

Aus der Klinik und Poliklinik für Nuklearmedizin
der Universität zu Köln
Direktor: Universitätsprofessor Dr. med. A. Drzezga

Die automatisierte Bewertung von Dopamin- Transporter-SPECT basierend auf künstlicher Intelligenz

Inaugural-Dissertation zur Erlangung der ärztlichen Doktorwürde
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Michelle Meier
aus Köln

promoviert am 18. März 2024

Gedruckt mit Genehmigung der Medizinischen Fakultät der Universität zu Köln
Druckjahr 2023

Dekan: Universitätsprofessor Dr. med. G. R. Fink
1. Gutachter: Universitätsprofessor Dr. med. T. van Eimeren
2. Gutachterin: Professorin Dr. med. H. Endepols

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich Unterstützungsleistungen von folgenden Personen erhalten:

Herr Universitätsprofessor Dr. med. Thilo van Eimeren
Leiter der Arbeitsgruppe Multimodale Bildgebung
Klinik und Poliklinik für Nuklearmedizin
Universität zu Köln

Herr PD Dr. med. Jochen Hammes
Wissenschaftlicher Mitarbeiter
Klinik und Poliklinik für Nuklearmedizin
Universität zu Köln

Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die dieser Arbeit zugrunde liegenden DAT-SPECT-Rohdatensätze wurde ohne meine Mitarbeit in der Klinik und Poliklinik für Nuklearmedizin der Universität zu Köln im Rahmen der diagnostischen Abklärung nach fachärztlicher Überweisung durch das dortige Personal erstellt.

Das Herunterladen der Rohdaten und die räumliche Normalisierung erfolgten unter direkter Anleitung zusammen mit Herrn PD Dr. med. Jochen Hammes. Die Analyse und manuelle Auswahl der Bilddateien erfolgten nach Erläuterung durch Herrn Universitätsprofessor Dr. med. Thilo van Eimeren und Herrn PD Dr. med. Jochen Hammes selbstständig durch mich. Die Präprozessierung erfolgte unter direkter Anleitung mit Herrn PD Dr. med. Jochen Hammes. Die Datenaugmentierung, das Trainieren und das Testen der Netzwerke erfolgten nach direkter Anleitung durch Herrn PD Dr. med. Jochen Hammes selbstständig durch mich. Die statistische Analyse und anschließende Auswertung der Daten unter Verwendung einer integrierten Microsoft® Excel®-Anwendung in Google® Drive und von SPSS® erfolgten selbstständig durch mich.

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Köln, den 10.10.2023

Unterschrift:

Danksagung

Ich möchte mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Bei Herrn Univ.-Prof. Dr. med. Thilo van Eimeren möchte ich mich für die Bereitstellung des Themas und die sorgfältige und zuverlässige Betreuung und Unterstützung bedanken.

Herrn PD Dr. med. Jochen Hammes gilt Dank für die intensive Einarbeitung und Begleitung bei der Projektbearbeitung.

Ich danke allen Mitarbeitenden der Klinik und Poliklinik für Nuklearmedizin der Universität zu Köln, geleitet von Univ.-Prof. Dr. med. Alexander Drzezga, in der ich die Arbeit angefertigt habe und in der die zu Grunde liegenden Bilddaten erstellt wurden.

Außerdem möchte ich mich bei dem Team der Arbeitsgruppe „Multimodale Bildgebung neuronaler Netzwerke“ für die freundliche Aufnahme und Unterstützung bedanken: Univ.-Prof. Dr. med. Thilo van Eimeren, PD Dr. Jochen Hammes, Dr. Gérard Nisal Bischof, Dr. med. Hendrik Theis, Dr. Kathrin Giehl, Dr. Merle C. Hönig, Omer Reiner, Adrian Asendorf, Verena Dzialas, Magdalena Banwinkler, Elena Döring und Julia Lothmann.

Außerdem bedanke ich mich herzlich bei meiner Familie, insbesondere bei meinen Eltern, und bei meinen Freundinnen und Freunden für ihre liebevolle Förderung.

Für meine Familie

Widmung

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	9
1. ZUSAMMENFASSUNG	11
2. EINLEITUNG	12
2.1 Thematische Einführung	12
2.2 Morbus Parkinson	13
2.2.1. Epidemiologie und Sozioökonomie	13
2.2.2. Ätiologie und Pathophysiologie	13
2.2.3. Neuropathologie	14
2.2.4. Die Basalganglien	15
2.2.5. Klinische Symptomatik	16
2.2.6. Diagnostische Prinzipien	18
2.2.7. Therapie und Prognose	19
2.3 Bildgebung des dopaminergen Systems	20
2.3.1. FP-CIT-SPECT - Funktionsweise und Untersuchungsablauf	20
2.3.2. FP-CIT-SPECT - Indikation und diagnostische Bewertung	23
2.3.3. ¹⁸ F-DOPA-PET	24
2.3.4. IBZM-SPECT	25
2.3.5. MIBG-Szintigraphie	25
2.3.6. Transkranielle Sonographie	26
2.4 Maschinelles Lernen	27
2.4.1. Überblick / Grundlagen	27
2.4.2. Varianten und Anwendungsgebiete	28
2.4.3. Tiefes Lernen	30
2.4.4. Künstliche neuronale Netzwerke	31
2.4.5. Netztopologien	34
2.4.6. Transfer-Lernen	35
2.4.7. Verfügbare Deep Learning-Instrumente	37
2.4.8. Microsoft® Cognitive Toolkit	37
2.4.9. TensorFlow™	38
2.4.10. Keras	39
2.5 Studienlage	39
2.6 Fragestellungen und Ziel der Arbeit	42

3.	MATERIAL UND METHODEN	43
3.1	Material	43
3.1.1.	Herkunft der Daten	43
3.1.2.	Verwendete Systeme	43
3.2	Methoden	44
3.2.1.	Herunterladen und erste Bearbeitung	44
3.2.2.	Auswahl und Analyse geeigneter Dateien	46
3.2.3.	Präprozessierung	49
3.2.4.	Datenaugmentierung	51
3.2.5.	Training und Testen der Netzwerke	52
3.2.6.	Statistische Analyse	57
4.	ERGEBNISSE	58
4.1	Klinische Parameter	58
4.2	Verarbeitung der Bilddateien	59
4.3	Ergebnisse der Klassifikation	64
5.	DISKUSSION	71
5.1	Datengrundlage	71
5.2	Klassifikationsleistung	74
5.3	Juristische Aspekte	75
5.4	Ethische Aspekte	77
5.5	Limitationen	78
5.6	Zusammenfassung und Ausblick	81
6.	LITERATURVERZEICHNIS	83
7.	ANHANG	92
7.1	Abbildungsverzeichnis	92
7.2	Tabellenverzeichnis	92

7.3	Skripts	93
8.	VORABVERÖFFENTLICHUNGEN VON ERGEBNISSEN	104

Abkürzungsverzeichnis

AI	artificial intelligence [englisch], künstliche Intelligenz (siehe unten)
APO E	Apolipoprotein E
AUC	area under the curve [englisch]
BGB	Bürgerliches Gesetzbuch
bzw.	beziehungsweise
ca.	circa
CBD	corticobasal degeneration [englisch], kortikobasale Degeneration
CI	Konfidenzintervall
(c)CT	(kraniale) Computertomographie
CNN	convolutional neural network [englisch], faltendes neuronales Netzwerk
CSV	comma separated value [englisch]
DAT	Dopamintransporter
d. h.	das heißt
DICOM	Digital Imaging and Communications in Medicine, [englisch]
DS-GVO	Datenschutzgrundverordnung
ENS	enterisches Nervensystem
ET	essentieller Tremor
EU	Europäische Union
FBP	filtered back projection [englisch], gefilterte Rückprojektion
FDG	2-Fluor-2-desoxy-D-Glucose
FN	falsch negativ
FP	falsch positiv
FP-CIT	N- ω -fluoropropyl-2 β -carbomethoxy-3 β -(4-iodophenyl) nortropane
GE	General Electric
IBZM	Iodobenzamid
ID	Identifikationsnummer
i. d. R.	in der Regel
IPS	idiopathisches Parkinson-Syndrom
JPEG	Joint Photographic Experts Group [englisch]
KI	künstliche Intelligenz
LBD	Lewy body dementia [englisch], Lewy-Körperchen-Demenz
L-DOPA	Levodopa, 3,4-Dihydroxy-Phenylalanin
MAO-B	Monoaminoxidase B
MCI	mild cognitive impairment [englisch]

MDCG	Medical Device Coordination Group [englisch], Koordinierungsgruppe Medizinprodukte
MDR	Medical Device Regulation [englisch], Medizinprodukteverordnung
MDS-UPDRS	Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale, [englisch]
MIBG	Metaiodbenzylguanidin
MIP	Maximumintensitätsprojektion
MPDG	Medizinprodukterechtdurchführungsgesetz
MSA	Multisystematrophie
(c)MRT	(kraniale) Magnetresonanztomographie
NIFTI	Neuroimaging Informatics Technology Initiative, [englisch]
PACS	Picture Archiving and Communication System, [englisch]
PET	Positronenemissionstomographie
PNS	peripheres Nervensystem
PPMI	Parkinson Progression Marker Initiative [englisch]
ProdHaftG	Produkthaftungsgesetz
PSP	progressive supranuclear palsy [englisch], progressive supranukleäre Blickparese
RN	richtig negativ
RNN	rekurrentes neuronales Netzwerk
RP	richtig positiv
ROI	region of interest, [englisch]
SEV	Sekundärelektionenvervielfacher
SPECT	single photon emission tomography [englisch], Einzelphotonenemissionstomographie
SPSS	Statistical Package for the Social Sciences, [englisch])
SWEDD	subject without evidence of dopaminergic deficit, [englisch]
VPN	virtuelles privates Netzwerk
vs.	versus
z.B.	zum Beispiel
ZNS	zentrales Nervensystem

1. Zusammenfassung

Bei der DAT-SPECT (dopamine transporter single photon emission computer tomography, [englisch]) handelt es sich um nuklearmedizinisches diagnostisches Verfahren, welches unter anderem bei dem klinischen Bild des Parkinsonismus Anwendung findet. Es bietet eine in vivo-Einschätzung über die Integrität dopaminerger Synapsen und unterliegt einer gewissen Interrater-Variabilität. Künstliche Intelligenz (KI) hat in den vergangenen Jahren in vielen Bereichen des Alltags Einzug gehalten. In der Medizin gibt es bereits jetzt Algorithmen, die die menschliche Beurteilung in ihrer Leistungsfähigkeit übertreffen. Hier ist besonders das Teilgebiet des tiefen Lernens in den Fokus gerückt. Ziel des Projektes war die Erprobung eines Ansatzes, mit dieser Methodik DAT-SPECT-Dateien automatisiert zu klassifizieren.

Es wurden retrospektiv aus einem Datensatz von 1679 DAT-SPECT-Aufnahmen nach diagnostischer Bewertung durch klinisch tätiges ärztliches Personal 442 Einzeldateien ausgewählt und binär klassifiziert. Anschließend wurde eine KI-Umgebung basierend auf Google Tensorflow® aufgesetzt. Es wurden verschiedene Kombinationen aus Fraktionen eines faltenden künstlichen neuronalen Netzwerkes, Durchlaufraten und verschiedenen Trainingsdatensätzen angewandt. Diese wurden danach mit unbekanntem Daten einem Testprozess unterzogen.

Es wurden Genauigkeiten von bis zu 90%, Sensitivitäten und Spezifitäten von jeweils bis zu 100% erreicht. Die positiven und negativen Prädiktionswerte lagen ebenfalls je bei bis zu 100%. Allerdings zeigten die Ergebnisse starke Schwankungen je nach benutzter Trainings- und Testkombination. Tendenziell schien eine höhere Netzwerkfraktion mit höherer Laufrate bessere Resultate zu erbringen.

Insgesamt illustriert die vorliegende Arbeit eine Möglichkeit, DAT-SPECT-Dateien automatisiert zu beurteilen. Auch wenn sowohl in der wissenschaftlichen Literatur wie auch in diesem Projekt teils beeindruckende Ergebnisse erzielt wurden, stellt die eingeschränkte Nachvollziehbarkeit sowie die bislang fehlende klinische Korrelation automatisierter Klassifikationen eine Herausforderung dar. Auch ergeben sich in diesem Kontext in Anbetracht der Verwendung sensibler Daten juristische wie ethische Fragestellungen, deren Bedeutung mit der zunehmenden Digitalisierung des Gesundheitssystems wachsen wird.

2. Einleitung

2.1 Thematische Einführung

Der Morbus Parkinson zählt zur Gruppe der neurodegenerativen Erkrankungen und betrifft die Basalganglien und das extrapyramidal-motorische System. Er basiert auf dem Untergang dopaminergener Neuronen. Der Morbus Parkinson stellt vor allem in zunehmendem Alter eine recht prävalente Entität mit mannigfaltigen Manifestationen dar, wobei bestimmte motorische Symptome typisch sind. Es treten gleichermaßen auch Symptome auf, die nicht die Motorik betreffen. Die Diagnostik ergibt sich in erster Linie aus der Zusammenschau der klinischen Symptomatik und einer eventuellen Responsivität auf Levodopa. Bei unklaren klinischen Befunden besteht die Möglichkeit, mithilfe der Dopamin-Transporter-SPECT die Diagnose zu erhärten. Hierbei erfolgt die Auswertung der Bilddaten bis dato zum einen visuell, zum anderen semiquantitativ.

Der Einsatz von auf künstlicher Intelligenz (KI) fußenden Algorithmen hat sich bereits in diversen Sparten im nicht-medizinischen Feld wie auch im medizinischen Bereich als erfolgreich erwiesen. Hierbei zeigt insbesondere die auf bildgebenden Verfahren gestützte medizinische Diagnostik besonderes Potential zur Anwendung von KI.

In dieser Arbeit wird nach entsprechender Präprozessierung von Dopamin-Transporter-SPECT-Dateien die Erprobung eines solchen Programmes dargestellt. Anschließend werden dessen Implementierbarkeit und Potential zur Analyse und Auswertung dieser Dateien untersucht. Zuerst werden die Grundlagen im Hinblick auf den Morbus Parkinson zusammenfassend dargestellt, wobei insbesondere die bildgebende Diagnostik genauer beleuchtet wird. Im Anschluss erfolgt eine Erläuterung des sogenannten maschinellen Lernens, dessen Anwendung im Methodenteil in Bezug auf das zu Grunde liegende Projekt erklärt wird. Nach Darstellung der gewonnenen Ergebnisse erfolgt anschließend eine Diskussion der Erkenntnisse. Hierbei werden die möglichen Bedeutungen und Stärken sowie potentielle Herausforderungen in Bezug auf die Anwendung von KI in der Bewertung von Dopamin-Transporter-SPECT-Dateien sowie in der Medizin im Allgemeinen berücksichtigt. Zuletzt wird ein Ausblick gegeben, inwiefern der Einsatz KI-basierter Algorithmen verbessert und in der Zukunft zum Einsatz kommen kann.

2.2 Morbus Parkinson

2.2.1. Epidemiologie und Sozioökonomie

Unter den neurodegenerativen Erkrankungen stellt der Morbus Parkinson nach dem Morbus Alzheimer die zweithäufigste Entität dar.¹ In Deutschland sind ungefähr 1-2 % der Menschen im Alter von 65 Jahren und älter betroffen.² Die Prävalenz steigt mit zunehmendem Alter deutlich an und Männer sind häufiger betroffen als Frauen.^{3,4} Im Jahr 2010 waren ca. 220 000 Personen in Deutschland an Morbus Parkinson erkrankt, wodurch im Durchschnitt Gesamtkosten in Höhe von ungefähr 13 000 € pro Person entstanden.⁵ Durch Faktoren wie den demographischen Wandel, gestiegenes Krankheitsbewusstsein und verbesserte diagnostische Methoden haben Inzidenz und Prävalenz der Erkrankung in den letzten Jahrzehnten weltweit in erheblichem Maße zugenommen. Eine Entwicklung im Sinne einer weiteren Erhöhung der Anzahl an Erkrankten ist wahrscheinlich.^{4,6}

2.2.2. Ätiologie und Pathophysiologie

Der Morbus Parkinson zählt zur Gruppe der Parkinson-Syndrome. Diese bezeichnen Erkrankungen, welche sich vor allem durch einen als Parkinson-Trias bezeichneten Symptomkomplex (siehe Kapitel 2.2.5) äußern. Man unterscheidet primäre, sekundäre und atypische Parkinson-Syndrome. Die primären Formen lassen sich wiederum in idiopathisch oder genetisch einteilen. Ätiologisch lassen sich die Parkinson-Syndrome also in insgesamt vier Kategorien klassifizieren.

