gender balance between the two groups could have affected the results and should be avoided in future studies. Among the PD patients, 20 had TD motor phenotype and 27 had PIGD motor phenotype. 11 patients with an indeterminate motor phenotype were excluded from motor phenotype classification. The limited data availability during the study was the reason for the small sample sizes. In future studies, it is recommended to include a larger number of instances of each class to improve the statistical power of the analysis.

5.6.2 Clinical Data

The clinical data contained missing values. Specifically, 5 participants had missing B-SIT scores, and 3 had missing MoCA scores. Additionally, 22 patients had missing age of diagnosis, six had missing age of symptoms and disease duration, nine had missing MDS-UPDRS scores, and eight had missing MDS-NMS scores. The presence of missing values can affect the accuracy and validity of study results, particularly if the missing values are not randomly distributed, leading to biased estimates of population values [109].

The presence of missing values influenced the selection of algorithms in this study. XGBoost was specifically chosen due to its inherent capability to handle missing values by default. On the other hand, LR, DT, RaF, and SVC were not employed on datasets containing missing values. If no missing values existed in the datasets, one of these models could potentially have outperformed XGBoost.

Although the sample size could have been reduced to exclude participants with missing data, this approach was not advisable due to the already small sample size and the large number of samples with missing values. Future studies should minimize the number of missing values and carefully consider how to handle them to ensure the accuracy and validity of the results.

5.6.3 MRI Acquisition

During the MRI scans used in this thesis, three different sequences were employed, resulting in a total scan time of over 16 minutes. Given the duration of the scans, it is unlikely that participants could remain perfectly still throughout the entire process. Consequently, head movements could have introduced errors in the resulting images. While co-registration and normalization techniques can help mitigate the impact of head movements to some extent, errors due to movement could still be present in the means calculated during the ROI analysis. Furthermore, other factors such as B_0 inhomogeneity, gradient artifacts, and RF noise may have also contributed to image errors. However, correction for B_0 inhomogeneity and gradient non-linearity was performed when generating images and maps with STAGE, leading to minimized effects.

5.6.4 Generating Images and Maps

The ME-GRE sequences produced various images and maps with different contrasts using STAGE. The image processing time per participant was approximately 15 minutes on a consumer-grade laptop equipped with an AMD Ryzen 7 4700U processor featuring eight cores. During this process, k maps were generated, with fixed values of T1WM and T1GM at 900 ms and 1600 ms, respectively. However, the chosen values of T1WM and T1GM may vary in different research studies and could potentially impact the results [44].

The SWI image generated had a lower resolution than the usual isotropic in-plane SWI scans [79]. However, it might be possible to double the SWI resolution by employing a split k-space coverage concept for the second echoes. Additionally, although not utilized in this study, STAGE allows MTC data to be included. Incorporating MTC data offers another approach to NM-MRI and should therefore be considered in future research endeavors [110].

5.6.5 Image Preprocessing

The images and maps generated by STAGE were preprocessed using co-registration and normalization techniques. Specifically, FSL FLIRT was used to co-register the STAGE images and maps with the MP-RAGE image. The T1WE image was chosen to create the transformation matrix, as it resembled the MP-RAGE image and provided satisfactory co-registration results. However, other STAGE images may be better suited for this task, and future studies should consider exploring alternative reference images and other co-registration tools.

The MP-RAGE image was normalized to the MNI152 T1 1mm brain template using SPM12. However, this standard procedure proved inadequate for three participants, potentially due to head movements during acquisition as discussed in Section 5.6.3. For two participants, normalization with FSL FLIRT was used instead. For the third participant, the origin and angle of the MP-RAGE image had to be manually set before normalization in SPM12 to obtain acceptable results. Future studies should explore alternative normalization tools to optimize the preprocessing of the STAGE images and maps. Additionally, it may be worth considering the possibility of cropping the images into smaller sizes and manually adjusting their angles.

5.6.6 Region of Interest Analysis

In this thesis, the SNpc atlas was a population-based probabilistic atlas generated through manual delineation, normalization, and averaging a 3T NM-MRI sequence [89]. The atlas was thresholded at a probability of 0.5 to include most of the voxels within the atlas and prevent overestimation of the SNpc region. The resulting 1x1x1 mm SNpc atlas had a volume of 500 mm^3 (left SNpc 235 mm^3 and right SNpc 265 mm^3). On the other hand, the LC atlas used was a population-based probabilistic atlas generated through manual delineation, normalization, and averaging a 7T MRI sequence with a spatial resolution of 0.4x0.4x0.5 mm [93].

Three versions of the LC atlas were created based on the slice-wise distribution of group-level probabilities. A zero threshold (0%) was applied to obtain a template (925 voxels, 115.625 mm^3) more sensitive to the LC spatial distribution. A liberal threshold (5%) was applied to obtain a template (705 voxels, 88.125 mm^3) which was sensitive to the LC spatial distribution while controlling for the potential noise seen in the caudal part of the LC. A stringent threshold (25%) was applied to the atlas to produce a template (284 voxels, 35.5 mm^3) more specific to the core LC locations.

In this thesis, the un-thresholded atlas was chosen to ensure that as much as possible of the LC, with a volume of 125.7 ± 59.3 mm^3 in HCs, was present inside the mask [111]. This was done partly due to possible normalization errors, which can significantly affect small structures such as the LC, and partly because of significant individual variations in LC volume.

The LC atlas had to be re-sampled to a resolution of 1x1x1 mm using FSL FLIRT. Again, a zero threshold (0%) was set to ensure that the entire LC was present inside the mask for all participants, even with a small normalization error. The resulting 1x1x1 mm LC atlas had a volume of 280 mm^3 (left LC 140 mm^3 and right LC 140 mm^3), about double the size of the actual LC.

The choice of thresholds could have significantly affected the results, as the mean intensity of all voxels in the mask depends on which voxels are included. The disadvantage of choosing zero thresholds is that they can introduce noise from the caudal part of the LC or neighboring structures in the dorsal pons ventral of the fourth ventricle. Future studies should experiment with more stringent thresholds of the LC mask.

Another potential avenue for future research involves investigating the size of the SNpc and the LC in study participants to identify potential differences between individuals with PD and HCs and variations between the PIGD and TD motor phenotypes. The size of the SNpc and the LC can be found by converting the masks into the participants' native space using the inverse transformation method employed during normalization.

The main advantage of atlas-based segmentation methods is the possibility of segmenting any brain structure available for multiple study participants without any additional cost [112]. In contrast, manual segmentation, which refers to the process where a human operator segments and labels an image by hand, is time-consuming and thus expensive. A disadvantage of atlas-based segmentation methods is that they are directly dependent on the quality of the registration method used. Registration errors are transferred directly to the ROI analysis when using atlases.

Another disadvantage is that atlases depend on the population they are constructed on [112]. The SNpc and LC atlas used in this thesis were constructed on healthy individuals. Several studies have shown that the volume of the SNpc and the LC is smaller in PD patients than in HCs due to neuronal degeneration [113][114][115]. Therefore, the atlases are better suited for HCs than PD patients. The SNpc atlas was also constructed on healthy individuals with a mean age of 39 ± 11 years. However, the mean age of the participants in this thesis was 69 ± 9 for PD patients and 65 ± 10 for HCs. A recent study showed that the NM region in the SNpc increase until the 30s and decreases into the 80s [116]. Therefore, the SNpc atlas used in this thesis may not be the best suited for either PD patients or HCs. Future studies should consider using alternative atlases for the SNpc and the LC with different biases or properties. Additionally, future studies should explore other segmentation methods, such as automatic segmentation methods.

Automatic segmentation methods, such as the Multimodal Image Segmentation Tool, have produced high-quality segmentations of the SNpc in 7T MRI, suggesting that automatic segmentations are at least as accurate as manual delineations and may be less sensitive to confounding differences in image intensity between participants [117]. Additionally, the deep neural network U-Net has segmented the SNpc with results that are more accurate and better correspond to theoretical assumptions about the volume of NM in the brain in different stages of PD than atlas-based methods [118]. A disadvantage of the deep learning approach for segmentation is that it requires a lot of labeled training data [100].

However, U-Net is optimized for segmenting biomedical images, so it does not need as much training data as standard convolutional neural networks [118]. Like the atlas-based approach, deep learning methods depend on the representation of the training data, meaning that training data should include not only HCs but also PD patients.

5.6.7 Laterality

In this thesis, the determination of laterality in the MRI images was based on statistical hypothesis testing. However, alternative measures for assessing laterality exist. Future studies should consider exploring alternative methods, such as the laterality index (LI), commonly used to indicate hemispheric dominance in functional MRI studies [119]. The LI is typically calculated as the difference between the activity levels in each hemisphere divided by the total activity across both hemispheres. In the context of this research, the LI could be defined as the difference in mean intensity between each part of the SNpc divided by the mean intensity of the entire SNpc. A similar approach could be applied to the LC.

5.6.8 Preprocessing Features

An *outlier* is a point that lies very far from the mean of the corresponding random variable [95]. Points with values very different from the mean value produce significant errors during training and may have disastrous effects. Outliers in small populations can be tricky to detect, and there is an increased risk of removing observations that are not outliers. No outliers were removed in this thesis due to the small sample size. However, this means that participants with values significantly different from the median value could have potentially introduced significant errors during training [95]. Therefore, future studies with larger sample sizes should consider the removal of outliers.

The features were scaled using standardization, subtracting the mean from each feature and dividing by the standard deviation. However, none of the MRI features followed a normal distribution, making the representation of feature values by the median more appropriate than the mean. Consequently, standardization may not be the most suitable scaling technique in this context. An alternative approach is a min-max normalization, which ensures that each feature is scaled to the [0,1] interval without assuming normality [120].

The MRI data did not contain any missing values. However, the clinical data did. Most classifiers, such as the logistic regression (LR) classifier, the decision tree (DT) classifier, the random forest (RaF) classifier, and the support vector classification (SVC), do not handle missing values by default. However, some classifiers, such as eXtreme Gradient Boosting (XGBoost), do this. When the dataset had no missing values, LR, DT, RF, and SVC classifiers were employed in addition to XGBoost. However, when the dataset included missing values, only XGBoost was utilized. This limitation could have been addressed by imputing the missing values.

Traditional techniques for dealing with missing data involve replacing the missing values with (a) zeros, (b) the unconditional mean computed from the available values of the respective feature, or (c) the conditional mean if the probability density function of the missing values given the observed data is estimated [95]. A more advanced approach is multivariate feature imputation, where each feature with missing values is modeled as a function of other features, and the estimated values are used for imputation [121]. Future studies should explore different imputation techniques and investigate whether alternative machine learning classifiers can outperform XGBoost in this context.