Das idiopathische Parkinson-Syndrom, auch als Morbus Parkinson bezeichnet, macht mit ca. 75% der Fälle den größten Anteil aus.^{7,8} Hierbei kommt es aus bislang ungeklärter Ursache zur Akkumulation des Proteins α -Synuclein im Bereich der dopaminergen Neurone der Substantia nigra sowie im Locus coeruleus. Dieser Prozess führt zum Untergang der Neurone. Besonders gefährdet sind melaninhaltige Nervenzellen, sodass deren Depigmentierung Rückschlüsse auf ein fortgeschrittenes Krankheitsstadium erlaubt.⁹ Auch eine Degeneration der cholinergen Neurone des Nucleus basalis ist möglich.¹⁰ Auf die neuropathologischen Veränderungen beim idiopathischen Parkinson-Syndrom wird im folgenden Unterkapitel detaillierter eingegangen. Bei den genetischen Formen des Parkinson-Syndroms bilden monogenetische Mutationen die ätiologische Grundlage der Erkrankung. Bislang identifizierte betroffene Gene sind bei den autosomal dominant vererbten Formen beispielsweise LRRK2 und SNCA (Synuclein alpha). Bei den autosomal rezessiv vererbten Formen sind unter anderem die Ubiquitin-Ligase Parkin oder die Proteine PINK1 und DJ-1 betroffen.^{11,12} Durch Genom-weite Analysen wird es in den nächsten Jahren voraussichtlich zur Identifizierung weiterer ursächlicher Mutationen kommen. Der Untergang der dopaminergen Neurone bei den

atypischen Parkinson-Syndromen wird durch andere neurologische Erkrankungen als den Morbus Parkinson hervorgerufen. Außerdem gibt es atypische Parkinson-Syndrome, bei denen nicht nur vorrangig die Substantia nigra, sondern auch weitere Regionen mit dopaminergen Neuronen im zentralen Nervensystem degenerieren. Dadurch kommt es zu zusätzlichen Symptomen, weshalb diese atypischen Parkinson-Syndrome auch als Parkinson-Plus-Syndrome bezeichnet werden.¹³ Hier gibt es vorrangig vier Entitäten zu unterscheiden: Die progressive supranukleäre Blickparese (progressive supranuclear palsy [englisch], PSP), die Multisystematrophie (MSA), die Lewy-Körperchen-Demenz (Lewy body dementia [englisch], LBD) und die kortikobasale Degeneration (corticobasal degeneration [englisch], CBD).¹⁴ Die PSP und die CBD sind wie auch der Morbus Alzheimer Tauopathien, basieren also vornehmlich auf einer pathologischen Akkumulation des Tau-Proteins.^{15,16} Bei der MSA gibt es zwei Subtypen: Den Parkinson- und den zerebellären Typ, respektive MSA-P und MSA-C. Bei der MSA und der LBD handelt es sich wie bei dem idiopathischen Parkinson-Syndrom um Synucleopathien.^{17,18}

Bei den sekundären Parkinson-Syndromen beruht die Symptomatik auf einer bekannten Ursache. Die häufigste Form ist das medikamentöse Parkinsonoid, welches durch Medikamente mit antidopaminergem Wirkung ausgelöst wird. Besonders häufig eingesetzte Präparate sind atypische Antipsychotika, Metoclopramid und Lithium sowie Kalziumantagonisten und Amiodaron.⁸ Außerdem ist der vaskuläre Parkinsonismus zu nennen, der durch multiple kleine Infarkte im Bereich der Basalganglien ausgelöst wird. Er wird in manchen Quellen auch den atypischen Parkinson-Syndromen zugeordnet.¹⁹ Andere mögliche Genesen sind posttraumatisch, toxisch, metabolisch oder entzündlich.²⁰

2.2.3. Neuropathologie

Neuropathologisch zeichnet sich der Morbus Parkinson durch sogenannte Lewy-Körperchen aus. Bei diesen handelt es sich um eosinophile intrazytoplasmatische Einschlusskörperchen. Sie bestehen vornehmlich aus dem immunhistochemisch nachweisbaren Protein α -Synuclein, dem proteolytischen Stressprotein Ubiquitin, phosphoryliertem Neurofilament und dem Protein Synphilin-1.^{21,22} α -Synuclein ist ein physiologisch vorliegendes synaptisches Protein, welches bei Gesunden nur in monomerer Form vorkommt. Bei Erkrankten kommt es durch Fehlfaltung erst zur Bildung von Oligomeren und schließlich zu amyloiden Fibrillen.^{23,24} Die Akkumulation des α -Synucleins und der anderen Proteine wirkt immunreaktiv. Durch neuroinflammatorische Prozesse kommt es aufgrund einer selektiven neuronalen Vulnerabilität vor allem zum Untergang der neuromelaninhaltigen dopaminergen Neurone. Dies führt konsekutiv zu einem dopaminergen Defizit. Hierbei ist besonders die Pars compacta der Substantia nigra betroffen, sodass es zu einem Mangel an Dopamin im Striatum kommt.

Die Neuropathologie betrifft allerdings nicht nur das zentrale Nervensystem (ZNS), sondern sie lässt sich ebenso im peripheren (PNS) sowie im enterischen Nervensystem (ENS) nachweisen. Von daher ist das idiopathische Parkinson-Syndrom besser gesagt als polysystemische denn als monosystemische Erkrankung anzusehen, wenngleich es das ZNS in besonderem Maße betrifft.²⁵

Das Ausmaß der neuropathologischen Veränderungen beim Morbus Parkinson wurde erstmals durch Heiko Braak beschrieben, welcher sechs Stadien der Progression definierte.^{26,27} In Stadium eins zeigten sich Einschlusskörperchen lediglich im Bereich der dorsalen motorischen Nuclei der Nervi glossopharyngei und vagi und/oder im intermediären Bereich der Zona reticularis sowie in den Bulbi olfactorii. Es sind dementsprechend nur Gebiete in der Medulla oblongata und die Riechbahn betroffen, jedoch keine kortikalen Strukturen. In Stadium zwei ist die Parkinson-Pathologie zusätzlich in verschiedenen Arealen des pontinen Tegmentum nachweisbar. In Stadium drei finden sich erstmals im Mesencephalon und dort besonders in der Substantia nigra Lewy-Körperchen. Kortikale Läsionen finden sich erst ab Stadium vier, in dem sie sich aber auf den temporalen Mesokortex und den Allokortex beschränken. Eine Beteiligung des Neokortex erlaubt je nach Ausmaß eine Einteilung in Stadium fünf oder sechs. Es fällt auf, dass die Ausbreitung der pathologischen Veränderungen vom Hirnstamm nach kranial Richtung Neokortex erfolgt.²⁷ Braak stellt die Hypothese auf, dass die Pathologie im Darm ihren Ursprung nimmt und sich neurotrop über intestinale Fasern des Nervus vagus (Truncus vagalis anterior und posterior) in Richtung des Gehirns ausbreitet.²⁸ Diese Verbindung wird als Darm-Hirn-Achse bezeichnet. Als Auslöser dieses Prozesses werden ein unbekanntes Pathogen und/oder ein Umwelt-assoziiertes Toxin diskutiert, welche entweder intestinal oder über die Riechschleimhaut aufgenommen werden.²⁹⁻³¹

2.2.4. Die Basalganglien

Die Basalganglien umfassen verschiedene subkortikale Kerngebiete. Ihre Bestandteile sind je nach Autor*in teilweise unterschiedlich eingegrenzt. Nach Mello und Villares sowie Haber und Gdowski setzen sie sich neuroanatomisch zusammen aus dem Nucleus caudatus, dem Nucleus lentiformis (bestehend aus Putamen und Globus pallidus externus und internus), dem Nucleus subthalamicus und der Substantia nigra (bestehend aus Pars reticulata und Pars compacta). Die größte Komponente bildet der Komplex aus Putamen und Nucleus caudatus, der als das Corpus striatum bezeichnet wird.^{32,33} Abbildung 1 visualisiert diese Struktur.

Es lässt sich in einen phylogenetisch alten, ventralen Anteil (Paleostriatum) und in einen entwicklungsgeschichtlich jüngeren dorsalen Anteil (Neostriatum) unterteilen. Sie werden

durch die Capsula interna voneinander getrennt. Das dorsale Striatum beinhaltet den Nucleus caudatus und das Putamen. Aus funktioneller Sicht wird außerdem der Nucleus accumbens zum Striatum (und damit ebenfalls zu den Basalganglien) gezählt, welcher den hauptsächlichsten Anteil des ventralen Striatum bildet. Die Hauptaufgabe des Striatums liegt in der Steuerung der Willkürmotorik.³⁴

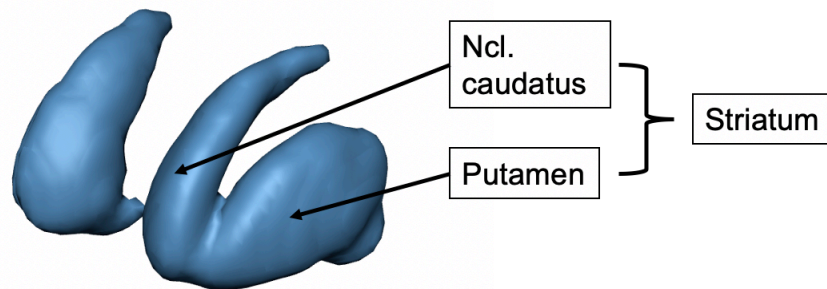


Abbildung 1: Darstellung des Striatums

Das Claustrum sowie das Corpus amygdaloideum gehören zwar phylogenetisch ebenfalls zu den Basalganglien, werden aber mittlerweile nicht mehr als Bestandteil dieses Kerngebietes angesehen.³⁵ Da der Übergang von anatomischen zu funktionellen Zusammenhängen fließend ist, ergibt sich hieraus der Hauptgrund für unterschiedliche Auffassungen bezüglich der Bestandteile der Basalganglien. Ihre Funktionen sind vielfältig: Die wesentliche Aufgabe besteht in der Steuerung von sowohl willkürlichen als auch unwillkürlichen Bewegungen, sie sind aber ebenfalls Bestandteil affektiver und kognitiver Funktionskreise. Von daher sind sie auch an der Regulation von Lernprozessen, Verhalten, Emotion und Motivation beteiligt.³⁶⁻³⁸ Die Basalganglien sind funktionell untereinander verbunden, wobei meist drei Schleifen unterschieden werden: Die motorische Schleife, die assoziative Schleife für die kognitiven Funktionen und die limbische Schleife.^{39,40} Im physiologischen Zustand befinden sich diese in einem austarierten Rückkopplungssystem, in dem Impulse verstärkt und zugleich komplementär gehemmt werden. Beim idiopathischen Parkinson-Syndrom liegt ein dopaminerges Defizit im Bereich des Striatum vor, sodass dieses Gleichgewicht gestört wird. Letztendlich kommt es aufgrund eines Wegfalls der thalamischen Disinhibition zu einer Imbalance insbesondere der motorischen Schleife, aus der sich die charakteristische klinische Konstellation ergibt.⁴¹

2.2.5. Klinische Symptomatik

Die charakteristische Symptomatik der Erkrankung ist gekennzeichnet durch die Parkinson-Trias. Sie umfasst die Symptome Rigor, Ruhetremor und Brady- bzw. Akinese. Beim Rigor handelt es sich um eine generalisierte Erhöhung des Muskeltonus. Dadurch kommt es bei

passiver Bewegung z.B. der Extremitäten zu einem wächsernen Widerstand, der ruckartig nachlässt. Dies wird auch als Zahnrad-Phänomen bezeichnet und tritt bei aktiver Bewegung der kontralateralen Seite verstärkt auf.^{7,42} Der Ruhetremor stellt eine rhythmische Zitterbewegung dar und ist auch als Pillen-Dreher- oder Münzenzähl-Tremor bekannt, da er vor allem die Hände betrifft. Er kann sich aber auch an den Füßen und selten auch am Kopf manifestieren. Der grobschlägige Tremor weist eine Frequenz von vier bis sechs Hertz auf und tritt in der Regel unilateral auf. Verstärkt wird er durch emotionale Anspannung und Müdigkeit, gemindert wird er bei der Initiierung willkürlicher Bewegungen. Im Schlaf fehlt er gänzlich.^{7,43}

Die Bradykinese bezeichnet eine Verlangsamung von Bewegungen, die initial besonders durch ein verringertes Mitschwingen der Arme während des Gehens auffällt. Bewegungen können außerdem eine verringerte Amplitude aufweisen (Hypokinese), die zu einem mikrographischen Schriftbild führt. Außerdem weisen erkrankte Patient*innen häufig eine posturale Instabilität auf. Damit ist eine verringerte Toleranz gegenüber ante-, retro oder lateropulsiven Einwirkungen gemeint. Die Patient*innen reagieren mit vermehrt notwendigen Ausfallschritten und können sogar stürzen.⁷ Die Sturzneigung wird durch andere Gangstörungen zusätzlich beeinflusst. So beklagen viele Betroffene plötzliche Unterbrechungen des Gehens („Einfrieren des Gangs“, „Freezing of gait [englisch]).⁴⁴ Zusätzlich kann die Fähigkeit, Willkürbewegungen zu initiieren und zu stoppen, vermindert sein (Akinese bzw. überschießende Bewegungen). Auch ein kleinschrittiges Gangbild sowie schlurfende Bewegungen sind häufige Beschwerden. Andere motorische Symptome sind Dystonie sowie Veränderungen der Stimme.⁴⁵ Je nach vorherrschenden motorischen Symptomen werden vier verschiedene Unterformen des idiopathischen Parkinson-Syndroms unterschieden: Der akinetisch-rigide Typ, der Äquivalenz-Typ, der Tremordominanz-Typ und der monosymptomatische Ruhetremor.⁷ Der Symptomkomplex der typischen motorischen Beschwerden beim Morbus Parkinson wird auch unter dem neurologischen Syndrom des Parkinsonismus subsumiert.⁴⁶ Er kann aber ebenso zahlreiche andere Ursachen haben (siehe Kapitel 2.2.2).

Von den motorischen Symptomen abzugrenzen sind die nicht-motorischen Symptome, die in vielfältiger Weise auftreten können. Hierzu zählen kognitive Störungen und Demenz, neuropsychiatrische Symptome, affektive Störungen und Schlafstörungen sowie eine autonome Dysfunktion und sensorische Defizite.^{47,48} Es können potentiell alle Bereiche des Körpers betroffen sein, sodass hier der multisystemische Charakter der Erkrankung offensichtlich wird, der weit über eine Bewegungsstörung hinausgeht.

2.2.6. Diagnostische Prinzipien

Die Diagnose wird in erster Linie anhand des charakteristischen klinischen Bildes gestellt. Die sogenannten UK Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria sind dabei der Goldstandard und geben ein dreistufiges Konzept zur klinischen Diagnosefindung vor. Die erste Stufe besteht im Vorhandensein eines Parkinson-Syndroms. Ein solches liegt vor, wenn neben einer Brady- bzw. Akinese mindestens ein weiteres der Symptome Rigor, Ruhetremor oder posturale Instabilität zusätzlich auftritt. Als zweite Stufe wird das Bestehen unterstützender Kriterien wie z.B. eine Asymmetrie der Symptomatik oder ein Ansprechen auf Levodopa genannt. Die dritte Stufe umfasst schließlich das Fehlen von diagnostischen Ausschlusskriterien für einen Morbus Parkinson. Dabei wird zwischen Anzeichen für ein symptomatisches (z.B. Exposition gegenüber einer symptomauslösenden Substanz) beziehungsweise für ein atypisches Parkinson-Syndrom (z.B. eine Kleinhirnsymptomatik) unterschieden.⁷ Darüber hinaus fallen zu den genannten Symptomen in der klinischen Untersuchung eine verminderte Fähigkeit auf, schnell alternierende Bewegungen auszuführen (Bradydiadochokinese) sowie ein unerschöpflicher Glabellareflex. Bei diesem schließen Betroffene unwillkürlich die Augen, wenn die Region zwischen den Augenbrauen beklopft wird.⁴⁵

Eine Herausforderung stellt die Differenzierung der Diagnose des idiopathischen Parkinson-Syndroms von nicht-idiopathischen Formen dar. Eine Möglichkeit besteht in der Durchführung eines sogenannten Levodopa- oder Apomorphin-Tests, bei dem eine Einmalgabe Levodopa bzw. Apomorphin verabreicht wird. Vor und nach der Applikation erfolgt eine Einschätzung der Symptomatik. Kommt es zu einer klinischen Verbesserung, ist der Test positiv und als Anhaltspunkt für einen Morbus Parkinson zu werten. Bei fehlender Levodopa-Responsivität ist dagegen die Diagnose einer anderen Form des Parkinson-Syndroms wahrscheinlicher.⁴⁹ Zudem ist ein positives Ergebnis ein Hinweis auf ein Ansprechen einer dopaminergen Behandlung. Außerdem wird standardmäßig eine kraniale Computertomographie (cCT) oder kraniale Magnetresonanztomographie (cMRT) durchgeführt. Diese Verfahren dienen dem Ausschluss sekundärer Parkinson-Syndrome durch strukturelle Auffälligkeiten, das idiopathische Parkinson-Syndrom zeigt einen weitgehend unauffälligen Befund.^{7,50,51} Bei klinisch unklaren Parkinson-Syndromen kommen insbesondere in der Frühphase spezifische bildgebende Verfahren zum Einsatz. Diese dienen der Integritätsprüfung des nigrostriatalen dopaminergen Systems (siehe Kapitel 2.3) und ermöglichen daher eine Aussage über das Vorliegen einer dortigen Neurodegeneration. Das Staging des idiopathischen Parkinson-Syndroms erfolgt entweder mit Hilfe der Skala nach Hoehn und Yahr²⁰ oder der Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-

UPDRS)⁵². Diese Skalen beziehen vor allem klinische Parameter ein und werden unter anderem zur Quantifizierung des Therapieerfolges herangezogen.

2.2.7. Therapie und Prognose

Die Therapie des idiopathischen Parkinson-Syndroms zielt darauf ab, das dopaminerge Defizit zu reduzieren. Als wirksamstes Medikament gilt Levodopa, eine Vorstufe (Prodrug) des Dopamins. Es wird oral appliziert und kann mit weiteren Substanzen wie Decarboxylase-Inhibitoren, Catechol-O-Methyl-Transferase-Hemmern oder Dopaminagonisten kombiniert werden.^{53,54} Levodopa zeigt initial bei einem hohen Anteil der Behandelten eine gute Wirksamkeit, besitzt aber ebenfalls ein breites Spektrum an unerwünschten Begleiterscheinungen. So treten bei über 50% der Behandelten motorische Fluktuationen und Dyskinesien auf. Darüber hinaus wird häufig eine Abnahme der Wirksamkeit („Wearing-off [englisch]) im Laufe der Zeit beobachtet, sodass immer höhere Dosen und kürzere Abstände zwischen den Einzeldosen für eine effektive Therapie notwendig werden.⁵⁵ Des Weiteren besteht die Möglichkeit, den Dopamin-Abbau mit Hilfe von Monoaminoxidase-B-Inhibitoren (MAO-B-Hemmern) zu verlangsamen.⁵⁶ Insgesamt ist die medikamentöse Therapie rein symptomatisch und kann Nebenwirkungen mit sich bringen, die einen eigenen Krankheitswert innehaben. Zeigt die medikamentöse Therapie keine ausreichende Wirksamkeit bzw. ergeben sich starke Nebenwirkungen, gibt es verschiedene interventionelle Verfahren als zusätzliche Option. Ziel ist neben einer verbesserten Symptomkontrolle eine Reduktion der Medikamentendosis. Zur Verfügung stehen unter anderem die intrajejunale Applikation von Levodopa, die subkutane Apomorphinpumpe und die tiefe Hirnstimulation des Nucleus subthalamicus.^{57,58} Außerdem wird der Ansatz von Antikörpern gegen alpha-Synuclein wissenschaftlich verfolgt. Bislang konnte allerdings keine Substanz dieser Klasse überzeugen, auch wenn der Grundgedanke sehr erfolgversprechend erscheint.^{59,60}

Das idiopathische Parkinson-Syndrom ist eine schleichend voranschreitende Erkrankung, die bislang nicht kurativ therapierbar ist. In der Rotterdam-Studie zeigte sich eine signifikante Erhöhung des Risikos für die Entwicklung einer Demenz, besonders bei Träger*innen eines Apolipoprotein E (APOE) ε2 Allels. Dabei handelt es sich um ein Protein, welches im Gehirn am Transport von Lipiden, der Immunregulation und der Modulation des Zellwachstums beteiligt ist. Der genaue Mechanismus hinter diesem Einflussfaktor ist noch unklar.⁶¹ Außerdem gibt es Hinweise darauf, dass ein Wearing-off und Dyskinesien bei Frauen früher auftreten und sie früher fortgeschrittene Krankheitsstadien erreichen. Dies lässt einen etwas schnelleren Krankheitsprogress vermuten.⁶²

2.3 Bildgebung des dopaminergen Systems

Im Folgenden werden verschiedene bildgebende Modalitäten zur diagnostischen Beurteilung des dopaminergen Systems vorgestellt. Dabei liegt der Fokus auf der FP-CIT-SPECT.

2.3.1. FP-CIT-SPECT - Funktionsweise und Untersuchungsablauf

Bei der Einzelphotonenemissionstomographie (single photon emission tomography [englisch], SPECT) handelt es sich um ein nuklearmedizinisches funktionelles bildgebendes Verfahren. Sie basiert auf dem Prinzip der Szintigraphie. Dabei erhält der/die Patient*in intravenös radioaktive Substanzen (Radiopharmaka) als Tracer, die je nach spezifischen Eigenschaften in den zu untersuchenden Regionen des Körpers aufgenommen werden.

Die Radionuklide senden Gamma-Strahlung aus, also eine Form der elektromagnetischen Strahlung, bei der ungeladene Photonen (Gamma-Quanten) emittiert werden. Die Photonen treffen zuerst auf den Kollimator, der bei einer Gamma-Kamera (Gamma-Detektor) aus einer Bleiplatte mit vielen kleinen Öffnungen besteht, ähnlich einer Lochplatte. Er wirkt wie ein Filter, der nur die senkrecht auftreffenden Photonen hindurch lässt. Diese treffen anschließend auf einen Natriumiodid-Szintillationskristall. Dabei entstehen Lichtblitze, deren Energie proportional zur Energie der Photonen ist. Das Lichtsignal gelangt nun über ein Lichtleitersystem zu einem sogenannten Sekundärelektronenvervielfacher (SEV, auch Photomultiplier genannt). Dieser verstärkt die Signale und wandelt sie in elektrische Impulse um. Deren Höhe ist proportional zu der Menge der empfangenen Lichtsignale. Entsprechend der Menge der aufgetroffenen Blitze lassen sich also Rückschlüsse auf die lokale Nuklidbelegung einer Körperregion ziehen. Das Ausgangssignal trifft auf die Auswerteelektronik, die die Positionskoordinaten bestimmt und diese als entsprechenden Bildpunkt in eine Matrix einträgt. Diese Information wird durch einen Rechner in ein farbvisualisiertes Bild bzw. in ein Graustufenbild überführt.⁶³ Abbildung 2 zeigt den schematischen Aufbau eines Gamma-Detektors. Aus Gründen der besseren Übersicht ist nur ein senkrecht auftreffendes Photon bzw. ein Lichtblitz dargestellt.

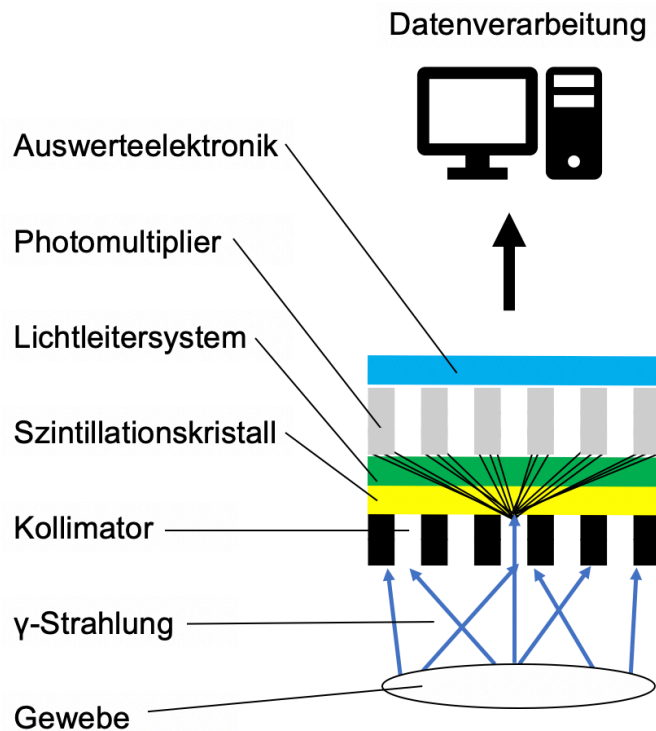


Abbildung 2: Aufbau einer Gamma-Kamera

Bei der SPECT sind im Gegensatz zur Szintigraphie die Detektoren der Gammakamera beweglich, sodass Signale aus unterschiedlichen Richtungen des Raumes detektiert werden können. Dabei rotieren einer bis drei Detektorköpfe langsam um den/die Patient*in und nehmen aus verschiedenen Positionen einzelne Bilder zu unterschiedlichen Zeitpunkten auf. Diese werden anschließend zu einem Datensatz verrechnet. Die SPECT ermöglicht eine dreidimensionale Darstellung der Nuklidbelegung im Gegensatz zur Szintigraphie, die eine planare Aufnahme liefert. Das Verfahren gibt also nicht nur Aufschluss darüber, ob eine Stoffwechselaktivität in einer Körperregion vorhanden ist, sondern erlaubt eine zusätzliche Differenzierung, wie diese Aktivität vor Ort verteilt ist und wie sie sich über die Zeit ändert.⁶⁴ Wie viele andere medizinische Bilddaten werden SPECT-Aufnahmen i. d. R. im DICOM-Format (Digital Imaging and Communications in Medicine, [englisch]) gespeichert. Es ist, wie bereits der Name suggeriert, für die Anwendung in der medizinischen Bildgebung entwickelt worden und dementsprechend außerhalb der Medizin wenig geläufig. Das DICOM-Format ist nur mit speziellen medizinischen Bildverarbeitungsprogrammen zu öffnen.⁶⁵

Die Dopamintransporter-Einzelphotonenemissionstomographie („dopamine transporter single photon emission tomography“ [englisch], DAT-SPECT) bezeichnet ein SPECT-Verfahren, welches die lokale Verfügbarkeit des präsynaptischen Dopamintransportersystems im Gehirn gezielt darstellt. Diese dient als Korrelat der Integrität der nigrostriatalen dopaminergen

Neurone. Der verwendete Tracer heißt [^{123}I]N- ω -fluoropropyl-2 β -carbomethoxy-3 β -(4-iodophenyl) nortropane, in Kurzform [^{123}I]FP-CIT oder Ioflupane, und ist ein Kokain-Derivat. Die Substanz ist in Deutschland ausschließlich über den Hersteller General Electric (GE) unter dem Handelsnamen DaTSCANTM erhältlich. Da es sich um einen Iod-haltigen Tracer handelt, werden vor der Injektion Natriumperchlorat-Tropfen verabreicht. Natriumperchlorat hemmt kompetitiv die Aufnahme des Iods in die Schilddrüse, sodass das Organ vor der Strahlenbelastung durch die radioaktive Substanz geschützt wird. Nach der intravenösen Applikation des ^{123}I -FP-CIT folgt eine drei- bis vierstündige Wartezeit, in der der Tracer an die Transporter der dopaminergen Neurone bindet (Uptake-Phase). Der/die Patient*in kann sich in dieser Zeit frei bewegen. Anschließend erfolgt die Akquisition der Aufnahmen durch die Gammakameras. Dieser Prozess kann bis zu 60 Minuten in Anspruch nehmen, in denen die zu untersuchende Person möglichst ruhig in dem Scanner liegen muss. Ein starker Tremor, vor allem im Bereich des Kopfes, kann zu Artefakten der Aufnahmen führen, die die diagnostische Aussagekraft erheblich einschränken können. Nach dem Ende der Untersuchung sollten Patient*innen ausgiebig Flüssigkeit zu sich nehmen und häufig die Harnblase entleeren, um eine schnelle Ausscheidung der Substanz und damit eine Reduktion der Strahlenexposition zu gewährleisten.⁶⁶ Bei der DAT-SPECT gibt es zwei Formen der Bildrekonstruktion: Die iterative Rekonstruktion und die gefilterte Rückprojektion (filtered back projection, abgekürzt FBP, [englisch]). Im Folgenden wird in Abbildung 3 eine Aufnahme eines Patienten in iterativer Rekonstruktion im Vergleich mit der FBP dargestellt.

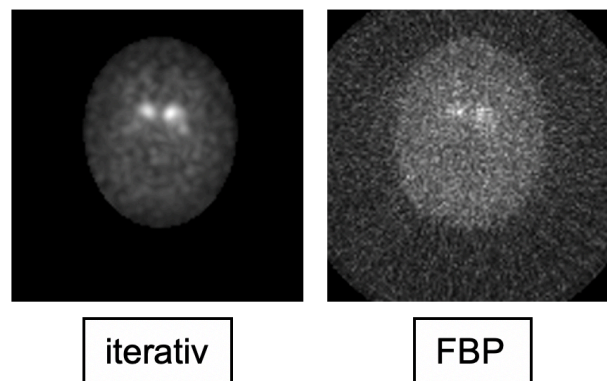


Abbildung 3: Iterative Rekonstruktion und gefilterte Rückprojektion im Vergleich

Die Dopamin-Transporter-SPECTs werden zwar als dreidimensionale Datensätze akquiriert, zur Befundung werden sie aber im Rahmen der klinischen Routine auf zweidimensionale Datensätze, sogenannte Maximumintensitätsprojektionen (MIPs), reduziert. Im Ergebnisteil ist ein entsprechendes Beispiel zu sehen (Kapitel 4.2, Abbildung 21).

2.3.2. FP-CIT-SPECT - Indikation und diagnostische Bewertung

Die DAT-SPECT wird bei klinisch unklaren Parkinson-Syndromen regelhaft als zusätzliches diagnostisches Mittel eingesetzt.⁶⁷ Die häufigste Indikation stellt die Differenzierung zwischen neurodegenerativ und nicht-neurodegenerativ bedingten Parkinson- oder Tremor-Syndromen dar. Die Auswertung erfolgt semiquantitativ und visuell. Dabei werden sogenannte Regions of Interest (ROIs, [englisch]) definiert, indem die nuklidbelegten Areale ausgemessen werden. Je nach Institution werden unterschiedliche ROIs verwendet, gemeinsamer Standard ist meist die Erstellung von ROIs für das gesamte Striatum sowie separat für das Caput des Nucleus caudatus und das Putamen. Manchmal werden zusätzliche ROIs für anteriores und posteriores Putamen bestimmt, die durch ihre kleine Größe allerdings eine höhere Anfälligkeit für Fehlinterpretationen bergen. Anschließend werden die ROIs in Bezug zu einer Referenzregion gesetzt und verschiedene Ratios berechnet, z.B. das Verhältnis von linkem zu rechtem Striatum oder die Putamen/Caudatus-Ratio. Für die verschiedenen Ratios existieren bislang keine festen Grenzwerte. Ihre isolierte Betrachtung bietet eine eher geringe diagnostische Aussagekraft.⁶⁶

Außerdem erfolgt eine visuelle Analyse der Nuklidbelegung. Physiologischerweise erscheint das Putamen in der FP-CIT-SPECT als komma-förmige Struktur in der Transversalebene. Liegt ein dopaminerges Defizit vor, nimmt die Belegung eine ovale bis rundliche Struktur an. Für ein idiopathisches Parkinson-Syndrom ist ein asymmetrisches Belegungsmuster zwischen linkem und rechtem Striatum mit putaminaler Betonung charakteristisch.⁶⁶ Dabei ist die klinisch stärker betroffene Seite kontralateral zur Verminderung der Nuklidbelegung. Bei einer rechtsdominanten klinischen Symptomatik wäre bei einem Morbus Parkinson also eine geringere Nuklidaufnahme des linken Putamens im Vergleich zum rechten zu erwarten. Die Asymmetrie bleibt sowohl klinisch als auch in der DAT-SPECT bei den meisten Patient*innen bis in späte Erkrankungsstadien bestehen.⁶⁸ Bei atypischen Parkinson-Syndromen können dagegen auch homogenere Muster des DAT-Defizits auftreten wie z.B. eine eher seitensymmetrische DAT-Bindung bei der PSP. Insgesamt müssen die visuelle und die semiquantitative Analyse zu übereinstimmenden Ergebnissen kommen, um eine stichhaltige Aussage treffen zu können.⁶⁶ In Abbildung 4 sind ein unauffälliger und ein auffälliger Bildbefund in iterativer Rekonstruktion zu sehen.

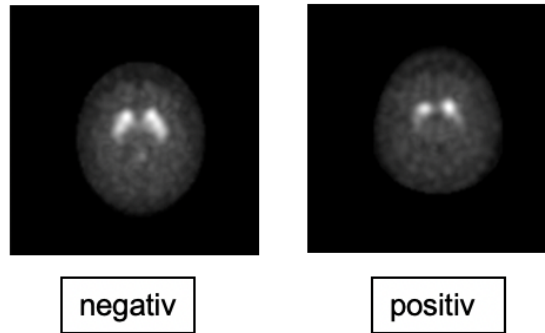


Abbildung 4: Unauffällige und auffällige DAT-SPECT im Vergleich

Eine ätiologische Einordnung oder eine Differenzierung zwischen typischem und atypischem Parkinson-Syndrom ist nicht möglich, sodass die DAT-SPECT nur in Zusammenschau mit Anamnese und weiteren Befunden zu einer konkreten Diagnose führen kann.^{66,69} Ein Normalbefund ist laut Postuma et al. als absolutes Ausschlusskriterium eines Morbus Parkinson zu werten.⁷⁰ Bestehen trotzdem klinische Symptome eines Parkinson-Syndroms, so spricht man von einem SWEDD („subject without evidence of dopaminergic deficit“, [englisch]).⁷¹

2.3.3. ¹⁸F-DOPA-PET

¹⁸F-Fluorodopa war der erste PET-Ligand, der zur Darstellung der Integrität des dopaminergen Transporter-Systems verwendet wurde. Die ¹⁸F-Dopa-PET zeigt wie die FP-CIT-SPECT den Zustand der präsynaptischen dopaminergen Neuronen an.^{68,72} Bei einem Untergang dopaminergener Neuronen zeigt sich im Bereich des Putamens eine verringerte Aufnahme des Tracers. Dies kommt durch eine reduzierte quantitative Aktivität der Dopa-Decarboxylase zustande, die ¹⁸F-Dopa zu ¹⁸F-Dopamin metabolisiert.⁷³ Die ¹⁸F-Dopa-PET eignet sich damit ebenfalls zur Differentialdiagnostik des idiopathischen Parkinson-Syndroms. Insbesondere bei der Abgrenzung zu nicht-neurodegenerativ bedingten Erkrankungen wie dem essentiellen Tremor (ET), medikamentös induzierten Parkinson-Syndromen und zur Dopa-responsiven Dystonie zeigt das Verfahren eine hohe Spezifität und Sensitivität.^{74,75} Bei atypischen Parkinson-Syndromen kann die Untersuchung Hinweise zur Diagnosesicherung geben. Eine detaillierte Einordnung ist jedoch nur mit Hilfe weiterer Modalitäten möglich, da den Krankheitsbildern ähnliche Prozesse der nigro-striatalen Degeneration zugrunde liegen.⁷⁶ Eine Besonderheit stellt die niedrige Test-Retest-Variabilität der ¹⁸F-Dopa-PET dar.⁷⁷ Sie eignet sich daher besonders zur Verlaufskontrolle der Progression bzw. zur Kontrolle des Therapieerfolges von Personen mit idiopathischen oder atypischen Parkinson-Syndromen. Außerdem bietet sie durch eine hohe örtliche Auflösung die Möglichkeit, dopaminerge Defizite exakt zu quantifizieren. Diese Eigenschaft ist insbesondere bei der Detektion von exogen

bedingten Läsionen oder genetisch bedingten Krankheitsbildern von Vorteil. Ein Nachteil liegt in den im Vergleich zur FP-CIT-SPECT höheren Kosten.⁷⁸

2.3.4. IBZM-SPECT

Die SPECT mit ¹²³I-Iodobenzamid (¹²³I-IBZM) ermöglicht eine quantitative Darstellung des postsynaptischen D2-Rezeptor-Status. Die Hauptindikation besteht in der Differenzialdiagnostik des idiopathischen Parkinson-Syndroms, bei dem die Integrität des postsynaptischen D2-Status meist erhalten ist. Dagegen zeigen sich bei anderen neurodegenerativ bedingten Parkinson-Syndromen wie z.B. der MSA, PSP oder CBD pathologische Veränderungen in dieser Untersuchung. Die Differenzierung dieser Krankheitsbilder vom Morbus Parkinson ist eine Eigenschaft, die die ¹²³I-IBZM-SPECT im Gegensatz zur FP-CIT-SPECT zeigt. Weitere Einsatzgebiete sind die Beurteilung der D2-Rezeptor-Blockade unter Behandlung mit Antipsychotika, die Abklärung eines Verdachts auf eine Chorea Huntington oder einen Morbus Wilson sowie die medikamentöse Therapieplanung bei Hypophysenadenomen.⁷⁹⁻⁸¹ Ein Nachteil der Methode ist die hohe Anfälligkeit für verfälschte Untersuchungsergebnisse durch dopaminerge Medikamente oder durch Dopaminantagonisten. Eine korrekte Interpretation der ¹²³I-IBZM-Aufnahmen ist von daher nur nach sorgfältiger Vorbereitung, gegebenenfalls durch Absetzen entsprechender Medikamente, möglich. Dafür bieten Untersuchungen des zerebralen Glukosestoffwechsels mittels ¹⁸F-Fluor-2-desoxy-D-Glucose-PET (¹⁸F-FDG-PET) und MRT-basierte Verfahren eine höhere Sensitivität und Spezifität bei der Differentialdiagnostik von Parkinson-Syndromen.^{68,79}

2.3.5. MIBG-Szintigraphie

Die Iod-123-Metaiodidbenzylguanidin-Szintigraphie (MIBG-Szintigraphie) ist ein nuklearmedizinisches Verfahren, das in der Tumordiagnostik, aber auch wie die IBZM-SPECT zur Differentialdiagnose von klinisch unklaren Parkinson-Syndromen eingesetzt wird.⁷ Das verwendete Pharmazeutikum ist das Noradrenalin-Analogon MIBG. Es wird mit meist mit dem Radioisotop Iod-123, manchmal auch mit Iod-131, markiert. Iod-123-MIBG bietet gegenüber Iod-131-MIBG die Vorteile einer geringeren Strahlenexposition, des zusätzlichen Einsatzes einer SPECT oder SPECT/CT sowie eine schnellere Verfügbarkeit der Untersuchungsergebnisse.⁸² Insbesondere zur Abgrenzung der MSA eignet sich dieses Verfahren.⁷ Interessanterweise wird dabei keine Aufnahme des Gehirns, sondern des Herzens angefertigt. Der Hintergrund besteht darin, dass beim fortgeschrittenen idiopathischen Parkinson-Syndrom das periphere autonome Nervensystem von neurodegenerativen Prozessen betroffen ist. Korrelat der verringerten noradrenergen Innervation des Myokards ist eine verringerte kardiale MIBG-Aufnahme im Vergleich zum übrigen Mediastinum (verringertes