5.6.9 Feature Selection

In this thesis, the selection of features was based on statistical testing and pairwise correlation. Evaluating features individually offers computational simplicity but may not be effective for complex problems and highly correlated features [95]. Alternatively, feature vector selection techniques evaluate the classification capabilities of feature vectors rather than individual features. Another approach is dimensionality reduction, which involves transforming a given set of features into a smaller set. In this thesis, dimensionality reduction was not performed, as the goal was to understand the importance of individual features in the classification process. However, if the objective was to build a good model without focusing on feature interpretation, dimensionality reduction techniques such as principal component analysis and independent component analysis could have been employed for dimensionality reduction.

5.6.10 Differentiating PD from HC and TD from PIGD

Datasets

Stratified sampling was employed to divide the data into training and test sets, allocating 80% of the data for training and 20% for testing. Stratified sampling helps to mitigate bias towards the majority class. There is no definitive guidance on the ideal data allocation for training and testing [122]. Determining a suitable train-test split depends on various factors, including dataset size, model complexity, and the nature of the problem. In this thesis, an 80:20 split ratio was selected as it is widely used and provides a significant portion of the data for training while still retaining enough for testing purposes [122].

Classification

This thesis utilized five machine learning classifiers: LR, DT, RF, SVC, and XGBoost. These classifiers were chosen for their well-established nature and diverse characteristics. LR is a simple probabilistic-based model. While LR can encounter overfitting with high-dimensional datasets, it performs well when the dataset can be linearly separated [102]. Regularization techniques such as L1 and L2 regularization can be applied to mitigate overfitting in such scenarios. One notable drawback of LR is its assumption of linearity between the dependent and independent variables.

On the other hand, SVC is particularly effective in high-dimensional spaces, and its behavior can vary depending on the choice of mathematical functions, known as kernels [102]. The most frequently used kernel in SVC is the linear function. However, SVC may encounter challenges with noisy data or overlapping classes.

DT is a non-parametric model that is easily visualized and simple to understand. It can capture interactions among variables. However, a drawback of DT is that it is prone to overfitting [123]. RF, on the other hand, is a well-known ensemble classification technique. Consequently, the RaF learning model, consisting of multiple decision trees, tends to yield higher accuracy than a single DT-based model [102]. RF is recognized as one of the most accurate learning algorithms available. However, a disadvantage of RF is that the resulting classifications can sometimes be challenging to interpret.

XGBoost is an ensemble learning algorithm [102]. It uses advanced techniques such as computing second-order gradients of the loss function to minimize loss and employing regularization to reduce overfitting, enhancing model generalization and performance.

XGBoost is known for its ability to handle large datasets efficiently and the built-in capability to handle missing values. However, it is worth noting that XGBoost is highly sensitive to outliers, as each classifier is compelled to correct errors made by preceding learners.

Alternative classifiers that could have been utilized in this thesis include K-nearest neighbors, adaptive boosting, and stochastic gradient descent [102]. Additionally, clustering algorithms such as k-means could have been employed [95].

In general, constructing an effective machine learning model involves not only selecting the appropriate algorithm, but also obtaining an optimal model architecture by fine-tuning its hyperparameters [124]. *Hyperparameters* are parameters used to configure a machine learning model or specify the algorithm used to minimize the loss function. The process of designing the ideal model architecture with an optimal hyperparameter configuration is known as hyperparameter tuning. While hyperparameter tuning was not prioritized in this thesis, it is strongly encouraged in future studies as it is considered a crucial component of building an effective machine-learning model, particularly for tree-based models with numerous hyperparameters.

Evaluation

The primary evaluation metric in this thesis was the WAF1 score. For this, participants with a probability score of 0.5 or higher were classified as PD or TD, while those with a probability score below 0.5 were classified as HC or PIGD. The receiver operating characteristics curve (ROC) and the precision-recall (PCR) plot can provide an overview of the model's performance across different thresholds [103]. The ROC plot illustrates the tradeoff between specificity and sensitivity, while the PRC plot shows precision values for corresponding recall values. It is recommended that future studies utilize ROC or PCR plots to determine the best threshold.

Permutation feature importance was employed to evaluate and identify the most important features. Permutation feature importance offers several advantages, including comparability across different problems and classifiers [105]. However, a major drawback of permutation feature importance is that the importance of an associated feature can decrease if a correlated feature is added, as it splits the importance between them. This can lead to incorrect estimates of feature importance when correlations exist among features. Given that many of the features in this thesis were correlated, there may be more suitable algorithms than permutation feature importance in this context.

An alternative algorithm is PIMP, which adapts the permutation feature importance by providing p-values for the importances [125]. Variance-based measures present another model-agnostic option for feature importance assessment. These measures, such as Sobol's indices and functional analysis of variance, assign higher importance to features that cause more significant variance in the prediction function [105]. Shapley additive explanations importance also shares similarities with variance-based importance measures. It is recommended for future studies to explore other methods for feature evaluation.

Chapter 6

Conclusions and Future Work

This thesis successfully demonstrated the use of multimodal MRI for differentiating PD patients from HCs and stratifying PD patients based on their motor phenotypes, specifically PIGD and TD. To achieve this, a robust pipeline was implemented, utilizing an atlas-based segmentation method to compute the mean intensity of the left SNpc, the right SNpc, the SNpc, the left LC, the right LC, and the LC in various MRI images and maps, including iron-sensitive images and maps. Two feature selection techniques were employed to reduce the number of MRI features while retaining their class-discriminatory information. Five machine learning classifiers were compared using various evaluation metrics, including the WAF1 score. The most important features were identified through the technique of permutation feature importance.

The mean intensity of the LC in the simGRE image emerged as a promising biomarker for distinguishing PD patients from HCs, especially when combined with the B-SIT score. Additional potential biomarkers for PD/HC classification include the mean intensity of the SNpc in the simDIR CSF image, the LC in the simFLAIR image, the LC in the mpSWIM image, and the LC in the CROWN R2* map. To ensure balanced data representation and prevent bias towards the PD patient class, future studies should include a larger number of HCs.

Promising biomarkers for differentiating PIGD patients from TD patients include the mean intensity of the SNpc and the LC in the simDIR GM image, the SNpc in the simDIR WM image, and the MoCA score. These findings are novel in the field and demonstrate the potential of this approach to contribute to more precise diagnosis of PD patients in a clinical setting. To further strengthen these results, it would be valuable to investigate their reproducibility in an independent cohort or a larger dataset.

Future studies should consider exploring additional NM-MRI techniques such as MTC [110]. Moreover, non-atlas-based segmentation methods, like deep learning segmentation with U-Net [118], should be investigated. Additionally, including the volume of the SNpc and the LC as additional features in the classification process should be considered in future research. Furthermore, it is recommended that future studies analyze the relationship between motor asymmetry and multimodal MRI laterality among patients with different motor phenotypes.

Code Availability

All the relevant code utilized to produce the results presented in this thesis can be accessed at the following URL: https://github.com/signehogstad/master. The repository will be regularly updated and supplemented with explanatory comments.

Appendix A

Mean of SNpc and LC for PD and HC

Tables A.1-A.6 show the mean intensities of the SNpc and the LC for PD patients and HCs in all images and maps generated in this thesis. The *p*-values were determined using the Mann-Whitney U test.

Table A.1: Mean intensities of the SNpc and the LC in the R2, the R2* MIP, the CROWN R2*, and the CROWN R2* A2 for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.*

Image/map	**Structure**		**Median** $(\mathbf{Q_1} - \mathbf{Q_3})$		**p-value**
			PD	**HC**	
R2*	SNpc	Total	**3178 (2893 - 3510)**	3071 (2819 - 3326)	0.3
		Left	**3153 (2804 - 3489)**	2982 (2818 - 3200)	0.4
		Right	**3148 (2910 - 3587)**	3123 (2686 - 3472)	0.3
	LC	Total	**1553 (1377 - 1729)**	1551 (1310 - 1659)	0.5
		Left	**1516 (1310 - 1631)**	1354 (1154 - 1634)	0.1
		Right	**1680 (1365 - 1853)**	1627 (1361 - 1836)	0.9
R2* MIP	SNpc	Total	**4045 (3689 - 4406)**	3849 (3700 - 4066)	0.2
		Left	**4034 (3619 - 4535)**	3720 (3504 - 4163)	0.1
		Right	**4049 (3678 - 4330)**	3978 (3619 - 4205)	0.6
	LC	Total	2532 (2360 - 2730)	**2567 (2389 - 2821)**	0.7
		Left	**2560 (2365 - 2754)**	2552 (2326 - 2759)	0.8
		Right	2538 (2345 - 2701)	**2564 (2386 - 2844)**	0.6
CROWN R2*	SNpc	Total	**2506 (2384 - 2699)**	2434 (2189 - 2661)	0.2
		Left	**2459 (2296 - 2670)**	2303 (2169 - 2750)	0.3
		Right	**2571 (2378 - 2818)**	2473 (2283 - 2583)	0.2
	LC	Total	1482 (1365 - 1706)	**1562 (1336 - 1767)**	0.7
		Left	1419 (1266 - 1735)	**1476 (1168 - 1641)**	1.0
		Right	1592 (1387 - 1763)	**1647 (1333 - 1877)**	0.6
CROWN R2* A2	SNpc	Total	**3002 (2687 - 3377)**	2766 (2442 - 3200)	0.1
		Left	**2935 (2650 - 3236)**	2627 (2485 - 3408)	0.2
		Right	**3085 (2706 - 3456)**	2825 (2611 - 3057)	0.2
	LC	Total	1512 (1346 - 1859)	**1617 (1274 - 1904)**	0.9
		Left	1454 (1288 - 1872)	**1562 (1153 - 1831)**	0.8
		Right	1637 (1357 - 1853)	**1758 (1301 - 2017)**	0.6

Table A.2: Mean intensities of the SNpc and the LC in the T2, the T2* MIP, the CROWN T2*, the CROWN T2* A2, the SWI, the SWI mIP, and the tSWI for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.*

Image/map	Structure		Median ($Q_1 - Q_3$)		p-value
			PD	HC	
T2*	SNpc	Total	3581 (3177 - 3922)	**3711 (3404 - 3913)**	0.5
		Left	3638 (3207 - 4063)	**3731 (3535 - 3924)**	0.1
		Right	**3455 (3085 - 3824)**	3420 (3182 - 4094)	0.6
	LC	Total	7662 (6622 - 9301)	**7752 (7125 - 9482)**	0.4
		Left	8067 (6916 - 9678)	**9076 (7542 - 10782)**	0.2
		Right	6585 (5816 - 9027)	**6800 (6329 - 8676)**	1.0
T2* MIP	SNpc	Total	2598 (2453 - 2851)	**2734 (2635 - 2895)**	0.2
		Left	2641 (2415 - 2935)	**2840 (2557 - 2930)**	0.1
		Right	2629 (2398 - 2896)	**2638 (2517 - 2911)**	0.6
	LC	Total	**4013 (3765 - 4338)**	3951 (3681 - 4241)	0.6
		Left	**4029 (3742 - 4289)**	3960 (3755 - 4331)	0.7
		Right	**4041 (3749 - 4345)**	3943 (3573 - 4252)	0.4
CROWN T2*	SNpc	Total	4270 (4032 - 4560)	**4372 (4041 - 4768)**	0.3
		Left	4365 (4005 - 4745)	**4593 (4012 - 4795)**	0.6
		Right	4143 (3785 - 4396)	**4157 (4054 - 4615)**	0.2
	LC	Total	**7603 (6107 - 8676)**	7114 (6333 - 8590)	0.9
		Left	**8195 (6064 - 9180)**	7496 (6844 - 10569)	1.0
		Right	**6531 (5929 - 8199)**	6366 (5551 - 8388)	0.7
CROWN T2* A2	SNpc	Total	3679 (3467 - 4165)	**4008 (3508 - 4462)**	0.2
		Left	3855 (3466 - 4260)	**4140 (3376 - 4567)**	0.5
		Right	3559 (3277 - 3992)	**3859 (3473 - 4096)**	0.2
	LC	Total	**7318 (5916 - 8529)**	6797 (5541 - 8812)	0.8
		Left	7850 (5792 - 9121)	**7144 (5880 - 10361)**	0.9
		Right	**6728 (5762 - 8120)**	6286 (5229 - 8635)	0.7
SWI	SNpc	Total	154 (141 - 171)	**169 (146 - 184)**	0.1
		Left	159 (145 - 179)	**173 (157 - 180)**	0.2
		Right	150 (135 - 168)	**168 (137 - 183)**	0.1
	LC	Total	229 (214 - 237)	**230 (215 - 236)**	0.9
		Left	228 (211 - 238)	**231 (214 - 237)**	0.9
		Right	228 (215 - 238)	**229 (206 - 237)**	0.8
SWI mIP	SNpc	Total	120 (107 - 129)	**121 (108 - 132)**	0.5
		Left	123 (108 - 131)	**126 (115 - 134)**	0.3
		Right	118 (106 - 130)	**119 (100 - 131)**	0.9
	LC	Total	**164 (154 - 177)**	161 (153 - 166)	0.2
		Left	**162 (156 - 176)**	160 (153 - 169)	0.3
		Right	**164 (154 - 177)**	159 (151 - 165)	0.3
tSWI	SNpc	Total	128 (106 - 143)	**138 (127 - 141)**	0.2
		Left	133 (115 - 147)	**139 (117 - 147)**	0.5
		Right	118 (103 - 142)	**133 (115 - 145)**	0.2
	LC	Total	227 (213 - 241)	**229 (214 - 241)**	0.9
		Left	223 (216 - 240)	**229 (215 - 242)**	0.6
		Right	226 (212 - 242)	**230 (210 - 241)**	0.8

Table A.3: Mean intensities of the SNpc and the LC in the tSWI mIP, the tSWI HPF, the meiSWIM, the meiSWIM filled, the meiSWIM filled MIP, the meiSWIM HPF, and the meiSWIM HPF filled for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.

Image/map	**Structure**		**Median ($Q_1 - Q_3$)**		**p-value**
			PD	**HC**	
tSWI mIP	SNpc	Total	96 (78 - 112)	**104 (86 - 116)**	0.2
		Left	99 (75 - 116)	**108 (90 - 117)**	0.3
		Right	90 (74 - 111)	**98 (87 - 111)**	0.3
	LC	Total	**166 (158 - 177)**	160 (148 - 168)	0.1
		Left	**167 (159 - 177)**	159 (148 - 170)	0.1
		Right	**166 (155 - 177)**	161 (149 - 170)	0.1
tSWI HPF	SNpc	Total	154 (143 - 164)	**161 (152 - 177)**	0.1
		Left	159 (144 - 171)	**165 (153 - 176)**	0.4
		Right	148 (138 - 164)	**159 (139 - 176)**	0.3
	LC	Total	229 (214 - 238)	229 (214 - 240)	1.0
		Left	227 (213 - 237)	**232 (214 - 239)**	0.6
		Right	**229 (213 - 238)**	226 (204 - 240)	0.8
meiSWIM	SNpc	Total	**80 (49 - 109)**	77 (59 - 89)	0.7
		Left	**83 (43 - 106)**	72 (50 - 101)	0.6
		Right	78 (53 - 113)	**86 (66 - 95)**	0.8
	LC	Total	**-2 (-13 - 9)**	-10 (-17 - 4)	0.2
		Left	**-4 (-15 - 8)**	-8 (-21 - 3)	0.2
		Right	**-4 (-12 - 7)**	-9 (-17 - 2)	0.2
meiSWIM filled	SNpc	Total	**82 (56 - 109)**	77 (67 - 89)	0.6
		Left	**84 (47 - 107)**	72 (54 - 101)	0.6
		Right	78 (62 - 113)	**86 (66 - 95)**	0.8
	LC	Total	**-2 (-12 - 9)**	-10 (-17 - 4)	0.1
		Left	**-4 (-15 - 8)**	-8 (-21 - 3)	0.2
		Right	**-4 (-12 - 7)**	-9 (-17 - 2)	0.2
meiSWIM filled MIP	SNpc	Total	**127 (102 - 160)**	112 (104 - 131)	0.3
		Left	**121 (98 - 159)**	116 (94 - 128)	0.2
		Right	**129 (101 - 153)**	118 (96 - 134)	0.4
	LC	Total	31 (24 - 42)	**33 (29 - 37)**	1.0
		Left	30 (24 - 41)	**32 (25 - 38)**	0.9
		Right	33 (25 - 42)	**34 (25 - 40)**	0.8
meiSWIM HPF	SNpc	Total	**24 (8 - 37)**	20 (15 - 29)	0.8
		Left	**19 (4 - 39)**	18 (9 - 24)	0.7
		Right	**22 (9 - 37)**	19 (13 - 36)	1.0
	LC	Total	**4 (0 - 10)**	2 (-2 - 7)	0.2
		Left	**5 (1 - 9)**	3 (-5 - 6)	0.1
		Right	3 (0 - 7)	3 (-1 - 7)	0.6
meiSWIM HPF filled	SNpc	Total	**24 (9 - 37)**	20 (15 - 29)	0.7
		Left	**19 (4 - 39)**	18 (10 - 23)	0.7
		Right	**23 (10 - 37)**	19 (13 - 36)	0.9
	LC	Total	**4 (0 - 10)**	2 (-2 - 7)	0.2
		Left	**5 (1 - 9)**	3 (-5 - 6)	0.1
		Right	**3 (0 - 7)**	3 (-1 - 7)	0.6

Table A.4: Mean intensities of the SNpc and the LC in the HPF, the mpSWIM, the pSWIM, the PrD, the CROWN PrD, the true PrD, and the CROWN true PrD for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.

Image/map	Structure		Median $(Q_1 - Q_3)$		p-value
			PD	HC	
HPF	SNpc	Total	**2072 (2037 - 2092)**	2048 (2031 - 2078)	0.2
		Left	**2056 (2034 - 2086)**	2054 (2025 - 2080)	0.6
		Right	**2083 (2059 - 2112)**	2066 (2030 - 2101)	0.3
	LC	Total	**2057 (2047 - 2068)**	2056 (2050 - 2067)	0.9
		Left	**2060 (2046 - 2072)**	2059 (2048 - 2070)	1.0
		Right	2057 (2042 - 2066)	**2063 (2045 - 2069)**	0.5
mpSWIM	SNpc	Total	**2245 (2196 - 2263)**	2234 (2208 - 2254)	0.6
		Left	**2239 (2197 - 2267)**	2236 (2209 - 2256)	0.7
		Right	**2240 (2201 - 2269)**	2234 (2216 - 2254)	0.6
	LC	Total	2135 (2124 - 2152)	**2138 (2124 - 2151)**	0.9
		Left	2133 (2120 - 2156)	**2142 (2118 - 2148)**	1.0
		Right	2139 (2123 - 2155)	**2146 (2129 - 2158)**	0.6
pSWIM	SNpc	Total	**2249 (2220 - 2277)**	2238 (2227 - 2257)	0.5
		Left	**2249 (2213 - 2275)**	2245 (2218 - 2266)	0.6
		Right	**2244 (2220 - 2282)**	2241 (2226 - 2256)	0.7
	LC	Total	2137 (2124 - 2155)	**2144 (2125 - 2155)**	0.4
		Left	2133 (2121 - 2157)	**2145 (2127 - 2155)**	0.3
		Right	2139 (2123 - 2155)	**2146 (2129 - 2158)**	0.5
PrD	SNpc	Total	617 (597 - 635)	**646 (594 - 667)**	0.1
		Left	622 (599 - 639)	**647 (596 - 670)**	0.2
		Right	611 (584 - 635)	**645 (596 - 664)**	0.1
	LC	Total	**720 (671 - 750)**	706 (664 - 748)	0.7
		Left	**735 (668 - 777)**	704 (683 - 806)	1.0
		Right	**690 (643 - 737)**	683 (643 - 735)	0.8
CROWN PrD	SNpc	Total	651 (632-676)	**669 (637 - 682)**	0.4
		Left	660 (636-688)	**673 (635 - 688)**	0.9
		Right	638 (621-669)	**669 (635 - 677)**	0.2
	LC	Total	**720 (681 - 751)**	703 (684 - 737)	0.7
		Left	**744 (684 - 767)**	728 (683 - 788)	1.0
		Right	685 (658 - 729)	**691 (658 - 730)**	1.0
True PrD	SNpc	Total	803 (757 - 834)	**814 (752 - 846)**	0.8
		Left	810 (763 - 843)	**812 (759 - 846)**	0.9
		Right	802 (732 - 833)	**817 (748 - 839)**	0.6
	LC	Total	**806 (766 - 835)**	793 (753 - 828)	0.6
		Left	**817 (773 - 853)**	793 (755 - 862)	0.6
		Right	792 (742 - 831)	**802 (743 - 824)**	0.9
CROWN true PrD	SNpc	Total	783 (755 - 803)	**785 (760 - 796)**	0.9
		Left	**788 (765 - 814)**	786 (768 - 796)	0.6
		Right	780 (742 - 797)	**785 (752 - 791)**	0.6
	LC	Total	**800 (783 - 824)**	793 (781 - 815)	0.6
		Left	**818 (780 - 839)**	817 (773 - 843)	0.9
		Right	**783 (758 - 809)**	777 (760 - 815)	0.9

Table A.5: Mean intensities of the SNpc and the LC in the CROWN true PrD A, the T1WE, the T1, the simGRE, the simGREa, the simFLAIR, and the simDIR GM for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.