Herz/Mediastinum-Verhältnis). Bei der MSA liegt ebenso eine schwere autonome Dysfunktion vor, jedoch liegt die Ursache zentral und nicht peripher. Dementsprechend ist die periphere autonome Innervation des Herzmuskels intakt.^{83,84} Dies äußert sich in einer regelrechten kardialen und mediastinalen Aufnahme von MIBG und in einem physiologischen Herz/Mediastinum-Quotient. Im frühen Stadium des Morbus Parkinson (Hoehn & Yahr-Stadium I) kann die postganglionäre kardiale Innervation noch unbeeinträchtigt sein, sodass die Methode in der frühen diagnostischen Differenzierung weniger aussagekräftig ist.^{85,86} Außerhalb der neurologischen Diagnostik finden sich weitere Einsatzgebiete der MIBG-Szintigraphie in der Onkologie. Dort stellt sie eine wichtige Säule dar in der Beurteilung von neuroendokrinen Tumoren und Tumoren, die ihren Ursprung im sympathischen Nervensystem haben. Hierzu zählen vor allem Neuroblastome, adrenale und extraadrenale Phäochromozytome und deren Metastasen. In der Regel wird auch dort das MIBG bevorzugt mit Iod-123 statt mit Iod-131 markiert. Dies erklärt sich unter anderem durch das oft junge Alter der Patient*innen. So treten z.B. Neuroblastome meistens bei Kindern auf, sodass einer möglichst geringen Strahlenexposition eine besondere Bedeutung zukommt.⁸²

2.3.6. Transkranielle Sonographie

Die transkranielle B-Bild-Sonographie ist eine noch recht neue Modalität in der neuronalen Bildgebung. Bei diesem Verfahren wird das Hirnparenchym durch ein temporales Knochenfenster per Ultraschall untersucht. Neben der Abgrenzung von einem intakten ZNS ermöglicht die transkranielle Sonographie die Differenzierung zwischen einem idiopathischen Parkinson-Syndrom und anderen neurologischen Erkrankungen wie dem Restless-Legs-Syndrom oder sekundären und atypischen Parkinson-Syndromen.^{87,88} Charakteristisch zeigt sich bei bis zu 90% der Personen mit einem Morbus Parkinson eine Hyperechogenität der Substantia nigra. Bereits im präsymptomatischen Stadium kann sich diese Veränderung zeigen, sodass sie zur Identifizierung von Menschen mit einem erhöhten Risiko für einen Morbus Parkinson dienen kann.⁸⁹ Neben einem hohen positiven prädiktiven Wert bietet die Methode den Vorteil, dass sie mit geringem technischem Aufwand, schnell und günstig durchzuführen ist und keinerlei Strahlenexposition mit sich bringt. Zusätzlich ist sie aus praktischer Sicht leicht erlernbar und non-invasiv. Die Größe des hyperechogenen Ultraschallsignals korreliert jedoch nicht mit dem Schweregrad der Erkrankung. Außerdem ist die transkranielle Sonographie nicht immer möglich, da nicht alle Menschen ein temporales Knochenfenster besitzen.⁹⁰

2.4 Maschinelles Lernen

2.4.1. Überblick / Grundlagen

Das maschinelle Lernen („machine learning“, [englisch]) beschreibt einen Teilbereich der künstlichen Intelligenz („artificial intelligence“, [englisch]). Diese ist definiert als ein Gebiet der Informatik, das sich mit der Fähigkeit von Maschinen beschäftigt, (menschliches) intelligentes Verhalten zu simulieren bzw. zu imitieren.^{91,92} Als Intelligenz bezeichnet die Encyclopædia Britannica die Befähigung, sich an sich verändernde Gegebenheiten und Rahmenbedingungen anzupassen.⁹³ Dazu zählen laut dem English Oxford Living Dictionary unter anderem Merkmale wie die visuelle Wahrnehmung, die Wiedererkennung z.B. von Sprache und das Treffen von Entscheidungen.⁹⁴ Zur künstlichen Intelligenz existieren zahlreiche weitere Definitionen, denen aber zusammenfassend gemeinsam ist, dass der Fokus sich auf die Imitation menschlichen Verhaltens begrenzt. Der Computer wird trotzdem als Maschine und nicht als Mensch wahrgenommen.

Das maschinelle Lernen zeichnet sich dadurch aus, dass Programme anhand beispielhaft vorgegebener Daten selbstständig lernen und daraufhin Entscheidungen und Vorhersagen treffen können. Die Computernetzwerke erkennen in einem Trainingsdatensatz Muster, Regel- und Gesetzmäßigkeiten und wenden diese Erkenntnisse an einem neuen, unbekanntem Testdatensatz an. Dieser Prozess wird auch als Generalisierung bezeichnet. Dabei werden die entsprechenden Algorithmen eigenständig durch Selbstlernprozesse fortlaufend optimiert. Es werden nach dem Vorbild des menschlichen Lernens Konklusionen gezogen und daraus Handlungen für die Lösung eines definierten Problems abgeleitet, die normalerweise ein Mensch vollziehen würde. Der häufig aufwendige manuelle Prozess des Erstellens von Regeln und Modellen entfällt.^{95,96} Maschinelles Lernen kann in den unterschiedlichsten Bereichen eingesetzt werden, z.B. im Marketing, in der Telekommunikation oder in der Automobilbranche. Es werden also verschiedenste Daten in unterschiedlichen Dateiformaten verarbeitet. Um ein ausgeglichenes Verhältnis zwischen dem Aufwand für die Entwicklung entsprechender Machine Learning-Anwendungen und den potentiellen Einsatzgebieten zu schaffen, ist eine Kompatibilität mit häufigen Dateiformaten wie z.B. GIF (Graphics Interchange Format, [englisch]) oder JPEG (Joint Photographic Experts Group, [englisch]), sinnvoller und wirtschaftlicher als der Fokus auf seltene Formate. Die meisten Anwendungen sind nicht primär für den medizinischen Bereich gestaltet worden, sodass die Vereinbarkeit mit den teilweise speziellen Dateiformaten eingeschränkt ist. Eine Herausforderung stellen außerdem dreidimensionale Formate dar, da diese Dateien eine für viele Algorithmen zu umfangreiche Dateigröße aufwiesen. Daher wird meist eine zweidimensionale Struktur der Datensätze angestrebt, da bei diesen die aufzubringende Rechenleistung und der notwendige

Arbeitsspeicher für die meisten Rechner im überschaubaren Rahmen bleiben. Zu große Datenmengen können zu unverhältnismäßig hohem zeitlichem Rechenaufwand führen.

2.4.2. Varianten und Anwendungsgebiete

Beim maschinellen Lernen lassen sich drei verschiedene untergeordnete Kategorien unterscheiden: Das überwachte Lernen („supervised learning“, [englisch]), das unüberwachte Lernen („unsupervised learning“, [englisch]) und das bestärkende Lernen („reinforcement learning“, [englisch]).

Das überwachte Lernen zeichnet sich dadurch aus, dass der Datensatz zum Trainieren des Netzwerkes gekennzeichnete („gelabelte“) Daten enthält. Dies bedeutet, dass die Daten kategorisiert und mit einer Information über die Zugehörigkeit zu einer der Kategorien versehen worden sind, z.B. befinden sich alle Dateien mit einem bestimmten Merkmal in einem Ordner oder sie sind durch ein Attribut im Dateinamen gekennzeichnet. Die Zusammenhänge zwischen den Datenfeldern sind verstanden, das Programm soll diese nun eigenständig extrahieren. Die erwünschten Ausgabewerte sind dementsprechend bekannt und werden Labels genannt. Während des Lernprozesses „übt“ der Algorithmus das Treffen von Vorhersagen an den Lerndaten. Da die korrekten Labels bekannt sind, besteht die Möglichkeit einer direkten Rückmeldung über die Richtigkeit der Prognose, daher die Bezeichnung „überwachtes Lernen“.⁹⁵ Einsatzgebiet des überwachten Lernens sind Problemstellungen der Klassifizierung, die die Einteilung des Datensatzes in verschiedene Klassen beschreibt. Das Netzwerk lernt anhand der Daten, welche Merkmale welche Klasse kennzeichnen und erkennt diese Merkmale dann in unbekanntem Daten wieder. So können Voraussagen über die Kategoriezugehörigkeit neuer Daten getroffen werden. Ein typisches und im Alltag häufig anzutreffendes Anwendungsgebiet des überwachten Lernens ist der Einsatz von Filtern, z.B. von E-Mail-Spam-Filtern. Dort hat ein Algorithmus mittels eines Satzes von E-Mails, die entweder als „Spam“ oder „Nicht-Spam“ gekennzeichnet worden sind, die jeweiligen Charakteristika extrahiert und kann neu eintreffende E-Mails entsprechend klassifizieren. Ein anderes Beispiel sind Filter, die in Online-Shops verwendet werden, um angebotene Produkte z.B. nach Farbe zu sortieren. Die Ausgabewerte der Klassifizierung sind diskret, es existieren also abzählbar viele Kategorien. Gibt es wie bei einem Spam-Filter zwei verschiedene mögliche Kategorien, handelt es sich um eine binäre Klassifizierung. Eine weitere Unterdisziplin des überwachten Lernens stellt die Lösung von Fragestellungen der Regression dar. Im Gegensatz zu den Ausgabewerten der Klassifizierung sind jene der Regression stetig, sie können also unendlich viele Klassen annehmen. Das Ergebnis gibt die Wahrscheinlichkeit für den Eintritt eines Ereignisses bzw. für die Zugehörigkeit zu einer Klasse an. Ein Beispiel für ein mögliches Regressionsproblem ist die Wettervorhersage.⁹⁵

Beim unüberwachten Lernen ist der Datensatz dagegen nicht mit Labels versehen. Ziel ist, dass der Lernalgorithmus vorher unbekannte Strukturen in unkategorisierten Daten erkennt und somit eine Gruppierung vornehmen kann. Hauptsächlich verwendet wird das unüberwachte Lernen zur Assoziationsanalyse („association rule learning“, [englisch]) und zur Ballungsanalyse, besser bekannt als Clustering. Die Assoziationsanalyse ist eine Methode, die nach Korrelationen zwischen verschiedenen Aktionen sucht und daraus sogenannte starke Regeln ableitet. Ein typisches Einsatzgebiet ist die Warenkorbanalyse zur Vorhersage von weiteren Online-Käufen. So ist es z.B. wahrscheinlich, dass der Käufer eines Bettes sich ebenfalls für einen Nachttisch interessieren könnte. Basierend auf dieser Vorhersage werden dann gezielte Angebote präsentiert. Clustering dagegen bedeutet, dass durch den Algorithmus Daten nach Attributen in vorher unbekannte Gruppen, sogenannte Cluster, eingeteilt werden. Dabei haben die Daten innerhalb eines Clusters untereinander verschiedene Merkmale gemeinsam, die sie wiederum von Daten, die anderen Clustern zugehörig sind, unterscheiden. Je homogener die Cluster in sich sind und je inhomogener sie untereinander sind, desto größer ist die Trennschärfe. Diese Methode wird z.B. zum Einsatz gezielter Marketingmaßnahmen benutzt, indem die Verbraucher anhand ihres Verhaltens in verschiedene Zielgruppen eingeteilt werden. Dies ist z.B. der Fall bei Werbemaßnahmen auf der Videoplattform YouTube, die sich an den zuvor angesehenen Videos orientieren. In der Medizin wird diese Methode verwendet, indem Symptomkomplexe von Patient*innen in Cluster eingeteilt werden. Darauf basierend können Erkrankungen mit einer großen Spannweite an Beschwerden vorher unbekannte Symptome zugeordnet werden. Auch können Personen mit einer vorher unklaren Symptomatik durch Clustering einer Diagnose zugeführt werden.⁹⁷

Das bestärkende Lernen als dritte Variante des maschinellen Lernens zielt darauf ab, ein System zu erstellen, welches die eigene Leistung durch Interaktion mit seiner Umgebung optimiert. Dieses System wird auch als Agent bezeichnet. Der Agent erhält über ein Belohnungssignal Rückmeldungen über die gegenwärtige Beschaffenheit seiner Umgebung, es besteht von daher eine gewisse Ähnlichkeit zum überwachten Lernen. Dieses Feedback besteht jedoch nicht in einer richtigen Kategoriebezeichnung oder einem korrekten Zahlenwert. Es ist eher indirekt über eine Belohnungsfunktion definiert, die Aufschluss darüber gibt, wie gut eine Aktion war. Der Agent erhält durch Interaktion mit seinem dynamischen Umfeld die Information, welche Handlungen besonders belohnt oder bestraft werden. Dabei sind geplante Aktionen genauso möglich wie ein Ansatz des bloßen Ausprobierens („Try-and-Error“ [englisch]). Einsatzgebiete des verstärkenden Lernens finden sich beispielsweise in der Robotik und bei Computerspielen. Dort können Spieler*innen sowohl gegen reale wie auch gegen virtuelle Gegner antreten. Bei einem Schachcomputer beispielsweise nimmt der Agent

Züge vor und evaluiert anschließend die Umgebung, die der Stellung auf dem Schachbrett entspricht. Die Belohnung wird am Ende des Spiels als Sieg oder Niederlage rückgemeldet.^{95,98} Abbildung 5 fasst die verschiedenen Varianten des maschinellen Lernens als Übersicht zusammen.

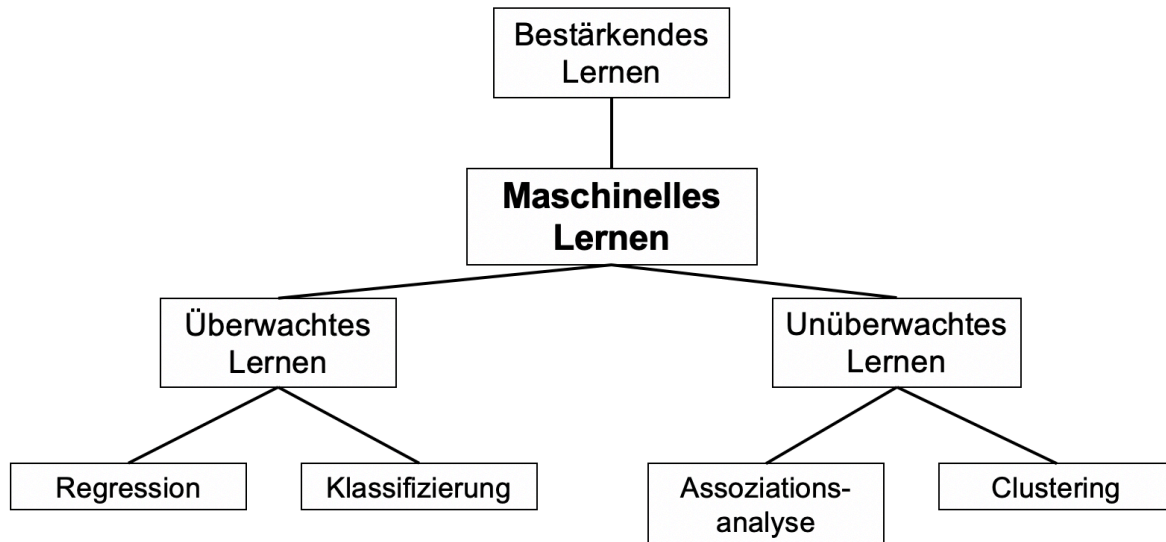


Abbildung 5: Varianten des maschinellen Lernens

2.4.3. Tiefes Lernen

Das tiefe Lernen („Deep Learning“, [englisch]) stellt eine spezifische Untergruppe des maschinellen Lernens dar. Abbildung 6 nimmt eine Einordnung in den Gesamtkontext der künstlichen Intelligenz vor. Tiefes Lernen ist dadurch charakterisiert, dass die Informationsverarbeitung und damit der Lernprozess durch ein schichtartig aufgebautes Modell schrittweise erfolgt. Dabei nimmt mit jeder Schicht die Relevanz der dargestellten Informationen zu. Bei den meisten verwendeten Algorithmen im Bereich des tiefen Lernens handelt es sich um sogenannte künstliche neuronale Netzwerke. Auf den detaillierten Aufbau wird im nächsten Unterkapitel eingegangen. Das tiefe Lernen eignet sich besonders für die Analyse großer Datenmengen.^{99,100} Dabei werden neben tiefen neuronalen Netzwerken („deep neural networks“, [englisch]) Entscheidungsbäume („Random Decision Forests“, [englisch]) verwendet. Sie zeichnen sich durch einen ebenfalls schichtweisen Aufbau aus, bestehen aber aus Entscheidungsbäumen, wie sie auch aus dem Bereich der Statistik bekannt sind („Random Decision Trees“, [englisch]).¹⁰¹

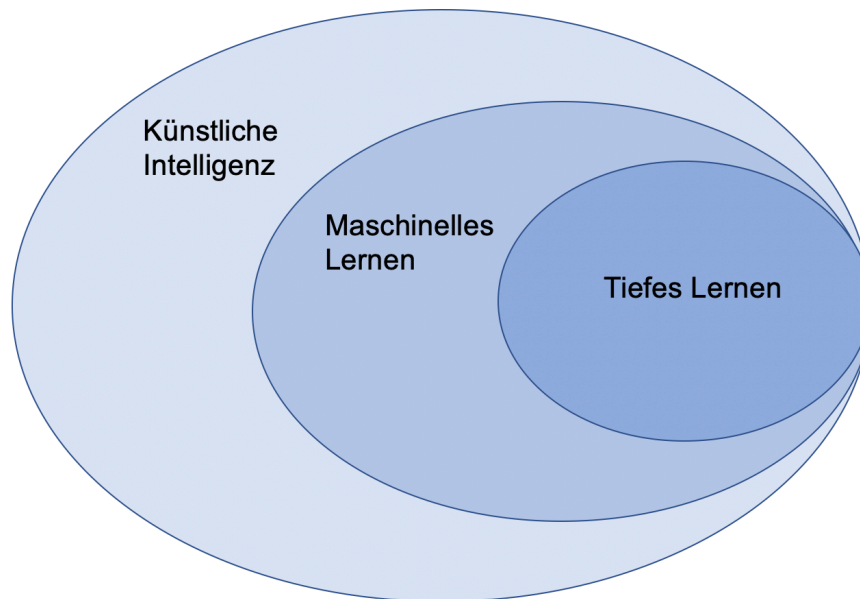


Abbildung 6: Tiefes Lernen im Gefüge der künstlichen Intelligenz nach Chollet (2018)

Eingesetzt wird tiefes Lernen, um aus umfangreichen Datensätzen Muster oder Trends zu extrahieren. Exemplarische Einsatzgebiete sind die Gesichts- oder Spracherkennung, wie sie auf vielen mobilen Endgeräten anzutreffen sind. Bei der Spracherkennung beispielsweise können Systeme durch tiefes Lernen kontinuierlich ihren Wortschatz um neue Wörter und Redewendungen ergänzen. Dieses Beispiel illustriert, dass Deep Learning-Algorithmen in der Lage sind, selbstständig Neues zu lernen, ohne dass eine externe Intervention in den Vorgang des Lernens notwendig ist. Dabei verknüpfen sie bereits Gelerntes mit neuen Informationen und generieren dadurch neue Lernprozesse. Der Mensch stellt der Maschine lediglich die zum Lernen notwendigen Datensätze zur Verfügung und sorgt für eine Dokumentation der verschiedenen Prozesse. Beim klassischen maschinellen Lernen greift man im Gegensatz dazu steuernd in die Informationsanalyse und in die Entscheidungsvorgänge ein und hat damit aktiven Einfluss auf die resultierenden Lernprozesse. Die Ergebnisse dieser Prozesse sind deshalb relativ gut nachvollziehbar, während es beim tiefen Lernen oftmals nicht möglich ist, die Bemessungsgrundlage einer durch die Maschine getroffenen Entscheidung zurückzuverfolgen.^{99,100}

2.4.4. Künstliche neuronale Netzwerke

Wie der Name bereits suggeriert, stellen künstliche neuronale Netzwerke ein abstrahiertes Modell der Funktionsweise des menschlichen Gehirns dar. So wie im Gehirn die Neuronen über ihre Dendriten und Axone miteinander verbunden sind, sind künstliche neuronale Netzwerke ebenfalls aus zahlreichen Knotenpunkten aufgebaut. Sie sind untereinander verknüpft, wobei diese Verknüpfungen durch Lernprozesse modifiziert werden können.

Während die Weitergabe von Informationen im lebenden Organismus über elektrische Impulse und die Freisetzung von Neurotransmittern erfolgt, basiert sie in den künstlichen Pendants auf Berechnungen durch den Computer. Hier eignen sich insbesondere Vektor- bzw. Matrizenrechnungen, da sich Bilddaten mittels dieser Konstrukte besonders gut darstellen lassen.⁹⁹

Aufgebaut sind künstliche neuronale Netzwerke aus künstlichen Neuronen, die auch als Knoten bezeichnet werden. Sie erhalten Eingangssignale in Form eines Eingangsvektors. Diese Information innerhalb eines Knotens unterliegt einem Gewichtungsprozess durch interne Parameter. Diese Parameter werden folglich auch als „weights“ [englisch] („Gewichtungen“) benannt. Konkret wird dieser Prozess durch die Übertragungsfunktion h definiert. Dabei wird der Eingangsvektor x mit einem Gewichtsvektor ω multipliziert. Liegt ein Bias b vor, wird dieser addiert und anschließend die Summe gebildet.¹⁰² Bei Neuronen im lebendigen Organismus wird erst bei Überschreitung einer bestimmten Schwellenspannung ein Aktionspotential ausgelöst und so die Erregung weitergeleitet. Der Hintergrund besteht darin, dass unterschwellige Reize nicht fortgeleitet werden, um eine Übererregung des Nervensystems mit unbedeutenden Reizen zu vermeiden. Bei künstlichen Neuronen existiert in Form der Aktivierungsfunktion φ ein ähnlicher Mechanismus. Dabei handelt es sich um eine mathematische Funktion, z.B. eine Sigmoidfunktion, in die das Ergebnis der Übertragungsfunktion eingesetzt wird. Das Ergebnis der Aktivierungsfunktion wiederum entspricht der Information, die der Knoten als Ausgangssignal y an den/die Knoten der folgenden Schicht ausgibt. Dabei bestimmt diese Funktion, ob und wenn ja in welchem Ausmaß Ausgangssignale an die Knoten der Folgeschicht weitergegeben werden. Ihr kommt also eine filterähnliche Funktion zu.¹⁰² Das Programm gibt eine vermutete Klassenzugehörigkeit an und kontrolliert diese anschließend anhand des Labels. Erwies sich die Vermutung als korrekt, werden die entsprechenden technischen Parameter stärker gewichtet. War sie inkorrekt, wurden die Parameter schwächer gewichtet. Abbildung 7 illustriert den Aufbau eines Knotens nach Deru und Ndiaye.¹⁰³

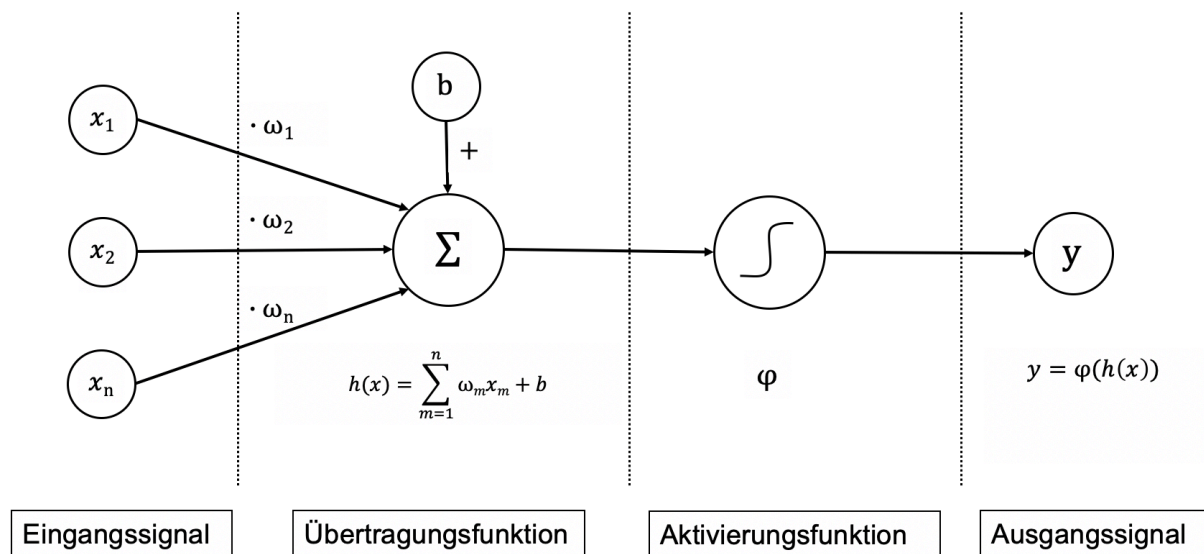


Abbildung 7: Aufbau eines Knotens nach Deru und Ndiaye (2019)

Man unterscheidet Eingangs- von Ausgangsneuronen, aus denen jeweils eine Eingangs- bzw. eine Ausgangsschicht aufgebaut ist. Die Breite der Schicht wird durch die Anzahl der künstlichen Neuronen pro Schicht bestimmt. Besteht ein solches Netz aus jeweils einer Ein- und Ausgabeschicht, so spricht man von einem einfachen künstlichen neuronalen Netzwerk. Zwischen Ein- und Ausgangsschicht können sich allerdings auch mehrere Schichten von Zwischenneuronen befinden. Liegt mindestens eine solche intermediäre Schicht vor, wird es als tiefes neuronales Netzwerk bezeichnet. Die Ein- bzw. Ausgabedaten der Zwischenschichten sind für den menschlichen Nutzer nicht ersichtlich. Sie werden daher auch als „versteckt(e) Schichte(n)“ bezeichnet („hidden layer(s)“, [englisch]). Die Ausgabedaten der Ausgangsschicht sind dagegen sichtbar. Sie entsprechen den Wahrscheinlichkeiten für die Zugehörigkeit zu verschiedenen Klassen, die vorgegeben oder nicht vorgegeben sein können. Je nachdem, wie die Gewichtungen in den einzelnen Knoten der unterschiedlichen Schichten festgelegt sind, liefert die Ausgangsschicht ein mehr oder weniger zutreffendes Ergebnis. Dementsprechend ist die Modifikation dieser internen Parameter ein bedeutender Bestandteil für das Erzielen eines möglichst guten Ergebnisses.^{102,104}

Die Tiefe eines Netzwerkes wird über die Anzahl der Schichten definiert: Je mehr Schichten vorhanden sind, desto höher ist die Tiefe. Daraus ergibt sich das Adjektiv „tief“ in der Bezeichnung „tiefes Lernen“. Mit zunehmender Anzahl an künstlichen Neuronen und Schichten steigt auch die Komplexität, mit der Sachverhalte untersucht werden können.¹⁰⁰ Außerdem gibt es künstliche neuronale Netzwerke in verschiedenen Fraktionen. Die Modellfraktion ist definiert als der prozentuale Anteil an der maximalen Ausdehnung des Netzwerkes, d. h. die Fraktionsversion 0.75 entspricht einer relativen Größe von 75% der

Ausdehnung. Der Vergleich dieser verschiedenen Netzwerkausführungen ist von Interesse, da bei maximaler Ausdehnung mehr Rechenleistung benötigt wird.¹⁰⁵ Abbildung 8 zeigt den Aufbau eines vierschichtigen tiefen neuronalen Netzwerkes mit zwei versteckten Schichten. Die Variable k steht dabei für die Knoten der Zwischenschichten.

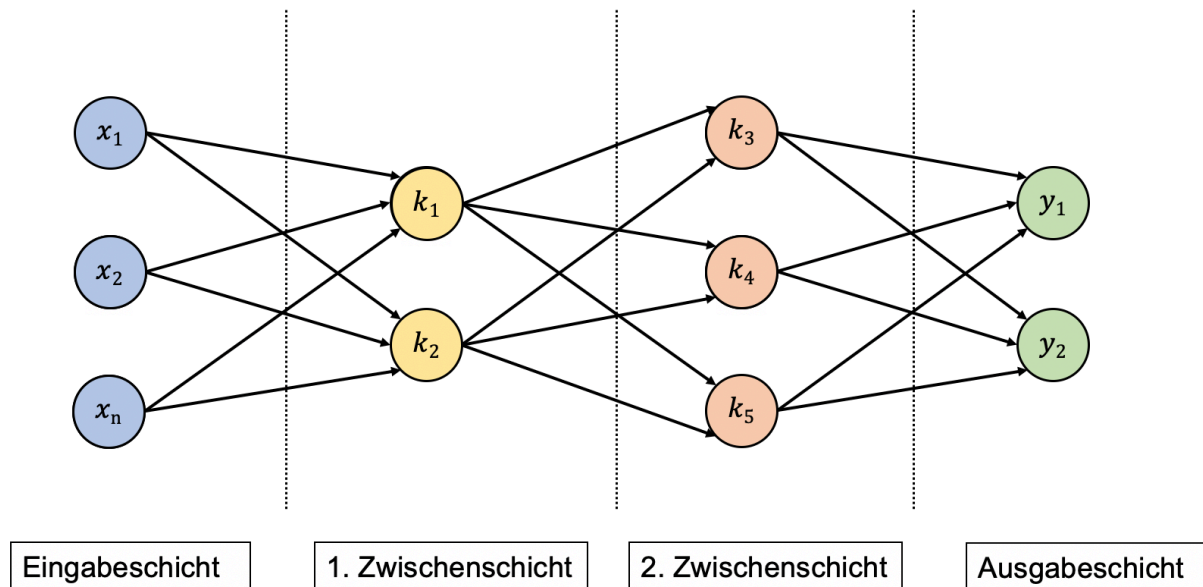


Abbildung 8: Aufbau eines tiefen neuronalen Netzwerkes

2.4.5. Netztopologien

Die Netztopologie beschreibt, auf welche Art und Weise die Knoten miteinander verknüpft sind. Dabei sind vorrangig drei Typen zu unterscheiden: Vorwärts gerichtete, rekurrente und faltende künstliche neuronale Netzwerke. Bei einem vorwärts gerichteten Netz, auch als Feed-Forward-Netz bezeichnet, geben die künstlichen Neuronen ihre Ausgabedaten immer an die Knoten der darauffolgenden Schicht weiter. Eine rückwärts gerichtete Informationsleitung im Sinne einer Übergabe der Daten an vorherige Knoten ist nicht möglich.¹⁰⁶

Bei rekurrenten neuronalen Netzen (RNN) ist die Vermittlung von Ausgabedaten sowohl an Knoten von vorherigen wie auch an jene von nachfolgenden Schichten durchführbar. Außerdem können Informationen innerhalb eines Knotens rückgemeldet werden. Dies ermöglicht schleifenartige Prozesse, die dem Netzwerk den Umgang mit komplexen Fragestellungen und Rückkoppelungsmechanismen erlauben. RNN verfügen zusätzlich über die Funktion eines „Forget-Gates“ [englisch], das das gezielte Merken oder Vergessen von Daten aus vorangegangenen Trainingsprozessen ermöglicht und damit eine Art selektives Gedächtnis darstellt. Rekurrente Netze haben einen breiten Anwendungsbereich, lassen sich aber besonders für Fragestellungen der Klassifikation oder Regression einsetzen.⁹⁶

Der dritte Netztyp umfasst die faltenden neuronalen Netzwerke („Convolutional Neural Networks“, kurz „convnets“ oder „CNN“, [englisch]). Sie erhalten ihre Eingangsdaten meist in Form einer mathematischen Matrix und berechnen ihre Ausgangsdaten durch Anwendung eines diskreten Faltungsprozesses. Die mathematische Faltung hat das Ziel, die Eingangsdaten auf ihre wesentlichen Merkmale zu reduzieren. Dadurch werden vor allem große Bilddateien kompakter und folglich Verarbeitungsaufwand und -dauer verringert. Faltende Netzwerke eignen sich hauptsächlich zur Bildklassifikation, da sich Bilddateien gut als Matrix darstellen lassen. Beim Lernprozess werden die Bilder in einzelne Abschnitte zerlegt. Anschließend extrahiert das Netzwerk Muster innerhalb dieser Segmente. Der Vorteil besteht in der Effizienz dieses Verfahrens: Hat ein CNN z.B. ein Muster in der linken unteren Ecke eines Bildes erkannt, kann es das gleiche Muster auch in einem anderen Bild erkennen, wenn es dort im rechten oberen Bereich liegt. Ein nicht-faltendes Netzwerk müsste dagegen in einem neuen Lernprozess das ganze Bild analysieren, um das Muster an der anderen Stelle zu erkennen. CNNs lernen also eher durch lokale als durch globale Mustererkennung. Aus diesem Grund können sie auch aus kleinen Datensätzen genug Muster extrahieren, um eine Generalisierung des Modells zu erreichen.⁹⁶

2.4.6. Transfer-Lernen

Das eigenhändige Erstellen der komplexen Architektur eines tiefen künstlichen neuronalen Netzwerkes sowie dessen Training bedeutet selbst bei großer Programmierexpertise einen hohen zeitlichen Aufwand bei unklaren Erfolgsaussichten. Daher ist es für eigene Deep Learning-Fragestellungen sinnvoller und vielversprechender, ein bereits vortrainiertes Netzwerk zu benutzen, das bestenfalls für ein ähnliches Problem aufgesetzt wurde, dieses anzupassen und es dann mit eigenen Daten zu trainieren. Dieses Vorgehen wird auch als Transfer Learning („Transfer-Lernen“, [englisch]) bezeichnet, da der Lernfortschritt eines bereits existierenden Modells auf eine neue Problemstellung transferiert wird. Als Vorbild dient auch hier das menschliche Lernen: So fällt es häufig leicht, eine neue Fähigkeit zu erlernen, wenn eine ähnliche bereits beherrscht wird. Jemandem, der z.B. schon ein Instrument spielt oder eine Fremdsprache beherrscht, fällt es häufig leichter, noch ein anderes Instrument bzw. eine weitere Fremdsprache zu erlernen.^{107,108}

Transfer-Lernen eignet sich besonders für die Analyse von unstrukturierten Daten wie Bild- oder Audiodateien. Unstrukturierte Daten verfügen im Gegensatz zu strukturierten Daten über keine standardisierte Formatierung oder anderen speziellen Bezeichnungen über den Inhalt der Daten. Strukturierte Daten besitzen dagegen eine oder mehrere Markierungen. Ein häufiges Beispiel für strukturierte Daten sind die Hintergrundinformationen einer Webseite, so kann z.B. eine Seite mit einem Rezept als „Rezept“ markiert sein und damit für Such- oder

Lernalgorithmen leichter zu identifizieren sein.¹⁰⁹ Beim Transfer-Lernen können sowohl eine wie auch mehrere Schichten des Modells neu trainiert werden. Sehr weit verbreitet ist das Re-Trainieren einzelner Schichten oder auch nur der letzten Schicht, also der klassifikationsbestimmenden Schicht. Dabei wird ein Modell verwendet, das bereits eine bestimmte Anzahl an Objekten erkennen kann. Ein Modell kann aber immer nur die Strukturen identifizieren, die es kennt. Außerdem kann es die neuen Daten auch nur in die Klassen einsortieren, die in der Klassifikationsschicht festgelegt sind. Konkret kann man z.B. einen Algorithmus mit Fotos von Hunden fünf verschiedener Rassen trainieren, sodass er anschließend ein neues Foto von einem Hund einer dieser fünf Rassen zuordnen könnte. Lässt man diesen Algorithmus aber ein Bild von einem Hund einer anderen, unbekanntem Rasse oder sogar einer Katze analysieren, wird er auch diese Bilddateien in eine der fünf erlernten Kategorien einteilen, auch wenn er das Prinzip der Mustererkennung in Bilddateien beherrscht. Fügt man dem Modell dagegen jeweils eine neue Lern- und Kategorisierungsschicht hinzu, die die gewünschten Klassen enthält, kann das Deep-Learning-Modell mit Hilfe eines neuen Datensatzes neue Merkmale erkennen und eine neue Klassifizierung vornehmen.¹¹⁰

Ein anderes Problem im Umgang mit Deep Learning-Programmen ist das sogenannte Overfitting („Überanpassung“, [englisch]). Dieses Phänomen tritt auf, wenn der Algorithmus seine Lernerkenntnisse nicht verallgemeinern und übertragen kann. Er erkennt in diesem Falle nur Dateien, die dem Lerndatensatz entstammen oder ihm in zahlreichen Merkmalen ähneln. Dabei misst er irrelevanten Merkmalen, die nicht zwischen den Klassen diskriminieren, eine zu hohe Bedeutung zu, anstatt Merkmalsmuster zu erkennen. Das Netzwerk enthält also sehr viele Parameter, von denen nicht alle für die Analyse relevant sind. Lässt man aber den Lerndatensatz durch alle Schichten des Modells passieren, werden alle diese Parameter an den Datensatz angepasst. Das Ergebnis ist zwar eine sehr gute Klassifikationsgenauigkeit im Trainingsprozess, jedoch auch eine geringe Generalisierbarkeit. Gerade bei eher kleinen Datensätzen mit einer begrenzten Zahl an unterscheidenden Parametern wird deswegen die Belassung der Zwischenschichten und das ausschließliche Re-Training der Klassifikationsschicht(en) bevorzugt. Außerdem führen zu monomorphe Dateien und zu kleine Trainingsdatensätze eher zu einer „Überanpassung“ des Modells. Ein ausreichend großer Trainingsdatensatz gewährleistet, dass das Programm lernt, (pathologische) Muster von interindividuellen Unterschieden zu unterscheiden. Andererseits muss der Testdatensatz ebenfalls groß genug sein, um über eine ausreichende Aussagekraft zu verfügen. Test- und Trainingsdatensätze müssen außerdem unbedingt voneinander getrennt gespeichert werden, um Überschneidungen auszuschließen. Bei Überschneidungen bzw. Dopplungen von Bildern zwischen Trainings- und Testdaten müsste das Programm ähnlich wie bei einem Memory-

Spiel nur bereits bekannte Bilder wiedererkennen, anstatt das Erlernete auf unbekannte Dateien anzuwenden. Des Weiteren müssen die verschiedenen möglichen Merkmalsvarianten in ähnlicher Häufigkeit vorliegen, um einen ausgeglichenen Lernprozess zu gewährleisten. Ansonsten würde das Programm die häufiger vertretene Kategorie besser kennen lernen und es zu einer möglichen Verzerrung der Klassifikation kommen.⁹⁶

2.4.7. Verfügbare Deep Learning-Instrumente

Wie aus den vorherigen Kapiteln ersichtlich wurde, handelt es sich bei künstlichen neuronalen Netzwerken um komplexe Gebilde, deren Erstellung erhebliche Kenntnisse verlangt. Das tiefe Lernen bedient allerdings ein breites Spektrum an Fragestellungen, die thematisch nicht zwangsläufig direkten Bezug zur Computerwissenschaft haben, wie z.B. Marketing-Projekte. Es ist also ein Feld, das auch für viele Menschen von Interesse ist, die hauptsächlich in einem fachfremden Gebiet arbeiten und daher über einen begrenzten informatischen Hintergrund verfügen. Aus diesem Grunde gibt es verschiedene Instrumente, die das Aufsetzen von Machine bzw. Deep Learning-Algorithmen nach dem Baukasten-Prinzip ermöglichen. Dabei sind die Rahmenbedingungen zur Etablierung eines eigenen künstlichen neuronalen Netzwerkes gegeben, sodass die Nutzer*innen diese nicht mehr selber erschaffen müssen. Stattdessen können sie bereits vorhandene Strukturen individuell an die Bedürfnisse ihres geplanten Vorhabens anpassen. Die verschiedenen Instrumente unterscheiden sich unter anderem in ihrer Komplexität, den anzuwendenden Programmiersprachen und ihrer Verfügbarkeit. Im Folgenden wird ein Überblick über ausgewählte Systeme zur Anwendung des tiefen Lernens gegeben.