Image/map	Structure		Median ($Q_1 - Q_3$)		p-value
			PD	HC	
CROWN true PrD A	SNpc	Total	780 (755 - 804)	780 (763 - 802)	0.8
		Left	**787 (761 - 818)**	786 (764 - 801)	0.4
		Right	773 (749 - 799)	**777 (749 - 801)**	0.7
	LC	Total	**802 (783 - 821)**	786 (780 - 817)	0.5
		Left	**818 (779 - 837)**	811 (773 - 842)	0.8
		Right	782 (760 - 809)	**784 (758 - 811)**	1.0
T1WE	SNpc	Total	422 (381 - 469)	**427 (368 - 441)**	0.7
		Left	419 (378 - 462)	**422 (363 - 445)**	0.7
		Right	426 (384 - 475)	**429 (372 - 438)**	0.6
	LC	Total	**395 (358 - 441)**	387 (354 - 412)	0.6
		Left	**397 (349 - 441)**	386 (354 - 413)	0.5
		Right	**400 (351 - 444)**	396 (355 - 432)	0.6
T1	SNpc	Total	**1095 (1031 - 1151)**	1081 (1052 - 1123)	0.9
		Left	**1124 (1049 - 1203)**	1112 (1057 - 1148)	0.6
		Right	1025 (1001 - 1112)	**1061 (990 - 1129)**	0.5
	LC	Total	1480 (1277 - 1754)	**1515 (1406 - 1671)**	0.8
		Left	1643 (1292 - 1870)	**1680 (1345 - 2096)**	0.7
		Right	1240 (1171 - 1609)	**1334 (1185 - 1593)**	0.8
simGRE	SNpc	Total	555 (542 - 573)	**564 (551 - 578)**	0.4
		Left	558 (545 - 575)	**562 (554 - 570)**	0.9
		Right	552 (538 - 572)	**564 (556 - 583)**	0.1
	LC	Total	**556 (520 - 565)**	531 (491 - 550)	0.02*
		Left	**554 (512 - 567)**	527 (474 - 550)	0.03*
		Right	**555 (538 - 566)**	541 (516 - 548)	0.008**
simGREa	SNpc	Total	**562 (547 - 571)**	560 (547 - 574)	0.8
		Left	**557 (543 - 570)**	554 (543 - 573)	0.9
		Right	**567 (552 - 576)**	565 (539 - 577)	0.7
	LC	Total	531 (496 - 563)	**556 (512 - 571)**	0.2
		Left	529 (478 - 563)	**559 (505 - 569)**	0.3
		Right	548 (520 - 562)	**559 (539 - 572)**	0.1
simFLAIR	SNpc	Total	457 (445 - 468)	**460 (443 - 474)**	0.8
		Left	**459 (444 - 472)**	456 (438 - 480)	0.9
		Right	454 (443 - 470)	**462 (446 - 468)**	0.4
	LC	Total	**492 (474 - 512)**	481 (456 - 500)	0.2
		Left	**491 (461 - 509)**	468 (439 - 499)	0.2
		Right	**502 (481 - 517)**	493 (483 - 506)	0.3
simDIR GM	SNpc	Total	69 (55 - 96)	**85 (65 - 122)**	0.2
		Left	80 (66 - 97)	**95 (72 - 116)**	0.2
		Right	66 (47 - 95)	**78 (55 - 123)**	0.2
	LC	Total	**120 (98 - 129)**	117 (102 - 126)	0.9
		Left	**117 (101 - 132)**	114 (99 - 125)	0.6
		Right	120 (96 - 131)	**125 (102 - 135)**	0.4

** Significant difference between PD and HC ($p < 0.05$).*

*** Highly significant difference between PD and HC ($p < 0.01$).*

Table A.6: Mean intensities of the SNpc and the LC in the simDIR WM, the simDIR CSF, the dSWI, the dSWI mIP, the MRA, and the k for PD and HC. The p-values were determined using the Mann-Whitney U test. The highest median for each structure in each image is highlighted in boldface.

Image/map	Structure		Median $(Q_1 - Q_3)$		p-value
			PD	HC	
simDIR WM	SNpc	Total	**229 (192 - 275)**	223 (173 - 254)	0.7
		Left	**215 (155 - 258)**	212 (172 - 239)	1.0
		Right	**248 (202 - 295)**	246 (168 - 288)	0.6
	LC	Total	105 (78 - 142)	**114 (99 - 134)**	0.9
		Left	77 (51 - 138)	**83 (58 - 121)**	1.0
		Right	122 (86 - 166)	**128 (93 - 155)**	0.9
simDIR CSF	SNpc	Total	**16 (3 - 52)**	2 (0 - 13)	0.02*
		Left	**6 (1 - 78)**	2 (0 - 11)	0.1
		Right	**2 (0 - 13)**	0 (0 - 2)	0.1
	LC	Total	103 (31 - 208)	**113 (66 - 229)**	0.7
		Left	135 (19 - 276)	**141 (45 - 362)**	0.6
		Right	7 (1 - 174)	**31 (5 - 184)**	0.5
dSWI	SNpc	Total	163 (151 - 174)	**173 (159 - 183)**	0.1
		Left	162 (151 - 179)	**172 (161 - 183)**	0.2
		Right	159 (146 - 175)	**168 (156 - 185)**	0.1
	LC	Total	**231 (217 - 241)**	230 (216 - 242)	0.8
		Left	229 (219 - 246)	**235 (217 - 240)**	1.0
		Right	**232 (216 - 244)**	229 (212 - 243)	0.7
dSWI mIP	SNpc	Total	115 (101 - 134)	**126 (106 - 138)**	0.3
		Left	116 (98 - 136)	**128 (109 - 140)**	0.2
		Right	116 (102 - 133)	**126 (107 - 133)**	0.4
	LC	Total	**172 (161 - 182)**	166 (156 - 179)	0.3
		Left	170 (161 - 181)	170 (157 - 185)	0.7
		Right	**173 (160 - 182)**	167 (155 - 174)	0.2
MRA	SNpc	Total	**376 (301 - 422)**	372 (301 - 433)	1.0
		Left	336 (283 - 455)	**404 (300 - 457)**	0.5
		Right	**323 (288 - 378)**	320 (288 - 358)	0.7
	LC	Total	352 (334 - 369)	**353 (339 - 366)**	0.8
		Left	351 (332 - 369)	**352 (343 - 360)**	0.9
		Right	**350 (331 - 368)**	348 (330 - 369)	0.9
k	SNpc	Total	**1256 (1228 - 1280)**	1240 (1196 - 1283)	0.2
		Left	**1255 (1228 - 1290)**	1234 (1191 - 1283)	0.2
		Right	**1258 (1231 - 1284)**	1241 (1201 - 1273)	0.1
	LC	Total	**1216 (1178 - 1251)**	1191 (1153 - 1235)	0.1
		Left	**1222 (1184 - 1257)**	1196 (1157 - 1230)	0.1
		Right	**1210 (1172 - 1246)**	1185 (1150 - 1235)	0.2

** Significant difference between PD and HC ($p < 0.05$).*

Appendix B

Mean of SNpc and LC for PIGD and TD

Tables B.1-B.6 show the mean intensities of the SN and the LC for PIGD and TD in all images and maps generated in this thesis. The p-values were determined using the Student's unpaired *t*-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations.

Table B.1: Mean intensities of the SNpc and the LC in the R2, the R2* MIP, and the CROWN R2* for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.*

Image/map	**Structure**		**Median ($Q_1 - Q_3$) / Mean±SD**		**p-value**
			PIGD	**TD**	
R2*	SNpc	Total	**3104 (2907 - 3542)**	3059 (2829 - 3374)	0.5
		Left	**3184 (2777 - 3463)**	3003 (2789 - 3366)	0.6
		Right	**3260 (2932 - 3692)**	3097 (2946 - 3518)	0.8
	LC	Total	1416 (1363 - 1694)	**1575 (1401 - 1798)**	0.4
		Left	1520±280	**1534±242**	0.9
		Right	1573±306	**1675±335**	0.3
R2* MIP	SNpc	Total	**4073 (3747 - 4404)**	3991 (3540 - 4319)	0.7
		Left	**4000 (3647 - 4441)**	3766 (3580 - 4449)	0.6
		Right	**4053 (3735 - 4278)**	4044 (3576 - 4368)	0.9
	LC	Total	2506 (2330 - 2679)	**2599 (2432 - 2929)**	0.1
		Left	2502 (2351 - 2740)	**2575 (2483 - 2878)**	0.1
		Right	2474 (2341 - 2655)	**2608 (2363 - 2720)**	0.3
CROWN R2*	SNpc	Total	**2574 (2446 - 2796)**	2474 (2280 - 2648)	0.2
		Left	**2469 (2308 - 2752)**	2408 (2209 - 2608)	0.3
		Right	**2644 (2405 - 2914)**	2517 (2328 - 2738)	0.2
	LC	Total	1551±256	**1555±254**	1.0
		Left	**1515±317**	1491±261	0.8
		Right	1586±315	**1619±301**	0.7

Table B.2: Mean intensities of the SNpc and the LC in the CROWN R2 A2, the T2*, the T2* MIP, the CROWN T2*, the CROWN T2* A2, the SWI, and the SWI mIP for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.*

Image/map	Structure		Median $(Q_1 - Q_3)$ / Mean±SD		p-value
			PIGD	TD	
CROWN R2* A2	SNpc	Total	**3038 (2683 - 3419)**	2966 (2586 - 3336)	0.4
		Left	**2926 (2661 - 3236)**	2924 (2535 - 3216)	0.4
		Right	**3163 (2782 - 3473)**	3090 (2653 - 3408)	0.6
	LC	Total	**1647±355**	1644±357	1.0
		Left	1445 (1353 - 2029)	**1708 (1278 - 1849)**	0.5
		Right	1679±416	**1699±420**	0.9
T2*	SNpc	Total	**3652±597**	3508±629	0.4
		Left	**3720±632**	3608±717	0.6
		Right	**3461 (3068 - 3861)**	3455 (3112 - 3699)	0.8
	LC	Total	**8095±1742**	7545±1562	0.3
		Left	**8379±2008**	8007±1696	0.5
		Right	**6641 (5856 - 9631)**	6611 (5649 - 8630)	0.3
T2* MIP	SNpc	Total	**2675±345**	2606±409	0.5
		Left	2645±395	**2662±429**	0.9
		Right	**2632 (2447 - 2885)**	2561 (2383 - 2915)	0.5
	LC	Total	**4109 (3792 - 4413)**	3895 (3644 - 4180)	0.1
		Left	**4053 (3727 - 4380)**	3942 (3710 - 4149)	0.2
		Right	**4066 (3819 - 4401)**	3916 (3733 - 4271)	0.3
CROWN T2*	SNpc	Total	4207 (4015 - 4465)	**4274 (4004 - 4535)**	0.9
		Left	4353 (4083 - 4703)	**4364 (3959 - 4816)**	1.0
		Right	3997 (3695 - 4292)	**4117 (3825 - 4436)**	0.4
	LC	Total	**7663±1685**	7389±1665	0.6
		Left	**7950±2198**	7865±1858	0.9
		Right	**6465 (5972 - 8369)**	6405 (5804 - 7893)	0.6
CROWN T2* A2	SNpc	Total	**3679 (3445 - 4181)**	3642 (3443 - 4062)	1.0
		Left	**3874±637**	3792±671	0.7
		Right	3520±567	**3539±548**	0.9
	LC	Total	**7411±1806**	7194±1895	0.7
		Left	**7615±2241**	7530±1978	0.9
		Right	**6699 (5677 - 8285)**	6175 (5680 - 7798)	0.8
SWI	SNpc	Total	152±20	**156±25**	0.5
		Left	158±21	**163±28**	0.5
		Right	147±23	**151±24**	0.6
	LC	Total	224±17	**225±21**	0.8
		Left	224±21	**225±21**	1.0
		Right	224±18	**226±25**	0.7
SWI mIP	SNpc	Total	116±18	**118±17**	0.8
		Left	117±20	**121±17**	0.5
		Right	**116±20**	115±19	0.9
	LC	Total	**165±17**	158±20	0.3
		Left	**166±19**	159±20	0.2
		Right	160 (154 - 172)	**165 (153 - 174)**	1.0

Table B.3: Mean intensities of the SNpc and the LC in the tSWI, the tSWI mIP, the tSWI HPF, the meiSWIM, the meiSWIM filled, the meiSWIM filled MIP, and the meiSWIM HPF for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.