2.4.8. Microsoft® Cognitive Toolkit

Das Microsoft® Cognitive Toolkit ist ein sogenanntes Toolkit, das durch das Softwareunternehmen Microsoft® entwickelt wurde. Ein Toolkit („Werkzeugkasten“, [englisch]) bezeichnet eine Programmbibliothek, die die Basiskomponenten zur Software-Entwicklung enthält. Toolkits ermöglichen die Verwendung des gleichen Programmcodes zwischen verschiedenen Applikationen, sind eher allgemein gehalten und damit nicht sehr anwendungsspezifisch. Es werden sowohl Algorithmen zum überwachten wie auch zum nicht überwachten und verstärkenden Lernen unterstützt. Das Toolkit ermöglicht außerdem die Anwendung von vorwärts und rückwärts gerichteten sowie von faltenden künstlichen neuronalen Netzwerken. Als Programmiersprache wird hauptsächlich C++ verwendet, es existiert aber auch eine Schnittstelle zur Anwendung von Python.¹¹¹ Python orientiert sich an der regulären englischen Sprache und ist dadurch recht intuitiv in der Anwendung. Es daher etwas leichter zu erlernen. C++ zeigt eine kompliziertere Syntax auf und ist eher eine

maschinennahe Programmiersprache und dadurch für Anfänger schwieriger nachzuvollziehen. Das Microsoft® Cognitive Toolkit ist ein sogenanntes Open-Source-Toolkit [englisch]. Open Source („offene Quelle“, [englisch]) bedeutet, dass der in Form von Programmiersprache verfasste Quelltext eines Programmes öffentlich zugänglich ist. Es kann von daher durch jede Person weltweit kostenlos verwendet werden. Microsoft® selbst setzt es bereits umfangreich bei seinen eigenen Anwendungen ein, so z.B. bei seinem sprachgesteuerten Assistenten Cortana®, der auf zahlreichen Computern mit einem Windows® -Betriebssystem installiert ist.^{111,112}

2.4.9. TensorFlow™

Bei TensorFlow™ handelt es sich um ein Produkt des Unternehmens Google®. Es wurde im Rahmen des Google Brain™ Projects entwickelt, das Teil der Forschungsabteilung des Konzerns ist und sich die Erstellung hochflexibler Deep Learning-Modelle zum Ziel gesetzt hat. Die Abteilung fing 2011 unter der Leitung von Jeffrey Dean, Greg Corrado und Andrew Ng an, sich mit der Anwendung umfangreicher tiefer neuronaler Netzwerke sowohl für den forschungsbezogenen wie auch für den firmeninternen Gebrauch zu befassen. Daraus folgte die Entwicklung des Software-Frameworks DistBelief. Ein Framework („Gerüst, Rahmen“, [englisch]) ist kein eigenständiges Programm, sondern eine vorprogrammierte Infrastruktur, in der bestimmte, häufig wiederkehrende Strukturen und Elemente schon enthalten sind. Die Verwendung eines Frameworks bringt eine Verringerung des Programmieraufwandes und damit einen erheblichen Gewinn an Zeit mit sich. DistBelief ist eine proprietäre Software, der Quellcode ist dementsprechend nicht öffentlich einsehbar. Im Gegensatz zu Open Source-Software steht DistBelief damit nur einem begrenzten Nutzerkreis zur Verfügung¹¹³.

Als Weiterentwicklung von DistBelief entstand schließlich das Framework TensorFlow™, das 2015 erstmals vorgestellt wurde. TensorFlow™ ist ebenfalls auf Anwendungen des maschinellen und tiefen Lernens spezialisiert. Es umfasst mehrere Softwarepakete und eröffnet unter anderem die Möglichkeit, künstliche neuronale Netzwerke zu trainieren. Es bietet mehr Leistung und eine höhere Spannweite an kompatiblen Hardware-Plattformen als DistBelief, z.B. ist auch eine Verwendung auf mobilen Endgeräten wie Smartphones möglich. Durch die Verwendung eines einzelnen Systems für verschiedene Plattformen wird die Anwendung vereinfacht und übersichtlicher, da nicht mehr zwischen verschiedenen Systemen gewechselt werden muss. Außerdem ist TensorFlow™ als Open-Source-Produkt im Gegensatz zu seinem Vorgängersystem einem breiten Anwenderkreis zugänglich. Die vorrangige Programmiersprache ist Python. Ein besonderes Merkmal ist die umfangreiche Bibliothek an bereits vortrainierten Modellen. Der Anwendungsfokus liegt auf der Text- und Sprachanalyse sowie Bilderkennung, durch die hohe Variabilität der anzuwendenden

Algorithmen wird TensorFlow™ aber auch in zahlreichen anderen Feldern wie z.B. der Robotik oder der Geographie verwendet.¹¹⁴ Google® selbst verwendet das System in zahlreichen seiner eigenen Produkte, unter anderem in dem Übersetzungsprogramm Google Translate™ oder beim Kartendienst Maps™.¹¹⁵ Außerdem stellt das Unternehmen den Nutzer*innen zahlreiche „Tutorials“ zur Verwendung von TensorFlow™ zur Verfügung. In diesen wird ähnlich einer Betriebsanleitung exemplarisch die Anwendung für verschiedene Fragestellungen erläutert.

2.4.10. Keras

Keras stellt ähnlich wie TensorFlow™ ein Framework dar und hat seinen Schwerpunkt im Bereich des maschinellen und tiefen Lernens. Es wurde ursprünglich für forschungsbezogene Anwendungen entwickelt und kann zusätzlich in Kombination mit dem Microsoft® Cognitive Toolkit oder auch mit TensorFlow™ und anderen Machine Learning-Anwendungen genutzt werden. Die Flexibilität zwischen Keras und den genannten Plattformen ist insofern vorteilhaft, dass in Keras geschriebener Code mit dem verwendeten Framework bzw. Toolkit kompatibel ist. Dies ermöglicht einen fließenden Übergang, wenn eine der Plattformen für eine spezifische Anwendung besonders geeignet oder schneller in der Verarbeitung ist als Keras. Die zu verwendende Programmiersprache ist ebenfalls Python. Keras ermöglicht die Verwendung von faltenden und rekurrenten Netzwerken und auch die Kombination dieser Netzwerktopologien. Ein Vorteil liegt in der Bedienerfreundlichkeit, die sowohl die Programmierschnittstelle wie auch die Baukasten-ähnliche Bereitstellung von Elementen zur Entwicklung von Modellen betrifft. Besonders erwähnenswert ist die NVIDIA CUDA Deep Neural Network Library, bei der es sich um eine auf tiefes Lernen spezialisierte Programmbibliothek handelt.⁹⁶

2.5 Studienlage

Betrachtet man die bisherige Verwendung von maschinellem bzw. tiefem Lernen in der Medizin, so fällt ein starker Fokus auf diagnostischen Methoden auf. Insbesondere bildgebende Verfahren eignen sich, da sie große Datenvolumina liefern, anhand derer die Maschine lernen kann. Eine explizite oder sogar zusätzliche Datenakquisition für den Algorithmus entfällt, da die Aufnahmen bei entsprechender Indikation ohnehin angefertigt werden und anschließend von Ärzt*innen und zusätzlich von der Maschine analysiert werden können. Besonders relevant ist dieser Aspekt bei Verfahren, die mit einer Strahlenbelastung einhergehen. Eine erhöhte Strahlenexposition zum Zweck der KI-Anwendung würde ein erhebliches ethisches Dilemma mit sich bringen und entsprechende Studien kaum umsetzbar machen. KI-Anwendungen werden bei medizinischen Bilddaten im Wesentlichen in zwei

Bereichen eingesetzt: In der Vorverarbeitung der Daten (z.B. für die Bildsegmentierung oder die Bildregistrierung) sowie in der diagnostischen Beurteilung. Choi et al. stellen einen Algorithmus basierend auf faltenden neuronalen Netzwerken vor, der die striatalen Strukturen im MRT nicht nur genauer, sondern auch schneller segmentiert als ein übliches Prozessierungsprogramm.¹¹⁶ Shao et al. bewerkstelligten die direkte Rekonstruktion cerebraler SPECT-Dateien von Rohdaten mittels eines selbst erstellten tiefen neuronalen Netzwerkes, sodass keine iterative Rekonstruktion notwendig war.¹¹⁷

In der diagnostischen Leistung übertreffen die Programme teilweise langjährige klinische Expert*innen. Besonders erfolgreich wurden solche Methoden bereits bei der Bewertung von Röntgen- und CT-Aufnahmen des Thorax in Hinblick auf Pathologien der Lunge eingesetzt. Es ergibt sich ein Kontrast zwischen gesundem Lungengewebe, das sich hypodens zeigt, und zahlreichen eher hyperdens erscheinenden Krankheitsprozessen. Epidemiologisch besonders relevant sind Bronchialkarzinome, sodass zahlreiche Untersuchungen zur Suche und Klassifizierung von Lungenrundherden existieren.¹¹⁸⁻¹²² Die hohe Sensitivität, Spezifität und Genauigkeit verschiedener Systeme hat dazu geführt, dass in manchen Praxen und Kliniken die Bewertung von CT-Thorax-Aufnahmen durch ein KI-gestütztes Programm zusätzlich zur Befundung durch radiologische Ärzt*innen zur Routine gehört. Ein anderes Feld sind die Methoden, bei denen im klinischen Alltag primär keine Bilddaten erhoben werden, diese aber z.B. mittels Fotografie ethisch vertretbar und mit geringem Aufwand erstellt werden können. Als Beispiel sind die Analyse von pigmentierten Hautveränderungen¹²³ oder die Bewertung endoskopischer Aufnahmen des Darmes, z.B. zur Detektion von Polypen, zu nennen.¹²⁴⁻¹²⁶ Die in der vorliegenden Arbeit verwendete Deep Learning-Programmibibliothek TensorFlow™ hat sich zur Beurteilung der medizinischen Bildgebung bereits in mehreren Arbeiten als erfolgreich einsetzbar gezeigt. Beispiele sind die Entwicklung eines Modells für die Einschätzung des Brustkrebs-Risikos anhand von Mammogrammen¹²⁷ und die Klassifikation von Thorax-Röntgen-Aufnahmen in Hinblick auf das Vorliegen mindestens einer von 14 möglichen Pathologien¹²⁸. In Hinblick auf die Analyse dreidimensionaler Bildgebung ist die Studienlage noch eher übersichtlich. Coy et al. verwendeten ein vortrainiertes faltendes Netzwerk, um in Multislice-CT-Aufnahmen des Abdomens renale Läsionen binär nach Onkozytomen oder klarzelligem Nierenzellkarzinomen zu klassifizieren. Die höchste erreichte durchschnittliche Genauigkeit betrug 76.0 % (95 % CI 69.0-82.0 %) und erscheint daher optimierungsbedürftig.¹²⁹

In der Diagnostik neurodegenerativer Erkrankungen wurden automatisierte Klassifizierungssysteme ebenfalls bislang vielfältig eingesetzt. Choi et al. präsentieren ein vierschichtiges faltendes Modell, welches eine Differenzierung von an Morbus Alzheimer

Erkrankten und nicht Betroffenen anhand der Nuklidbelegung im FDG-PET vornehmen kann (AUC 0.94, 95% Konfidenzintervall (CI) 0.89-0.98). Außerdem ist es in der Lage, bei Patient*innen mit einer leichten kognitiven Einschränkung (MCI, mild cognitive impairment, [englisch]) eine prognostische Aussage über einen Progress im Sinne eines Morbus Alzheimer zu treffen. Zusätzlich konnte es anhand der FDG-Daten Betroffene eines Morbus Parkinson mit Demenz von jenen ohne Demenz differenzieren.¹³⁰ Qiu et al. haben ein faltendes künstliches neuronales Netzwerk zur Kategorisierung kranialer MRT-Daten verwendet. Dabei wurde zwischen Personen ohne Einschränkung, leichter kognitiver Einschränkung, Morbus Alzheimer sowie nicht-Alzheimer-Demenzformen unterschieden. Die Trennschärfe lag im Bereich von ca. 65-85%. Erst der Einsatz eines hybriden Modells sowie die Kombination bildgebender und klinischer Daten brachten Ergebnisse von über 90%.¹³¹

Auch die Beurteilung von DAT-SPECT-Aufnahmen mittels KI-Methoden wurde bereits in mehreren Publikationen untersucht.¹³²⁻¹³⁸ Zhang und Kagen entwickelten ein einschichtiges künstliches neuronales Netzwerk in TensorFlow™, das solche Bilddateien im Hinblick auf eine dopaminerge Minderbelegung analysieren kann. Dabei erreichte es eine durchschnittliche Genauigkeit von $93.8 \pm 4.7\%$, eine durchschnittliche Sensitivität von $97.4 \pm 4.3\%$ und eine durchschnittliche Spezifität von $82.2 \pm 2.1\%$. Allerdings kann der Algorithmus nur jeweils eine einzelne axiale Aufnahme aus dem dreidimensionalen Datensatz verarbeiten, die vorher in einem zusätzlichen Arbeitsschritt extrahiert werden muss.¹³⁴ Kim et al. stellten in ihrer Arbeit von 2018 einen Algorithmus vor, der ein 48 Schichten tiefes faltendes künstliches neuronales Netzwerk erprobt. Es gelang, durch Transfer-Lernen auffällige und unauffällige DAT-SPECTs mit einer Sensitivität von 96.3% und einer Spezifität von 66.7% zu differenzieren. Die AUC lag bei 0.87. Diese Veröffentlichung lieferte den Beweis, dass auch bei einer kleinen Kohorte (n=108) beeindruckende Ergebnisse möglich sind.¹³² Magesh et al. trainierten ein 16-schichtiges faltendes Netzwerk mittels Transfer-Lernen mit DAT-SPECTs und erreichten dabei eine Genauigkeit von 95.2%, eine Sensitivität von 97.5% und eine Spezifität von 90.9%. Dabei wurden die Dateien wie auch in der vorliegenden Arbeit binär klassifiziert und ebenfalls ausschließlich Bilddaten berücksichtigt.¹³⁶ Außerdem werden in den meisten vergleichbaren Veröffentlichungen ebenfalls rekonstruierte und damit zweidimensionale Daten verwendet, vor allem in jenen vor 2020.^{133,135,136,139} Dies ist auf die zu jenem Zeitpunkt eingeschränkte Verfügbarkeit von 3D-kompatiblen Netzwerken zurückzuführen. Leung et al. benutzten dagegen nicht rekonstruierte, das heißt Projektionsdaten, von 659 Personen und trainierten damit ein dreidimensionales faltendes Netzwerk, eine noch sehr neue Form solcher Netzwerke. Dabei wurde eine Richtigklassifikationsrate von 97% erreicht und der Prozess der Rekonstruktion entfiel.¹³⁷ Es fällt zudem auf, dass zunehmend klinische Aspekte zusätzlich zum bildmorphologischen Befund berücksichtigt werden. So verwendeten Zhao et al. in ihrer

Arbeit von 2022 ebenfalls ein dreidimensionales tiefes faltendes Netzwerk und ließen neben demographischen Daten den Wert auf der UPDRS-Skala sowie das Stadium nach Hoehn & Yahr einfließen. Dies kommt zum einen dem menschlichen Vorgehen näher als die alleinige Bildanalyse, zum anderen erlaubt es zusätzliche differentialdiagnostische Aussagen, z.B. bei atypischen Parkinson-Syndromen. Zhao et al. erreichten auf diese Weise Trennschärfen von 90.0-95.3%, Sensitivitätswerte von 78.6-93.7% und Spezifitätswerte von 88.4-97.5%, um das idiopathische Parkinson-Syndrom, die MSA und die PSP voneinander zu unterscheiden.¹³⁸

2.6 Fragestellungen und Ziel der Arbeit

Die Beweggründe für eine Analyse von FP-CIT-SPECT-Aufnahmen mittels tiefen Lernens ergeben sich zum einen aus den Eigenschaften dieser Daten: Es sind große Datenmengen vor Ort verfügbar, die nach einem übersichtlichen System klassifiziert sind. Die Klassifikation ist binär, der Algorithmus muss also lediglich zwei verschiedene Muster erkennen können. Zum anderen ließ sich der limitierten Kompatibilität zahlreicher Deep Learning-Algorithmen auf zweidimensionale Daten dank der klinisch etablierten Erstellung von MIPs begegnen. Im Hinblick auf die Auswahl und Anwendung eines Modells kam dem Projekt zugute, dass im Internet viele vortrainierte Netzwerke zur Bildklassifizierung frei verfügbar und in ihrer Handhabung erläutert sind.