Image/map	**Structure**		**Median ($Q_1 - Q_3$) / Mean±SD**		**p-value**
			PIGD	**TD**	
tSWI	SNpc	Total	126±27	126±33	0.9
		Left	131±31	131±38	1.0
		Right	120±28	**123±32**	0.8
	LC	Total	224±17	224±24	1.0
		Left	224±19	224±25	0.9
		Right	224±19	**225±25**	0.9
tSWI mIP	SNpc	Total	**97±28**	90±30	0.5
		Left	**107 (86 - 116)**	86 (72 - 120)	0.4
		Right	**95±30**	88±28	0.4
	LC	Total	**166±16**	160±21	0.3
		Left	**167 (158 - 175)**	162 (156 - 175)	0.5
		Right	166 (158 - 174)	**167 (154 - 174)**	0.6
tSWI HPF	SNpc	Total	151±21	**155±26**	0.5
		Left	156±22	**160±30**	0.6
		Right	146±23	**151±25**	0.5
	LC	Total	224±16	**225±22**	1.0
		Left	224±18	224±22	0.9
		Right	224±18	**225±24**	0.9
meiSWIM	SNpc	Total	80 (60 - 100)	**81 (48 - 112)**	0.9
		Left	**83 (44 - 91)**	75 (47 - 109)	0.9
		Right	**88±48**	82±46	0.7
	LC	Total	**-2±11**	-5±19	0.5
		Left	**-3±13**	-6±20	0.6
		Right	**-2±13**	-5±19	0.5
meiSWIM filled	SNpc	Total	80 (60 - 100)	**81 (54 - 112)**	0.9
		Left	**83 (46 - 93)**	75 (51 - 109)	0.8
		Right	**88±47**	85±43	0.8
	LC	Total	**-2±11**	-5±19	0.6
		Left	**-3±13**	-5±20	0.6
		Right	**-2±13**	-4±20	0.6
meiSWIM filled MIP	SNpc	Total	117 (99 - 147)	**133 (102 - 161)**	0.3
		Left	115 (95 - 139)	**129 (97 - 164)**	0.3
		Right	129 (101 - 142)	**130 (110 - 156)**	0.5
	LC	Total	**34 (27 - 41)**	31 (27 - 44)	0.9
		Left	**32 (24 - 41)**	31 (26 - 38)	0.7
		Right	34 (26 - 40)	34 (27 - 42)	0.7
meiSWIM HPF	SNpc	Total	**25 (11 - 36)**	21 (7 - 37)	0.4
		Left	**21 (8 - 38)**	14 (1 - 38)	0.4
		Right	**24 (14 - 41)**	19 (7 - 36)	0.3
	LC	Total	**4±5**	3±7	0.6
		Left	**5±6**	4±8	0.7
		Right	**4±6**	3±8	0.5

Table B.4: Mean intensities of the SNpc and the LC in the meiSWIM HPD filled, the HPF, the mpSWIM, the pSWIM, the PrD, the CROWN PrD, and the true PrD for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.

Image/map	Structure		Median ($Q_1 - Q_3$) / Mean±SD		p-value
			PIGD	TD	
meiSWIM HPF filled	SNpc	Total	**26 (10 - 36)**	21 (11 - 37)	0.5
		Left	**21 (7 - 38)**	14 (3 - 38)	0.5
		Right	**30±26**	24±23	0.4
	LC	Total	**4±5**	3±8	0.5
		Left	**5±6**	4±8	0.7
		Right	**4±6**	3±8	0.5
HPF	SNpc	Total	**2070 (2035 - 2088)**	2066 (2049 - 2091)	0.8
		Left	2048 (2040 - 2079)	**2052 (2010 - 2089)**	0.7
		Right	2077±62	**2089±56**	0.5
	LC	Total	**2058 (2043 - 2068)**	2055 (2048 - 2064)	0.6
		Left	2058 (2041 - 2074)	**2062 (2050 - 2068)**	0.9
		Right	**2057±22**	2051±22	0.3
mpSWIM	SNpc	Total	**2249 (2196 - 2258)**	2236 (2201 - 2264)	1.0
		Left	**2240 (2185 - 2262)**	2221 (2205 - 2258)	0.9
		Right	**2241 (2202 - 2260)**	2239 (2202 - 2274)	0.7
	LC	Total	**2140 (2122 - 2159)**	2133 (2125 - 2145)	0.6
		Left	2133 (2117 - 2159)	**2137 (2124 - 2144)**	0.8
		Right	**2146 (2129 - 2161)**	2129 (2120 - 2143)	0.1
pSWIM	SNpc	Total	**2252 (2228 - 2282)**	2244 (2212 - 2275)	0.3
		Left	**2253 (2225 - 2273)**	2228 (2208 - 2258)	0.2
		Right	**2245 (2226 - 2271)**	2242 (2214 - 2287)	1.0
	LC	Total	**2141 (2122 - 2159)**	2133 (2125 - 2145)	0.6
		Left	2133 (2119 - 2159)	**2137 (2124 - 2144)**	0.9
		Right	**2146 (2129 - 2160)**	2129 (2120 - 2143)	0.1
PrD	SNpc	Total	616 (589 - 624)	**617 (599 - 642)**	0.3
		Left	613±35	**622±36**	0.4
		Right	596±39	**614±33**	0.1
	LC	Total	708±62	**712±62**	0.9
		Left	722±81	**734±71**	0.6
		Right	**694±78**	689±66	0.8
CROWN PrD	SNpc	Total	646±22	**660±31**	0.1
		Left	657±28	**666±32**	0.3
		Right	635±29	**654±32**	0.04*
	LC	Total	714±51	**717±50**	0.9
		Left	727±68	**738±60**	0.6
		Right	684 (657 - 725)	684 (655 - 735)	0.9
true PrD	SNpc	Total	781±50	**809±68**	0.1
		Left	782 (747 - 838)	**804 (765 - 852)**	0.5
		Right	767±62	**804±70**	0.1
	LC	Total	791±54	**798±60**	0.6
		Left	802±71	**813±66**	0.6
		Right	780±63	**784±66**	0.8

** Significant difference between PIGD and TD ($p < 0.05$).*

Table B.5: Mean intensities of the SNpc and the LC in the CROWN true PrD, the CROWN true PrD A, the T1WE, the T1, the simGRE, the simGREa, and the simFLAIR for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.

Image/map	Structure		Median $(Q_1 - Q_3)$ / Mean±SD		p-value
			PIGD	TD	
CROWN true PrD	SNpc	Total	773±29	**788±34**	0.1
		Left	786±39	**796±37**	0.4
		Right	761±34	**781±36**	0.1
	LC	Total	798±37	**800±38**	0.8
		Left	806±51	**814±46**	0.6
		Right	774 (756 - 803)	**779 (758 - 810)**	1.0
CROWN true PrD A	SNpc	Total	772±26	**787±34**	0.1
		Left	785±38	**795±39**	0.4
		Right	760±30	**781±35**	0.03*
	LC	Total	796±38	**801±38**	0.7
		Left	805±52	**815±47**	0.5
		Right	772 (754 - 800)	**780 (759 - 812)**	0.8
T1WE	SNpc	Total	422 (395 - 499)	**435 (389 - 473)**	0.6
		Left	**439±70**	419±61	0.3
		Right	423 (400 - 507)	**439 (387 - 477)**	0.6
	LC	Total	392 (367 - 474)	**417 (346 - 443)**	0.6
		Left	390 (361 - 469)	**411 (346 - 445)**	0.6
		Right	**420±71**	400±61	0.3
T1	SNpc	Total	1081±78	**1097±72**	0.5
		Left	**1146±135**	1129±86	0.6
		Right	1014 (990 - 1070)	**1092 (1007 - 1125)**	0.1
	LC	Total	**1564±361**	1515±333	0.6
		Left	1638±463	**1655±424**	0.9
		Right	**1236 (1139 - 1644)**	1234 (1165 - 1495)	1.0
simGRE	SNpc	Total	**558±21**	554±23	0.5
		Left	**559±19**	558±24	0.9
		Right	**557±28**	549±25	0.3
	LC	Total	556 (518 - 562)	**560 (541 - 571)**	0.1
		Left	558 (512 - 565)	**559 (544 - 568)**	0.3
		Right	555 (544 - 562)	**559 (545 - 571)**	0.2
simGREa	SNpc	Total	553±22	**566±17**	0.03*
		Left	550±22	**563±17**	0.03*
		Right	556±28	**568±18**	0.1
	LC	Total	**540 (486 - 556)**	523 (496 - 547)	0.5
		Left	**530 (495 - 559)**	513 (469 - 541)	0.3
		Right	**548 (520 - 557)**	542 (522 - 559)	1.0
simFLAIR	SNpc	Total	455±19	455±19	1.0
		Left	456±19	**458±21**	0.7
		Right	**454±20**	453±19	0.8
	LC	Total	487±28	**489±31**	0.8
		Left	**482±33**	480±38	0.8
		Right	498 (481 - 515)	**499 (485 - 519)**	0.6

** Significant difference between PIGD and TD ($p < 0.05$).*

Table B.6: Mean intensities of the SNpc and the LC in the simDIR GM, the simDIR WM, the simDIR CSF, the dSWI, the dSWI mIP, the MRA, the HPF, and the k for PIGD and TD. The p-values were calculated using the Student's unpaired t-test for populations with normal distribution and the Mann-Whitney U test for non-normally distributed populations. The highest median/mean for each structure in each image is highlighted in boldface.

Image/map	Structure		Median ($Q_1 - Q_3$) / Mean±SD		p-value
			PIGD	TD	
simDIR GM	SNpc	Total	66±21	**89±30**	0.004**
		Left	74±26	**95±33**	0.02*
		Right	58±25	**83±30**	0.003**
	LC	Total	105±29	**127±25**	0.008**
		Left	105±31	**126±26**	0.02*
		Right	105±29	**128±25**	0.006**
simDIR WM	SNpc	Total	**250±59**	217±60	0.1
		Left	**223±75**	202±60	0.3
		Right	**274±65**	231±69	0.04*
	LC	Total	**120 (92 - 145)**	92 (72 - 143)	0.1
		Left	**93 (62 - 139)**	64 (43 - 139)	0.1
		Right	**142±59**	127±63	0.4
simDIR CSF	SNpc	Total	**34 (5 - 85)**	8 (2 - 31)	0.03*
		Left	**23 (1 - 106)**	4 (1 - 56)	0.2
		Right	2 (0 - 39)	2 (0 - 7)	0.3
	LC	Total	**184 (9 - 283)**	75 (38 - 180)	0.3
		Left	109 (13 - 339)	**124 (23 - 257)**	0.9
		Right	**7 (2 - 190)**	3 (0 - 87)	0.3
dSWI	SNpc	Total	158±17	**164±24**	0.3
		Left	161±17	**166±27**	0.4
		Right	155 (141 - 167)	**163 (154 - 186)**	0.2
	LC	Total	228±16	228±22	0.9
		Left	**229±19**	228±21	0.9
		Right	228±18	228±25	1.0
dSWI mIP	SNpc	Total	109±30	**117±24**	0.3
		Left	**115 (94 - 131)**	114 (106 - 137)	0.4
		Right	110±30	**116±25**	0.4
	LC	Total	168 (160 - 182)	**169 (159 - 176)**	0.6
		Left	**168 (158 - 179)**	166 (160 - 179)	0.8
		Right	**174 (160 - 182)**	168 (156 - 180)	0.3
MRA	SNpc	Total	369±88	**374±83**	0.9
		Left	329 (278 - 437)	**338 (283 - 453)**	0.9
		Right	325 (293 - 420)	**341 (299 - 427)**	0.6
	LC	Total	346 (326 - 363)	**355 (349 - 379)**	0.1
		Left	345 (328 - 357)	**358 (347 - 381)**	0.1
		Right	347 (322 - 362)	**356 (344 - 377)**	0.2
k	SNpc	Total	**1273±50**	1240±46	0.03*
		Left	**1270±50**	1238±46	0.03*
		Right	**1275±51**	1242±47	0.03*
	LC	Total	**1228±52**	1193±52	0.03*
		Left	**1234±52**	1199±51	0.03*
		Right	**1222±52**	1188±52	0.03*

** Significant difference between PIGD and TD ($p < 0.05$).*

*** Highly significant difference between PIGD and TD ($p < 0.01$).*

Bibliography

[1] Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. Nat Rev Dis Primers. 2017 Mar;3(1):1-21. doi:10.1038/nrdp.2017.13.

[2] Balestrino R, Schapira AHV. Parkinson disease. Eur J Neurol. 2020 Oct;27(1):27-42. doi:10.1111/ENE.14108.

[3] Berg D, Borghammer P, Fereshtehnejad SM, Heinzel S, Horsager J, Schaeffer E, et al. Prodromal Parkinson disease subtypes - key to understanding heterogeneity. Nat Rev Neurol. 2021 Apr;17(6):349-61. doi:10.1038/s41582-021-00486-9.

[4] Neuro-SysMed. Parkinsons sykdom (PS);. Accessed: 2022-10-17. Available from: https://neuro-sysmed.no/dis-clinic-stud/parkinsons-disease/.

[5] Goetz CG, Fahn S, Martinez-Martin P. Movement Disorder Society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan. Mov Disord. 2008 Nov;22(1):41-7. doi:10.1002/mds.22340.

[6] Rajput AH, Voll A, Rajput ML, Robinson CA, Rajput A. Course in Parkinson disease subtypes: A 39-year clinicopathologic study. Neurology. 2009 Jul;73(3):206-12. doi:10.1212/WNL.0b013e3181ae7af1.

[7] Jankovic J, McDermott M, Carter J, Gauthier S, Goetz C, Golbe L, et al. Variable expression of Parkinson's disease: A base-line analysis of the DATATOP cohort. The Parkinson Study Group. Neurology. 1990 Oct;40(10):1529-34. doi:10.1212/WNL.40.10.1529.

[8] Jellinger KA. The pathomechanisms underlying Parkinson's disease. Expert Rev Neurother. 2014 Jan;14(2):199-215. doi:10.1586/14737175.2014.877842.

[9] Fearnley JM, Lees AJ. Ageing and parkinson's disease: Substantia nigra regional selectivity. Brain. 1991 Oct;114(5):2283-301. doi:10.1093/BRAIN/114.5.2283.

[10] Gesi M, Soldani P, Giorgi FS, Santinami A, Bonaccorsi I, Fornai F. The role of the locus coeruleus in the development of Parkinson's disease. Neurosci Biobehav Rev. 2000 Aug;24(6):655-68. doi:10.1016/S0149-7634(00)00028-2.

[11] Zarow C, Lyness SA, Mortimer JA, Chui HC. Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. Arch Neurol. 2003 Mar;60(3):337-41. doi:10.1001/archneur.60.3.337.

[12] Schneider E, Ng KM, Yeoh CS, Rumpel H, Fook-Chong S, Li HH, et al. Susceptibility-weighted MRI of extrapyramidal brain structures in Parkinsonian disorders. Medicine (Baltimore). 2016 Jun;95(26):e3730. doi:10.1097/MD.0000000000003730.

[13] Berg D, Hochstrasser H. Iron metabolism in Parkinsonian syndromes. Mov Disord. 2006 Jun;21(9):1299-310. doi:10.1002/mds.21020.

[14] Youdim MBH, Ben-Shachar D, Riederer P. The possible role of iron in the etiopathology of parkinson's disease. Mov Disord. 1993;8(1):1-12. doi:10.1002/mds.870080102.

[15] Zecca L, Stroppolo A, Gatti A, Tampellini D, Toscani M, Gallorini M, et al. The role of iron and copper molecules in the neuronal vulnerability of locus coeruleus and substantia nigra during aging. Proc Natl Acad Sci U S A. 2004 Jun;101:9843-8. doi:10.1073/pnas.0403495101.

[16] Zucca FA, C Bellei SG, Terreni MR, Gallorini M, E Rizzio GP, Albertini A, et al. Neuromelanin and iron in human locus coeruleus and substantia nigra during aging: consequences for neuronal vulnerability. J Neural Transm. 2006 Apr;113:757-67. doi:10.1007/s00702-006-0453-2.

[17] Zucca FA, Segura-Aguilar J, Ferrari E, Muñoz P, Paris I, Sulzer D, et al. Interactions of iron, dopamine and neuromelanin pathways in brain aging and Parkinson's disease. Prog Neurobiol. 2017 Aug;155:96-119. doi:10.1016/j.pneurobio.2015.09.012.

[18] Zecca L, Tampellini D, Gerlach M, Riederer P, Fariello RG, Sulzer D. Substantia nigra neuromelanin: structure, synthesis, and molecular behaviour. Mol Pathol. 2001 Dec;54(6):414-8. PMCID:PMC1187132.

[19] Pizarro-Galleguillos BM, Kunert L, Brüggemann N, Prasuhn J. Iron- and Neuromelanin-Weighted Neuroimaging to Study Mitochondrial Dysfunction in Patients with Parkinson's Disease. Int J Mol Sci. 2022 Nov;23(22). doi:10.3390/ijms232213678.

[20] Hirsch E, Graybiel AM, Agid YA. Melanized dopaminergic neurons are differentially susceptible to degeneration in Parkinsons disease. Nature. 1988 Jul;334(6180):345-8. doi:10.1038/334345a0.

[21] Kastner A, Hirsch EC, Lejeune O, Javoy-Agid F, Rascol O, Agid Y. Is the vulnerability of neurons in the substantia nigra of patients with Parkinsons disease related to their neuromelanin content? J Neurochem. 1992 Sep;59(3):1080-9. doi:10.1111/j.1471-4159.1992.tb08350.x.

[22] Brooks DJ. Imaging dopamine transporters in Parkinsons disease. Biomark Med. 2010 Oct;4(5):651-60. doi:10.2217/bmm.10.86.

[23] Boonstra JT, Michielse S, Temel Y, Hoogland G, Jahanshahi A. Neuroimaging Detectable Differences between Parkinson's Disease Motor Subtypes: A Systematic Review. Mov Disord Clin Pract. 2020 Oct;8(2):175-92. doi:10.1002/mdc3.13107.

[24] Hustad E, Aasly JO. Clinical and Imaging Markers of Prodromal Parkinson's Disease. Front Neurol. 2020 May;11:395. doi:10.3389/fneur.2020.00395.

[25] Kasban H, El-Bendary MAM, Salama DH. A comparative study of medical imaging techniques. International Journal of Information Science and Intelligent System. 2015;4(2):37-58. Available from: https://ilearn.th-deg.de/pluginfile.php/480243/mod_book/chapter/8248/updated_JXIJSIS2015.pdf.

[26] Haacke EM, Cheng NYC, House MJ, Liu Q, Neelavalli J, Ogg RJ, et al. Imaging iron stores in the brain using magnetic resonance imaging. Magn Reson Imaging. 2005 Jan;23(1):1-25. doi:10.1016/j.mri.2004.10.001.

[27] Trujillo P, Summers P, Ferrari E, Zucca FA, Sturini M, Mainardi LT, et al. Contrast mechanisms associated with neuromelanin-MRI. Magn Reson Med. 2017 Nov;78(5):1790-800. doi:10.1002/mrm.26584.

[28] Feraco P, Gagliardo C, Tona GL, Bruno E, Dangelo C, Marrale M, et al. Imaging of Substantia Nigra in Parkinson's Disease: A Narrative Review. Brain Sci. 2021 Jun;11(6):769. doi:10.3390/brainsci11060769.

[29] Delenclos M, Jones DR, McLean PJ, Uitti RJ. Biomarkers in Parkinson's disease: Advances and strategies. Parkinsonism Relat Disord. 2016 Jan;22(1):106-10. doi:10.1016/j.parkreldis.2015.09.048.

[30] Mehanna R, Lai E. Deep brain stimulation in Parkinson's disease. Transl Neurodegener. 2013 Nov;2(1):22. doi:10.1186/2047-9158-2-22.

[31] Kirsch W, McAuley G, Holshouser B, Petersen F, Ayaz M, Vinters HV, et al. Serial Susceptibility Weighted MRI measures brain iron and microbleeds in dementia. J Alzheimers Dis. 2009 Jul;17(3):599-609. doi:10.3233/JAD-2009-1073.

[32] Xiang Y, Gong T, Wu J, Li J, Chen Y, Wang Y, et al. Subtypes evaluation of motor dysfunction in Parkinson's disease using neuromelanin-sensitive magnetic resonance imaging. Neurosci Lett. 2017 Jan;638:145-50. doi:10.1016/j.neulet.2016.12.036.

[33] Depierreux F, Parmentier E, Baquero KA, Balteau E, Phillips C, Maquet P, et al. Validation of new brain MRI biomarkers in Parkinson's Disease: the use of quantitative multi-parameter mapping. Front Neurosci Conference Abstract: Belgian Brain Congress 2018 - Belgian Brain Council. 2019 Jan. doi:10.3389/conf.fnins.2018.95.00028.