Gemeinsam ist den meisten Arbeiten, dass die Vorbereitung der Bilddaten eine erhebliche Herausforderung darstellt, damit ein Netzwerk überhaupt in die Lage gelangen kann, sie zu verarbeiten. Die Programmbibliothek TensorFlow™ erscheint in der Hinsicht als geeignete Anwendung. Die Anzahl möglicher Klassifikationsklassen liegt meist im einstelligen Bereich und ist damit eher klein. Außerdem wird meistens nur ein Netzwerk bzw. eine Ausführung eines Netzwerkes verwendet, sodass der Vergleich von verschiedenen Netzwerken oder Netzwerkkonfigurationen zu einer identischen Fragestellung oft nicht möglich ist. Ziel der Arbeit war daher neben der Erprobung eines Deep Learning-Ansatzes zur automatisierten Klassifizierung von DAT-SPECT der Vergleich von verschiedenen Kombinationen aus Netzwerkversionen, Trainingsparametern und Datensatzkonfigurationen, als es in bisher publizierten Studien der Fall ist.

3. Material und Methoden

3.1 Material

3.1.1. Herkunft der Daten

Es wurden die im Zeitraum vom 20.01.2012 bis zum 11.04.2018 in der Klinik und Poliklinik für Nuklearmedizin der Uniklinik Köln durchgeführten kraniellen Dopamin-Transporter-SPECTs erfasst. Bei dem verwendeten Tracer handelte es sich um DaTScan® des Herstellers GE (General Electric) und zur anschließenden Detektion wurde als Gamma-Kamera das Modell Picker Prism 3000 verwendet. Das Bildmaterial ist durch Assistenzärzt*innen der Nuklearmedizin diagnostisch beurteilt und anschließend ober- bzw. chefärztlich vidiert worden. Die Befunde wurden im Patientendatenverwaltungssystem ris als Dokument gespeichert. Es wurden für den genannten Zeitraum insgesamt 1679 Bilddatensätze gefunden.

3.1.2. Verwendete Systeme

In Tabelle 1 sind die verwendeten Hardwarebestandteile, in Tabelle 2 ist die Software aufgeführt.

Komponente	Bezeichnung
Prozessor	2,3 GHz Dual-Core Intel Core i5
Betriebssystem	macOS Catalina Version 10.15.4
Arbeitsspeicher	8 GB 2133 MHz LPDDR3
Festplatte	256 GB SSD

Tabelle 1: Verwendete Hardware

Software	Version	Verwendung
Microsoft® Excel®	16.40	Erstellen von Tabellen, Durchführung von Berechnungen
Microsoft® Remote Desktop	10.3.7	Etablierung einer Server-Verbindung
Cendio® ThinLinc-Client	4.9.0	Etablierung einer Server-Verbindung
Cisco® AnyConnect Secure Mobility Client	4.10.02086	Etablierung einer VPN-Verbindung
MathWorks® MATLAB	R2018b	Erstellung und Ausführung von Codes bzw. Skripten
Microsoft® Visual Studio Code	1.30.2	Erstellung und Ausführung von Codes bzw. Skripten
TensorFlow™	1.7.0	Programm-bibliothek für tiefes Lernen
IBM® SPSS®	28.0.1.1	Statistische Analyse

Tabelle 2: Verwendete Software

3.2 Methoden

Kurz zusammengefasst wurden zuerst Bilddateien heruntergeladen und ersten Verarbeitungsschritten unterzogen. Anschließend erfolgte eine Auswahl geeigneter Dateien nach morphologischen und klinischen Gesichtspunkten. Nach weiterer Verarbeitung der Bilder folgten schließlich das Trainieren und anschließende Testen der Netzwerke. Abbildung 9 gibt einen allgemeinen Überblick über das methodische Vorgehen. Anschließend werden die einzelnen Abläufe detaillierter erläutert.

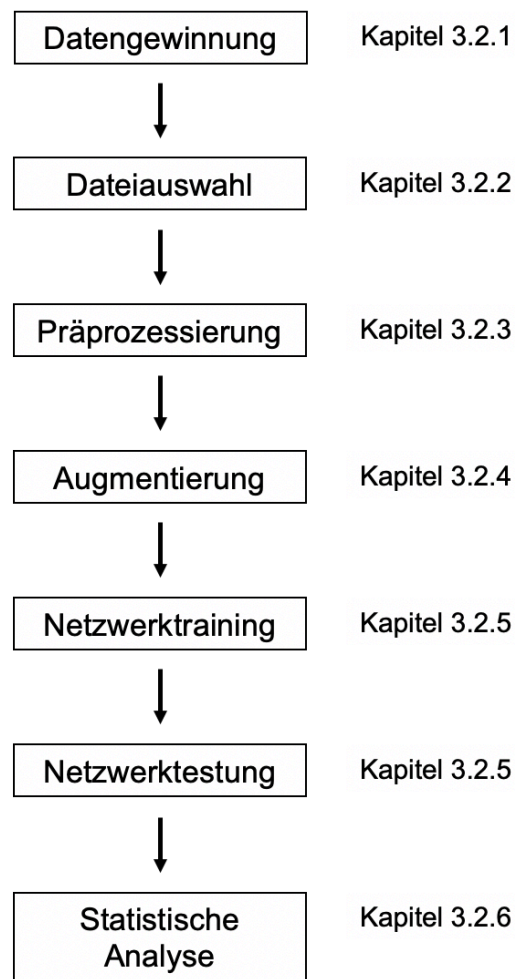


Abbildung 9: Methodisches Vorgehen

3.2.1. Herunterladen und erste Bearbeitung

Die Rohbilddateien wurden im DICOM-Format auf einen Universitäts-internen Rechner heruntergeladen. Der Zugriff auf den Server erfolgte über eine Passwort-geschützte VPN-Verbindung (virtuelles privates Netzwerk) mittels des Programms Cisco® AnyConnect über das Wissenschaftsnetz der uk-it. Für den Dateinamen wurde folgendes Format gewählt: ID_Nachname_Vorname_Geburtsdatum_Untersuchungsdatum. Die ID (Identifikationsnummer) entspricht einer Nummerierung bzw. Laufnummer, die jede einzelne

Bilddatei erhielt. Bei der genannten Gesamtanzahl an Dateien liegen die Identifikationsnummern dementsprechend als eine Zahl zwischen 1 und 1679 vor. Nach dem Herunterladen der Bilddateien mussten diese in eine Form geführt werden, die gut visuell beurteilbar und mit Machine Learning-Systemen technisch kompatibel ist.

Es wurden die dreidimensionalen Rohdaten heruntergeladen, um im Verlauf eine räumliche Normalisierung durchführen zu können (siehe unten). Von daher musste der Schritt der MIP-Erstellung noch durchgeführt werden. Dafür wurde das Skript CreatePlaneSums.m in MATLAB erstellt. Die DICOM-Dateien wurden durch das Programm gelesen und in Form von dreidimensionalen Matrizen verarbeitet. Als Farbschema wurde eine Grauwerte-Skala gewählt. Anschließend wurden Bildvariablen mit dem Wert 1 entfernt, sodass eine Kompression bzw. Stauchung der Daten durch Reduktion um die dritte Ebene resultierten. Anschließend wurde der Zusatz „planeMIP“ im Dateinamen ergänzt und die Daten in das GIF-Dateiformat (Graphics Interchange Format, [englisch]) konvertiert. Diese wurden separat in einem neuen Ordner abgespeichert, um eine adäquate Unterscheidbarkeit zu den Ausgangsdateien zu gewährleisten. Im Ergebnisteil wird ein Datensatz vor und nach Erstellung der MIP demonstriert (Kapitel 4.2, Abbildung 21). Der Bilddatensatz befand sich nun in einer auf einem herkömmlichen Rechner gut zu handhabenden Form, jedoch waren die Bilddateien zu diesem Zeitpunkt ausschließlich chronologisch nach dem Untersuchungsdatum sortiert. Folglich bestand der nächste Schritt in der Erfassung klinischer und diagnostischer Informationen der untersuchten Personen. Abbildung 10 fasst die bis hier genannten Schritte zusammen.

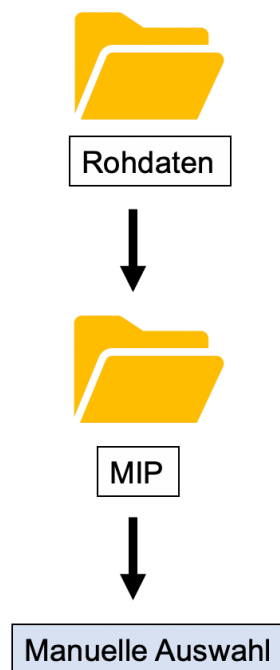


Abbildung 10: Verarbeitung der Rohdaten

3.2.2. Auswahl und Analyse geeigneter Dateien

Es wurden insgesamt 1679 Bilddateien erfasst, die von 529 Patient*innen der Klinik und Poliklinik für Nuklearmedizin der Uniklinik Köln aus einem Zeitraum über 6,23 Jahre stammen.

Folgende Kriterien wurden für die Auswahl geeigneter Dateien festgelegt:

- Es sollte nur das Gehirn abgebildet sein, d. h. Bilddateien, auf denen zusätzliche Bereiche wie z.B. Teile der Nase mit abgebildet waren, wurden nicht eingeschlossen.
- Um das Gehirn herum sollte der gesamte Bereich schwarz sein. Artefakte, meist in Form gräulicher Ringe um das Hirn herum, bedingten einen Ausschluss der entsprechenden Datei.
- Die gewünschte Bildebene war die axiale Lage. Dateien in koronarer oder sagittaler Ebene wurden ausgeschlossen.
- Das Gehirn sollte grob mittig im Bild liegen.
- Pro Patient*in sollte maximal ein Bilddatensatz pro Rekonstruktionsmethode gewählt werden, d. h. maximal ein Datensatz in iterativer Rekonstruktion und einer in gefilterter Rückprojektion. Anhand des Verhältnisses der Anzahl von untersuchten Patient*innen zur Zahl initial vorliegender Datensätze lässt sich bereits erahnen, dass häufig mehrere Datensätze pro untersuchter Person (teilweise > 10) vorlagen.
- Dopplungen von Patient*innen durch Mehrfachuntersuchungen, z.B. im Rahmen von Verlaufskontrollen, sollten vermieden werden. Es kamen in solchen Fällen nur die Bilddateien von einer Untersuchung für die Auswahl in Frage.

Die Bilddateien in Maximumintensitätsprojektion wurden einzeln nacheinander betrachtet und anschließend je nach Zutreffen der Selektionskriterien manuell ausgewählt. In Abbildung 11 werden Beispiele für Bilddateien gezeigt, die ausgeschlossen wurden, Abbildung 12 demonstriert exemplarisch geeignete Dateien. Die oberen beiden Dateien in Abbildung 12 zeigen unauffällige Bildbefunde, die unteren beiden sind dagegen als auffällig zu beurteilen.

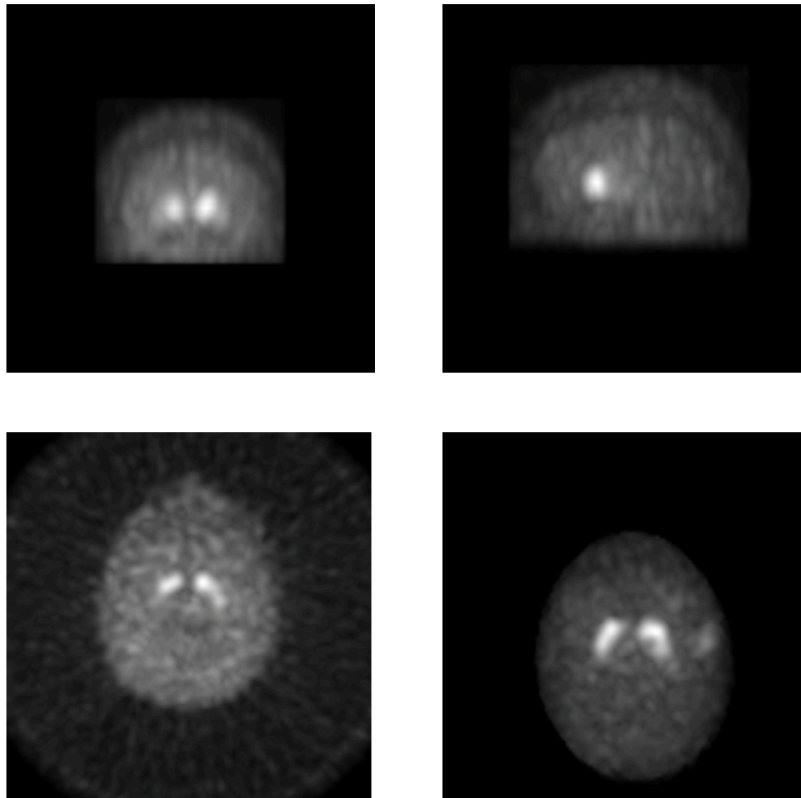


Abbildung 11: Exemplarische Darstellung ausgeschlossener Bilddateien

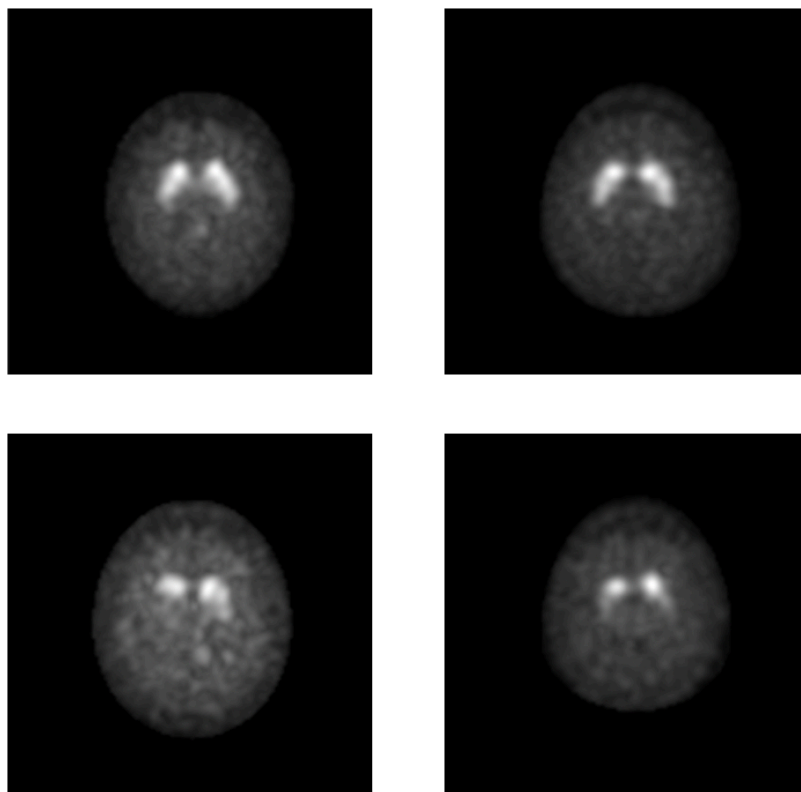


Abbildung 12: Exemplarische Darstellung eingeschlossener Bilddateien

Trafen die gewünschten Kriterien bei einer Bilddatei zu, wurden anschließend das Vorliegen eines dopaminergen Defizits (ja/nein/unklar) sowie die Rekonstruktionsmethode (iterativ/gefilderte Rückprojektion) erfasst. Dafür wurden nach der bildmorphologischen Beurteilung die jeweiligen schriftlichen ärztlichen Befunde im Patientendatenverwaltungssystem ris auf einem Computer im Kontrollbereich der Klinik für Nuklearmedizin gesichtet und die abschließende Beurteilung kopiert. Bei unklaren Befunden erfolgte eine Reevaluation mit PD Dr. Hammes und anschließender Zuordnung in eine diagnostische Kategorie. Schließlich wurden die erhobenen Charakteristika Bild für Bild in eine in Google® Drive erstellte Tabelle eingetragen. Diese weist eine Abmessung von 7x1680 Zellen auf und ist in ihrer genaueren Konfiguration ausschnittsweise in Abbildung 13 dargestellt. Die Kategorisierung der Dateien war elementare Voraussetzung für die Vergabe der Labels (siehe unten).

	A	B	C	D	E	F	G
1	ID	Auswertung	Rekonstruktion	Dopaminerges Defizit	Beurteilung	unklar	ID selber Patient iterativ
74	73	nein					
75	74	ja	iterativ	ja	Pathologische Dopamin-Transporter-Scintigraphie mit dopaminergem Defizit bds, rechts deutlicher ausgeprägt als links, passend zu der klinisch linksdominanten Symptomatik. Der Befund ist vereinbar mit einem IPS.		
76	75	ja	FBP	ja	Pathologische Dopamin-Transporter-Scintigraphie mit dopaminergem Defizit bds, rechts deutlicher ausgeprägt als links, passend zu der klinisch linksdominanten Symptomatik. Der Befund ist vereinbar mit einem IPS.		74

Abbildung 13: Exemplarische Darstellung der Erfassung geeigneter Dateien und klinischer Informationen

Es lagen letztlich 191 Bilder (101 mit positivem, 90 mit negativem Befund) in gefilterter Rückprojektion vor, sodass die Dateimenge tendenziell zu klein für eine Einbeziehung in die Folgeprozesse war. Insbesondere die geringe Anzahl der Bilddateien mit negativem Befund hätte nur einen sehr kleinen Testdatensatz erlaubt. In Anbetracht der teilweise schlecht zu identifizierenden Putamina im Vergleich zum umgebenden Hirngewebe in der gefilterten Rückprojektion bestand außerdem die Gefahr, dass diese eher einen verzerrenden Effekt auf den Algorithmus haben könnten. Aus diesen Gründen sollten schließlich nur Bilder in iterativer Rekonstruktion eingeschlossen werden. Schließlich wurden die gewählten Bilddateien (n=442) extrahiert und in einen Ordner übertragen, der entsprechend des Vorliegens eines dopaminergen Defizits zwei Unterordner enthielt. Mit Hilfe der Informationen aus der Tabelle konnten die Dateien dem jeweils korrekten Subordner zugeordnet werden. Dabei befanden sich die bildmorphologisch auffälligen Bilder im Unterordner mit dem Label „pos“ für „positiv“ (250 Dateien), die unauffälligen im Unterordner „neg“ für „negativ“ (192 Dateien). Diese Unterteilung wurde für die anschließenden Folgeprozesse beibehalten. Abbildung 14 visualisiert diese initiale Datenstruktur, die die Bildinformationen mit der entsprechenden diagnostischen Beurteilung verknüpft.