[34] Smith NB, Webb A. Introduction to medical imaging: Physics, engineering and clinical applications. Cambridge university press; 2011.

[35] Brown RW, Cheng YCN, Haacke EM, Thompson MR, Venkatesan R. Magnetic Resonance Imaging: Physical Principles and Sequence Design. 2nd ed. John Wiley & Sons, Inc.; 2014. Available from: https://www.home-workforyou.com/static_media/uploadedfiles/[Robert_W_Brown;_Yu-Chung_N_Cheng;_E__Mark_Haacke;(z-lib.org).pdf.

[36] Elster AD. IR-Prepped Sequences - Questions and Answers in MRI;. Accessed: 2022-12-15. Available from: https://mriquestions.com/ir-prepped-sequences.html.

[37] Hendrick RE. Glossary of MR Terms (5th ed); 2005. American College of Radiology. Accessed: 2023-05-24. Available from: https://mriquestions.com/uploads/3/4/5/7/34572113/acr_glossaryofmrterms.pdf.

[38] Mugler JP, Brookeman JR. Rapid three-dimensional T1-weighted MR imaging with the MP-RAGE sequence. J Magn Reson Imaging. 1991 Sep-Oct;1(5):561-7. doi:10.1002/jmri.1880010509.

[39] Elster AD. MP-RAGE v MP2RAGE - Questions and Answers in MRI;. Accessed: 2022-12-15. Available from: https://mriquestions.com/mp-rage-v-mr2rage.html.

[40] Elster AD. Ernst Angle - Questions and Answers in MRI;. Accessed: 2022-12-15. Available from: https://mriquestions.com/optimal-flip-angle.html.

[41] Elster AD. Dual/Multi-echo GRE - Questions and Answers in MRI;. Accessed: 2023-05-24. Available from: https://mriquestions.com/multi-echo-gre.html.

[42] Elster AD. Flow-compensation - Questions and Answers in MRI;. Accessed: 2023-05-24. Available from: https://mriquestions.com/flow-compensation.html.

[43] Chen Y, Liu S, Wang Y, Kang Y, Haacke EM. STrategically Acquired Gradient Echo (STAGE) imaging, part I: Creating enhanced T1 contrast and standardized susceptibility weighted imaging and quantitative susceptibility mapping. Magn Reson Imaging. 2018 Feb;46:130-9. doi:10.1016/J.MRI.2017.10.005.

[44] Wang Y, Chen Y, Wu D, Wang Y, Sethi SK, Yang G, et al. STrategically Acquired Gradient Echo (STAGE) imaging, part II: Correcting for RF inhomogeneities in estimating T1 and proton density. Magn Reson Imaging. 2018 Feb;46:140-50. doi:10.1016/J.MRI.2017.10.006.

[45] Buch S (SpinTech MRI). STAGE images overview;. Email to: Signe Hogstad. 2023-05-12.

[46] Wood JC, Ghugre N. Magnetic resonance imaging assessment of excess iron in thalassemia, sickle cell disease and other iron overload diseases. Hemoglobin. 2008;32(1-2):85-96. doi:10.1080/03630260701699912.

[47] Haacke EM, Ayaz M, Khan A, Manova ES, Krishnamurthy B, Gollapalli L, et al. Establishing a baseline phase behavior in magnetic resonance imaging to determine normal vs. abnormal iron content in the brain. J Magn Reson Imaging. 2007 Aug;26(2):265-4. doi:10.1002/jmri.22987.

[48] Langkammer C, Pirpamer L, Seiler S, Deistung A, Schweser F, Franthal S, et al. Quantitative Susceptibility Mapping in Parkinson's Disease. PLoS One. 2016 Sep;11(9):e0162460. doi:10.1371/journal.pone.0162460.

[49] Cheng Z, He N, Huang P, Li Y, Tang R, Sethi SK, et al. Imaging the Nigrosome 1 in the substantia nigra using susceptibility weighted imaging and quantitative susceptibility mapping: An application to Parkinson's disease. Neuroimage Clin. 2020 Nov;25:102103. doi:10.1016/j.nicl.2019.102103.

[50] Parkinson J. An essay on the shaking palsy. 1817. Journal Neuropsychiatry Clin Neurosci. 2002 Spring;14(2):223-36. doi:10.1176/jnp.14.2.223.

[51] Chen W, Hopfner F, Becktepe JS, Deuschl G. Rest tremor revisited: Parkinsons disease and other disorders. Transl Neurodegener. 2017 Jun;6(1). doi:10.1186/S40035-017-0086-4.

[52] National Institute of Neurological Disorders and Stroke. Tremor Fact Sheet;. Accessed: 2022-10-17. Available from: https://www.ninds.nih.gov/tremor-fact-sheet.

[53] Groot MH, Jagt-Willems HC, Campen JPCM, Lems WF, Beijnen JH, Lamoth CJC. A flexed posture in elderly patients is associated with impairments in postural control during walking. Gait Posture. 2014 Feb;39(2):767-72. doi:10.1016/J.GAITPOST.2013.10.015.

[54] Nonnekes J, Giladi N, Guha A, Fietzek UM, Bloem BR, Ruzicka E. Gait festination in parkinsonism: Introduction of two phenotypes. J Neurol. 2019 Feb;266(2):426-30. doi:10.1007/S00415-018-9146-7.

[55] Jankovic J, Tan EK. Parkinson's disease: Etiopathogenesis and treatment. J Neurol Neurosurg Psychiatry. 2020 Aug;91(8):795-808. doi:10.1136/JNNP-2019-322338.

[56] Mayo Clinic Staff. Parkinson's disease;. Accessed: 2022-11-11. Available from: https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055?p=1.

[57] National Institute on Aging. What Is Lewy Body Dementia? Causes, Symptoms, and Treatments;. Accessed: 2022-11-11. Available from: https://www.nia.nih.gov/health/what-lewy-body-dementia-causes-symptoms-and-treatments.

[58] Lanciego JL, Luquin N, Obeso JA. Functional Neuroanatomy of the Basal Ganglia. Cold Spring Harb Perspect Med. 2012 Dec;2(12). doi:10.1101/cshperspect.a009621.

[59] Berardelli A, Rothwell JC, Thompson PD, Hallett M. Pathophysiology of bradykinesia in Parkinson's disease. Brain. 2001 Nov;124(11):2131-46. doi:10.1093/brain/124.11.2131.

[60] Chinta SJ, Andersen JK. Dopaminergic neurons. Int J Biochem Cell Biol. 2005;37(5):942-6. doi:10.1016/j.biocel.2004.09.009.

[61] Mochizuki H, Choong CJ, Baba K. Parkinsons disease and iron. J Neural Transm (Vienna). 2020 Feb;127(2):181-7. doi:10.1007/s00702-020-02149-3.

[62] Francisco S. Melanins: skin pigments and much more - Types, structural models, biological functions, and formation routes. New Journal of Science. 2014 Mar;2014:1-28. doi:10.1155/2014/498276.

[63] Zucca FA, Basso E, Cupaioli FA, Ferrari E, Sulzer D, Casella L, et al. Neuromelanin of the human substantia nigra: An update. Neurotox Res. 2014 Oct;25:13-23. doi:10.1007/s12640-013-9435-y.

[64] Sasaki M, Shibata E, Tohyama K, Takahashi J, Otsuka K, Tsuchiya K, et al. Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. Neuroreport. 2006 Jul;17:1215-8. doi:10.1097/01.WNR.0000227984.84927.A7.

[65] Tokyo Metropolitian Institute of Medical Science. EBAN Disease Neuropathology;. Accessed: 2022-11-15. Available from: https://pathologycenter.jp/english/sample-room/disease.html.

[66] Martin-Bastida A, Lao-Kaim NP, Loane C, Politis M, Roussakis AA, Valle-Guzman N, et al. Motor associations of iron accumulation in deep grey matter nuclei in Parkinson's disease: A cross-sectional study of iron-related magnetic resonance imaging susceptibility. Eur J Neurol. 2017 Feb;24(2):357-65. doi:10.1111/ene.13208.

[67] Simon DK, Tanner CM, Brundin P. Parkinson Disease Epidemiology, Pathology, Genetics and Pathophysiology. Clin Geriatr Med. 2020 Feb;36(1):1-12. doi:10.1016/J.CGER.2019.08.002.

[68] Ohtsuka C, Sasaki M, Konno K, Koide M, Kato K, Takahashi J, et al. Changes in substantia nigra and locus coeruleus in patients with early-stage Parkinsons disease using neuromelanin-sensitive MR imaging. Neurosci Lett. 2013 Apr;541:93-8. doi:10.1016/j.neulet.2013.02.012.

[69] Savitt D, Jankovic J. Targeting α-Synuclein in Parkinsons Disease: Progress Towards the Development of Disease-Modifying Therapeutics. Drugs. 2019 Jun;79(8):797-810. doi:10.1007/S40265-019-01104-1.

[70] Brakedal B, Dölle C, Riemer F, Ma Y, Nido GS, Skeie GO, et al. The NADPARK study: A randomized phase I trial of nicotinamide riboside supplementation in Parkinson's disease. Cell Metab. 2022 Mar;34(3):396-407.e6. doi:10.1016/j.cmet.2022.02.001.

[71] Thenganatt MA, Jankovic J. Parkinson Disease Subtypes. JAMA Neurol. 2014 Apr;71(4):499-504. doi:10.1001/JAMANEUROL.2013.6233.

[72] Mestre TA, Fereshtehnejad SM, Berg D, Bohnen NI, Dujardin K, Erro R, et al. Parkinsons Disease Subtypes: Critical Appraisal and Recommendations. J Parkinsons Dis. 2021 1;11:395-404. doi:10.3233/JPD-202472.

[73] DiMarzio M, Madhavan R, Joel S, Hancu I, Fiveland E, Prusik J, et al. Use of Functional Magnetic Resonance Imaging to Assess How Motor Phenotypes of Parkinsons Disease Respond to Deep Brain Stimulation. Neuromodulation. 2020 Jun;23(4):515-24. doi:10.1111/NER.13160.

[74] Zaidel A, Arkadir D, Israel Z, Bergman H. Akinetorigid vs. tremor syndromes in Parkinsonism. Curr Opin Neurol. 2009 Aug;22(4):387-93. doi:10.1097/WCO.0b013e32832d9d67.

[75] Barber TR, Klein JC, Mackay CE, Hu MTM. Neuroimaging in pre-motor Parkinson's disease. NeuroImage Clin. 2017 Apr;15:215-27. doi:10.1016/J.NICL.2017.04.011.

[76] Blazejewska AI, Schwarz ST, Pitiot A, Stephenson MC, Lowe J, Bajaj N, et al. Visualization of nigrosome 1 and its loss in PD. Neurology. 2013 Aug;81(6):534-40. doi:10.1212/WNL.0b013e31829e6fd2.

[77] Schwarz ST, Afzal M, Morgan PS, Bajaj N, Gowland PA, Auer DP. The Swallow Tail Appearance of the Healthy Nigrosome - A New Accurate Test of Parkinson's Disease: A Case-Control and Retrospective Cross-Sectional MRI Study at 3T. PLoS One. 2014;9(4):e93814. doi:10.1371/JOURNAL.PONE.0093814.

[78] Pluijm M, Cassidy C, Ms MZ, Wallert E, Bruin K, Booij J, et al. Reliability and Reproducibility of Neuromelanin-Sensitive Imaging of the Substantia Nigra: A Comparison of Three Different Sequences. J Magn Reson Imaging. 2021 Mar;53(3):712-21. doi:10.1002/jmri.27384.

[79] Haacke EM, Chen Y, Utriainen D, Wu B, Wang Y, Xia S, et al. STrategically Acquired Gradient Echo (STAGE) imaging, part III: Technical advances and clinical applications of a rapid multi-contrast multi-parametric brain imaging method. Magn Reson Imaging. 2020 Jan;65:15-26. doi:10.1016/J.MRI.2019.09.006.

[80] Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. 2002 Nov;17(3):143-55. doi:10.1002/hbm.10062.

[81] Witoszynskyj S, Rauscher A, Reichenbach JR, Barth M. Phase unwrapping of MR images using Phi UN - A fast and robust region growing algorithm. Med Image Anal. 2009 Apr;13(2):257-68. doi:10.1016/j.media.2008.10.004.

[82] Schweser F, Deistung A, Lehr BW, Reichenbach JR. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism? Neuroimage. 2011 Feb;54(4):2789-807. doi:10.1016/j.neuroimage.2010.10.070.

[83] Haacke EM, Tang J, Neelavalli J, Cheng YCN. Susceptibility mapping as a means to visualize veins and quantify oxygen saturation. J Magn Reson Imaging. 2010 Sep;32(3):663-76. doi:10.1002/jmri.22276.

[84] Jahn A. Chapter 3: Coregistration; 2019. Accessed: 2023-02-28. Available from: https://andysbrainbook.readthedocs.io/en/latest/SPM/SPM_Short_Course/SPM_04_Preprocessing/03_SPM_Coregistration.html.

[85] Jenkinson M, Bannister P, Brady JM, Smith SM. Improved Optimisation for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage. 2002 Oct;2:825-41. doi:10.1016/s1053-8119(02)91132-8.

[86] Jenkinson M, Smith SM. A Global Optimisation Method for Robust Affine Registration of Brain Images. Med Image Anal. 2001 Jun;5(2):143-56. doi:10.1016/s1361-8415(01)00036-6.

[87] Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage. 2009 Oct;48(1):63-72. doi:10.1016/j.neuroimage.2009.06.060.

[88] The Wellcome Centre for Human Neuroimaging. SPM12; 2023. Accessed: 2023-02-28. Available from: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/.

[89] Safai A, Prasad S, Chougule T, Saini J, Pal PK, Ingalhalikar M. Microstructural abnormalities of substantia nigra in Parkinson's disease: A neuromelanin sensitive MRI atlas based study. Human Brain Mapping. 2020 Apr;41:1323-33. doi:10.1002/HBM.24878.

[90] Safai A. NMS-SNc-atlas; 2019. Accessed: 2022-09-22. Available from: https://github.com/apoorvasafai/NMS-SNc-atlas.

[91] Jenkinson M. Fslutils; 2020. Accessed: 2022-09-23. Available from: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Fslutils.

[92] Ye R, Rua C, O'Callaghan C, Jones PS, Hezemans FH, Kaalund SS, et al. An in vivo probabilistic atlas of the human locus coeruleus at ultra-high field. Neuroimage. 2021 Jan;225:117487. doi:10.1016/j.neuroimage.2020.117487.

[93] NeuroImaging Tools and Resources Collaboratory. An in vivo 7T Probabilistic Atlas of the Human Locus Coeruleus;. Accessed: 2022-09-22. Available from: https://www.nitrc.org/projects/lc_7t_prob/.

[94] Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive Statistics and Normality Tests for Statistical Data. Ann Card Anaesth. 2019 Jan-Mar;22(1):67-72. doi:10.4103/aca.ACA_157_18.

[95] Koutroumbas K, Theodoridis S. Pattern recognition. 4th ed. Academic Press; 2008. Available from: https://darmanto.akakom.ac.id/pengenalanpola/-Pattern%20Recognition%204th%20Ed.%20(2009).pdf.

[96] Antoneli F, Passos FM, Lopes LR, Briones MRS. A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies. PLoS One. 2018 Jan;13(1):e0190826. doi:10.1371/journal.pone.0190826.

[97] Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. Indian J Dermatol. 2016 May-Jun;61(3):251-60. doi:10.4103/0019-5154.182416.

[98] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30. Available from: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

[99] Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions. Computational Statistics and Data Analysis. 2008 Aug;52(12):5186-201. doi:10.1016/j.csda.2007.11.008.

[100] Sharma AK. Text book of correlations and regression. 1st ed. Discovery Publishing House; 2005. Available from: https://-books.google.no/books?hl=nolr=id=obb7_6k6XDQCoi=fndpg=PA1dq=pearson+correlation+coefficient+bookots=SeJLtsNDfesig=xRbOdW-krIiu2i62G9Dm2Gx58sEredir_esc=yv=onepageq=pearson%20correlation%20coefficient%20bookf=false.

[101] Doan QH, Mai SH, Do QT, Thai DK. A cluster-based data splitting method for small sample and class imbalance problems in impact damage classification. Applied Soft Computing. 2022 May;120. doi:10.1016/j.asoc.2022.108628.

[102] Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021 Mar;2(3):160. doi:10.1007/s42979-021-00592-x.

[103] Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLoS One. 2015 Mar;10(3). doi:10.1371/journal.pone.0118432.

[104] Breiman L. Random Forests. Machine Learning. 2001;45:5-32. doi:10.1023/a:1010933404324.

[105] Molnar C. Interpretable Machine Learning. 2nd ed.; 2022. Available from: https://christophm.github.io/interpretable-ml-book.

[106] Scikit-learn. 4.2. Permutation feature importance;. Accessed: 2023-05-25. Available from: https://scikit-learn.org/stable/modules/permutation_importance.html.

[107] Riederer P, Jellinger KA, Kolber P, Hipp G, Sian-Hülsmann J, Krüger R. Lateralisation in Parkinson disease. Cell and Tissue Research. 2018 Apr;373:297-312. doi:10.1007/s00441-018-2832-z.

[108] Minati L, Grisoli M, Carella F, Simone TD, Bruzzone MG, , et al. Imaging Degeneration of the Substantia Nigra in Parkinson Disease with Inversion-Recovery MR Imaging. AJNR Am J Neuroradiol. 2007 Feb;28(2):309-13. PMCID:PMC7977418.

[109] Donders ART, Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 2006 Oct;59(10):1087-91. doi:10.1016/j.jclinepi.2006.01.014.

[110] He N, Chen Y, LeWitt PA, Yan F, Haacke EM. Application of Neuromelanin MR Imaging in Parkinson Disease. Journal of Magnetic Resonance Imaging. 2022 Aug;57(2):337-52. doi:10.1002/jmri.28414.

[111] Morris S, Tan A, Smith DA, Grehl M, Han-Huang K, Naidich TP, et al. Submillimeter variation in human locus coeruleus is associated with dimensional measures of psychopathology: An in vivo ultra-high field 7-Tesla MRI study. NeuroImage: Clinical. 2020 Jan;25:102148. doi:10.1016/j.nicl.2019.102148.

[112] Despotovi I, Goossens B, Philips W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Comput Math Methods Med. 2015 Mar;2015:450341. doi:10.1155/2015/450341.

[113] Hwang KS, Langley J, Tripathi R, Hu XP, Huddleston DE. In vivo detection of substantia nigra and locus coeruleus volume loss in Parkinson's disease using neuromelanin-sensitive MRI: Replication in two cohorts. PLoS One. 2023 Apr;2023:e0282684. doi:10.1371/journal.pone.0282684.

[114] Castellanos G, Fernández-Seara MA, Lorenzo-Betancor O, Ortega-Cubero S, Puigvert M, Uranga J, et al. Automated Neuromelanin Imaging as a Diagnostic Biomarker for Parkinson's Disease. Movement Disorders. 2015 Jun;30(7):945-52. doi:10.1002/mds.26201.

[115] Vitali P, Pan MI, Palesi F, Germani G, Faggioli A, Anzalone N, et al. Substantia Nigra Volumetry with 3-T MRI in De Novo and Advanced Parkinson Disease. Radiology. 2020 Aug;296(2):401-10. doi:10.1148/radiol.2020191235.

[116] Chen Y, Gong T, Sun C, Yang A, Gao F, Chen T, et al. Regional age-related changes of neuromelanin and iron in the substantia nigra based on neuromelanin accumulation and iron deposition. Eur Radiol. 2023 May;33(5):3704-14. doi:10.1007/s00330-023-09411-8.

[117] Visser E, Keuken MC, Forstmann BU, Jenkinson M. Automated segmentation of the substantia nigra, subthalamic nucleus and red nucleus in 7 T data at young and old age. Neuroimage. 2016 Oct;139:324-36. doi:10.1016/j.neuroimage.2016.06.039.

[118] Krupicka R, Marecek S, Mala C, Lang M, Klempir O, Duspivova T, et al. Automatic Substantia Nigra Segmentation in Neuromelanin-Sensitive MRI by Deep Neural Network in Patients With Prodromal and Manifest Synucleinopathy. Physiol Res. 2019;68(4):453-8. doi:10.33549/physiolres.934380.

[119] Seghier ML. Laterality index in functional MRI: methodological issues. Magn Reson Imaging. 2008 Jan;26(5):594-601. doi:10.1016/j.mri.2007.10.010.

[120] Belyadi H, Haghighat A. Machine Learning Guide for Oil and Gas Using Python; 2021. doi:10.1016/B978-0-12-821929-4.00001-9.

[121] Scikit-learn. 6.4. Imputation of missing values;. Accessed: 2023-05-28. Available from: https://scikit-learn.org/stable/modules/impute.htmlimpute.

[122] Joseph VR. Optimal ratio for data splitting. Statistical analysis and data mining. 2022 Aug;15(4):531-8. doi:10.1002/sam.11583.

[123] Gupta B, Rawat A, Jain A, Arora A, Dhami N. Analysis of Various Decision Tree Algorithms for Classification in Data Mining. International Journal of Computer Applications. 2017 Apr;163(8). Accessed: 2023-05-28. Available from: https://www.ijcaonline.org/archives/volume163/number8/gupta-2017-ijca-913660.pdf.

[124] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing. 2020 Nov;415:295-316. doi:10.1016/j.neucom.2020.07.061.

[125] Altmann A, Toloi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010 May;26(10):1340-7. doi:10.1093/bioinformatics/btq134.