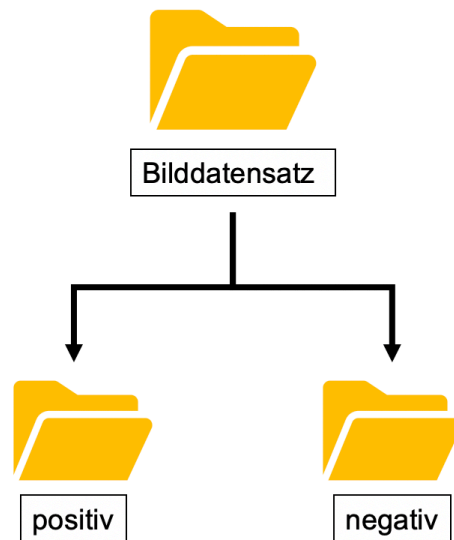


Abbildung 14: Initiale Datenstruktur

3.2.3. Präprozessierung

Nach der Auswahl geeigneter Dateien galt es, die Daten weiter für das Machine Learning-System vorzubereiten. Die folgenden Prozesse erfolgten wie zuvor auf einem arbeitsgruppeninternen Passwort-geschützten Server. Der Zugriff erfolgte auch hier innerhalb des Wissenschaftsnetzes der Uniklinik Köln mittels Cisco® AnyConnect, anschließend wurde eine VPN-Verbindung unter Verwendung von Cendio® ThinLinc hergestellt.

Die Rohdaten wurden zuerst für den Normalisierungsprozess mittels des Programms MATLAB (Skript CreateNiftiFromDicom.m) in das NIFTI-Format (Neuroimaging Informatics Technology Initiative, [englisch]) für die weitere Verarbeitung importiert. Dieses erwies sich als besonders geeignet, da das gewünschte Template zur Normalisierung (siehe unten) ebenfalls in diesem Dateiformat vorlag. Anschließend wurde eine Markierung der Dateien vorgenommen, um ihre räumliche Orientierung für die spätere Normalisierung anzuzeigen. Es wurde die vordere Kommissur markiert, da diese anatomische Struktur relativ klar abgrenzbar ist und sich somit gut als Referenzregion eignet. Außerdem unterliegt sie einer geringen interindividuellen Variabilität und ist auch im Hinblick auf Unterschiede zwischen Patient*innen mit bzw. ohne ein dopaminerges Defizit wenig anfällig. Nach Erfassung der Koordinaten dieser Struktur wurden diese den Bilddateien hinzugefügt. Die so markierten Bilder erhielten das Kürzel „OS“ für „origin set“ („Ursprung markiert“, [englisch]) im Dateinamen. Die Abläufe dieses Vorgangs wurden im Skript SetOriginInNiftis.m definiert.

Zusätzlich war eine Änderung der Dateinamen notwendig, da manche Patient*innen in ihrem Namen sprachliche Besonderheiten wie z.B. Umlaute, Akzente oder Tremata auf einzelnen

Vokalen enthielten. Die meisten der verwendeten Computerprogramme folgen jedoch dem angloamerikanischen Alphabet, in dem diese nicht regelhaft vorkommen. Um eine Kompatibilität mit dem geplanten weiteren Vorgehen zu gewährleisten, wurde das Skript RemoveAccents.m im Programm MATLAB erstellt. Es hat die entsprechenden Anpassungen jeweils für Klein- und Großbuchstaben vorgenommen wie z.B. ein Ersetzen des „ä“ durch „ae“. Hierfür wurden die Umlaute und Vokale mit Akzenten in einem Umlautvektor aufgeführt. In einem Ersatzvektor wurden die Buchstaben in der gewünschten Form definiert, d. h. als Vokalfolge bei Umlauten bzw. ohne Akzent. Anschließend wurden die betreffenden Buchstaben entsprechend ersetzt. Die derartig bearbeiteten Dateien erhielten in ihrem Namen das Attribut „NoAccent“ (ohne Akzent, [englisch]) und wurden in einem neuen Ordner gespeichert. Anschließend wurde in MATLAB das Skript BatchNormalizeDatScans.m zur Durchführung des Normalisierungsprozesses erstellt. Die Normalisierung erfolgte mittels eines FP-CIT-SPECT-Templates, welches im NIfTI-Format vorlag. Bei einem Template handelt es sich um einen standardisierten Raum, ähnlich einer Schablone. Abbildung 15 demonstriert das Erscheinungsbild des Templates in axialer Ebene.

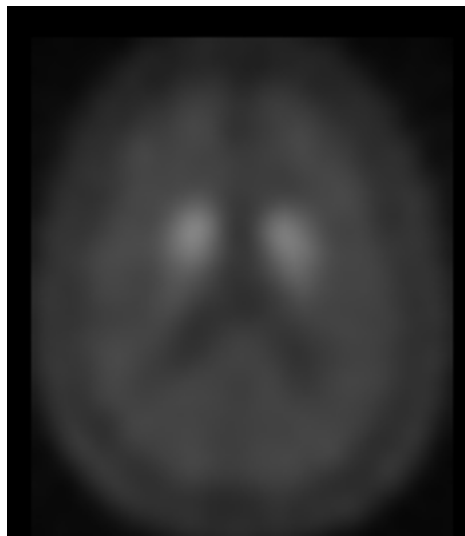


Abbildung 15: Verwendetes Template in axialer Ebene

Bei der Normalisierung wurden die Bilddateien in diesen standardisierten Bereich überführt, sodass die räumlichen Verhältnisse nun dem durch das Template vorgegebenen Standard entsprachen. Somit kam es zu einer Reduktion interindividueller Unterschiede wie z.B. unterschiedlich großer Hirnvolumina oder Proportionen der Cortizes, die für die diagnostische Beurteilung der dopaminergen Integrität kaum relevant sind. Zur Unterscheidung der normalisierten von den nicht normalisierten Bildern wurde für erstere das Kürzel „w“ den Dateinamen an erster Position hinzugefügt. Im Ergebnisteil sind normalisierte und nicht normalisierte Dateien im Vergleich dargestellt. Nach der Normalisierung wurden die Dateien

schließlich in das GIF-Format konvertiert und in einem separaten Ordner abgespeichert. Anschließend wurden aus den normalisierten Daten durch ein analoges Vorgehen wie bei den nicht normalisierten Bilddaten (siehe oben) Maximumintensitätsprojektionen erstellt.

Die nicht normalisierten wie auch die normalisierten Bilddateien lagen nun im GIF-Format vor. Der exemplarische Datensatz für das Trainieren der künstlichen neuronalen Netzwerke befand sich allerdings im JPEG-Format, sodass die Vermeidung einer späteren etwaigen Inkompatibilität während des Trainingsprozesses für einen Formatwechsel sprach. Die Bilddateien wurden deshalb in MATLAB (Skript MIPConvertGifToJpg.m) in das JPEG-Format konvertiert.

3.2.4. Datenaugmentierung

Aufgrund der eher seltenen Durchführung und des vergleichsweise kleineren Umfangs an verfügbarem Bildmaterial der DAT-SPECT kam die Überlegung auf, das vorhandene Bildmaterial mit bildbearbeitenden Methoden zu vermehren und dadurch mehr Lerndaten zu akquirieren. Es galt zu beachten, dass die dabei gewonnenen Dateien nicht verzerrt würden, sondern einen Bezug zu realistischen Bildaufnahmen gewährleisten. Es wurden verschiedene Methoden für diese sogenannte Datenaugmentierung angewandt: Zum einen wurden die Bilddateien dem bereits erläuterten Normalisierungsprozess zugeführt, zum anderen wurden sie gespiegelt und gedreht.

Der erste Gedanke bestand darin, die Bilddateien zu spiegeln. Für diesen Prozess wurde in MATLAB das Skript `dataAugmentation_flipImages.m` erstellt. Die ausgewählten Bilder wurden entlang der Sagittalachse gespiegelt und in einem neuen Bildordner unter Beibehaltung der Unterteilung in zwei Unterordner gespeichert. Zur Differenzierung der Dateien untereinander wurde bei den gespiegelten Bildern dem Dateinamen der Zusatz „flipped“ („gespiegelt“, [englisch]) hinzugefügt. Allein durch diesen Schritt konnte die Anzahl der verfügbaren Bilddateien verdoppelt werden. Außerdem wurden die Bilddateien um jeweils 10° einmal im und einmal gegen den Uhrzeigersinn rotiert. Da Patient*innen manchmal etwas schräg im Scanner liegen, ist eine leichte Drehung der Bilder ein realitätsnaher Einfluss, mit dem ein Programm ebenfalls umgehen können sollte. Für die Rotation der Dateien wurde in MATLAB das Skript `dataAugmentation_rotateImages.m` geschrieben. Die rotierten Dateien erhielten das Suffix CL für „clockwise“ („im Uhrzeigersinn“, [englisch]) bzw. CC für „counterclockwise“ („gegen den Uhrzeigersinn“, [englisch]) in der Bezeichnung. Anschließend wurden sie in einem separaten Ordner abgespeichert. Da die Rotation in zwei verschiedene Richtungen erfolgte, konnte durch diesen Schritt eine Verdreifachung des Datensatzes erzielt werden. Insgesamt wurde bei den nicht normalisierten wie auch mit den normalisierten Dateien je eine Spiegelung

sowie eine Rotation durchgeführt. Im Ergebnisteil werden sie exemplarisch demonstriert. Abbildung 16 fasst die bisher erläuterten methodischen Abläufe zusammen und gibt einen Ausblick auf die Folgeschritte.

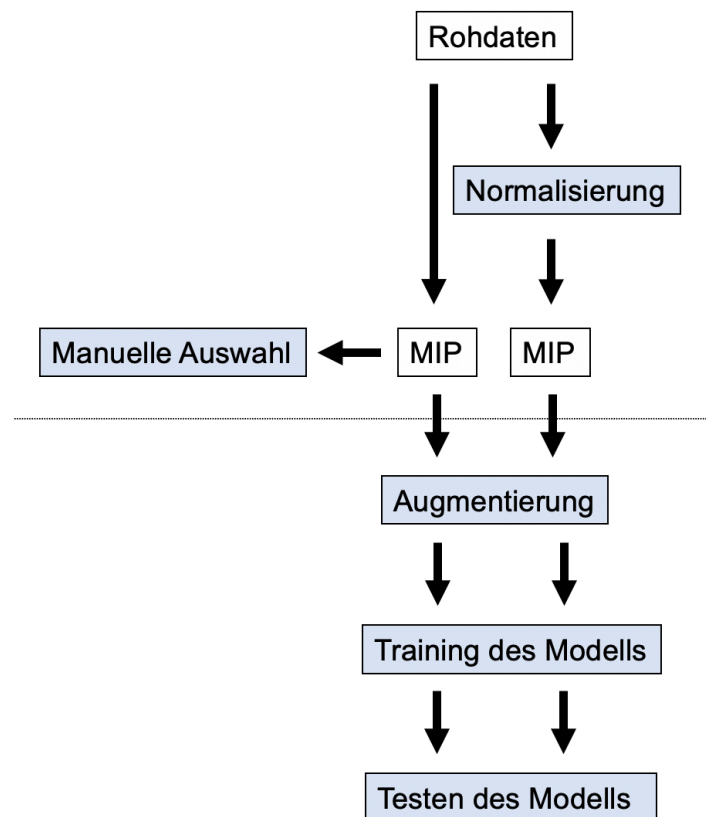


Abbildung 16: Vorbereitung der Bilddateien

3.2.5. Training und Testen der Netzwerke

Nach der Datenaugmentierung standen jeweils vier nicht normalisierte bzw. normalisierte, d. h. insgesamt acht verschiedene Datensätze zur Verfügung. Aus diesen wurden nun Trainings- und Testdatensätze erstellt. Dies sowie das anschließende Training und Testen der Netzwerke wurden mittels des Quelltexteditors Visual Studio Code von Microsoft® auf einem arbeitsgruppeninternen passwortgeschützten Server durchgeführt.

Für die Erstellung der Test- und Trainingsdatensätze wurden Extraktionsskripts in MATLAB geschrieben. Im Anhang ist exemplarisch das Skript für den normalisierten und rotierten Datensatz hinterlegt (Skript `Extract_40_random_Normalized_Rotated.m`). Als Dateiumfang wurde für die Testdatensätze eine Größe von 40 Bilddateien festgelegt. Das jeweilige Skript extrahiert zufällig je 20 Bilddateien aus den jeweils zwei Unterordnern, d. h. es werden 20 als auffällig und 20 als unauffällig beurteilte Bilder gewonnen. Die extrahierten Dateien wurden in neu erstellten Ordnern als Testdatensätze gespeichert. Die verbliebenen Bilder wurden

ebenfalls in separaten Ordnern als Trainingsdatensätze abgespeichert. Am Ende der Datenaugmentation lagen dementsprechend acht Trainings- und acht Testdatensätze vor, die Trainingsdatensätze jeweils unter Beibehaltung der strukturellen Unterteilung in zwei Unterordner. In den Testdatensätzen befanden sich die unauffälligen Bilddateien zusammen mit den auffälligen, da die Klassifikation in positiv und negativ im Rahmen des Testprozesses durch den Algorithmus erfolgen sollte. Abbildung 17 zeigt die unterschiedlichen Datenstrukturen im Vergleich.

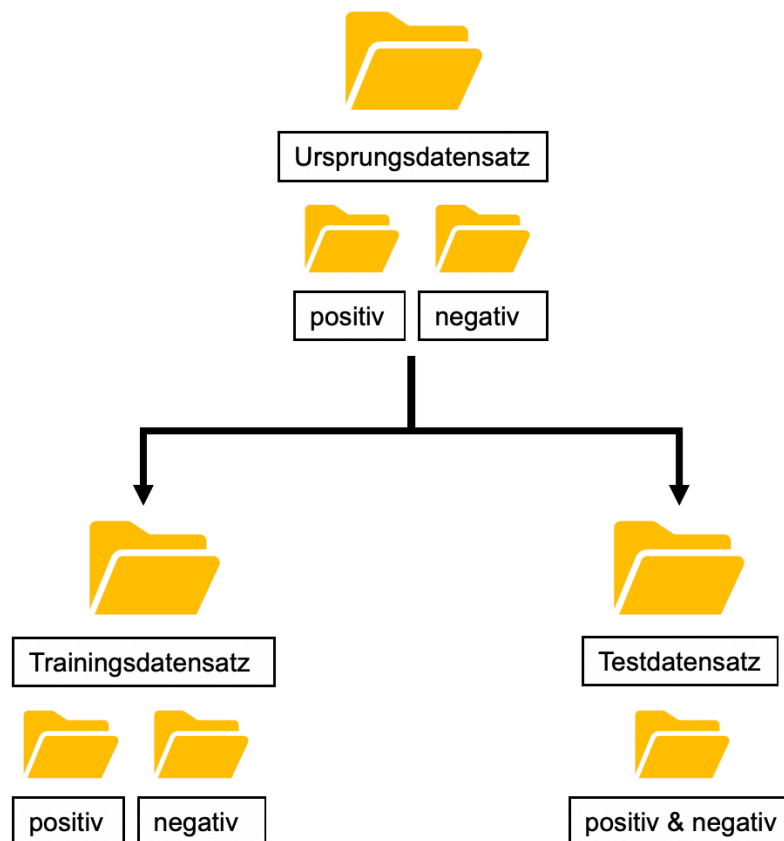


Abbildung 17: Datenstruktur der Trainings- und Testdatensätze

Für die nächsten Schritte wurde orientierend auf ein Tutorial des Unternehmens Google® zugegriffen. In diesem wurde schrittweise erläutert, wie man ein vortrainiertes künstliches neuronales Netzwerk trainiert, um Bilder von verschiedenen Blumenarten zu klassifizieren.¹⁰⁵ Nun galt es herauszufinden, inwiefern dieser Prozess auch mit DAT-SPECT-Dateien möglich ist und wenn ja, mit welchem Erfolg. Es wurde zuerst die Programm-bibliothek TensorFlow™ (Version 1.7.0) heruntergeladen und installiert. Nach Definition der Bildauflösung auf 128x128 Pixel wurde die Netzwerkarchitektur definiert. Bei dem verwendeten Modell handelte es sich um verschiedene Ausführungen des MobileNet Version 1.0. Dieses ist ein faltendes künstliches neuronales Netzwerk bestehend aus 28 Schichten und in unterschiedlichen Modellfraktionen erhältlich.¹⁴⁰

Es wurde bereits mit dem ImageNet Bilddatensatz vortrainiert. Dieser liegt im Internet für Forschungszwecke frei verfügbar vor und besteht aus über 14 Millionen Bilddateien aus mehr als 21 000 Kategorien, denen die Dateien seitens der Entwickelnden per Hand zugeordnet worden sind.¹⁴¹

Es gab insgesamt vier verschiedene Netzwerkausführungen, die sich in der Größe der Modellfraktion unterschieden (1.0, 0.75, 0.50 oder 0.25). Anschließend wurden weitere Konfigurationen für das Training des Netzwerkes festgelegt. Die Eingabeschicht enthielt je 128x128x1 Knoten, d. h. insgesamt 16 384 Knoten. Außerdem wurde eine Ausgabeschicht hinzugefügt, die aus zwei Knoten entsprechend den beiden möglichen Ausprägungen der Variable „dopaminerges Defizit“ (positiv oder negativ) bestand. Danach wurde der jeweilige Datensatz definiert, mit dem das Netzwerk neu trainiert werden sollte. Die virtuelle Netzwerkkumgebung wurde mittels „activate“-Befehl im Terminal gestartet. Der Erfolg dieses Vorgangs zeigte sich daran, dass im Terminal anschließend die Abkürzung „venv“ für „virtual environment“ ([englisch]) vor der Eingabezeile erschien. Dies war technische Voraussetzung für die Durchführung des Trainings.

Anschließend wurde der Lernprozess mittels Ausführung des jeweiligen Trainingsskripts gestartet. Dabei teilte der Rechner den Lerndatensatz selbstständig in einen Trainingsdatensatz (80% der Bilddateien) und in einen Validierungsdatsatz (20% der Bilddateien). Die Trainingsdaten durchliefen das Modell und wurden dabei auf Unterschiede analysiert. Da sie in Unterordnern sortiert vorlagen, wurde dem Algorithmus vorgegeben, welche Bilder einer Kategorie zugehörig sind. Das Programm analysierte nun Bildpunkt für Bildpunkt, in welchen Charakteristika sich die Bilder eines Ordners ähnelten und in welchen sie sich von denen eines anderen unterschieden. Nach jedem Trainingsdurchlauf evaluierte das Modell die Bilder des bis dahin unbekanntes Validierungsdatsatzes anhand der Labels. Die Vorgänge des Lernens und der anschließenden internen Validierung wurde je nach vorher definierter Anzahl an Trainingsdurchläufen entweder 500 oder 1000 Mal wiederholt. Die Zugehörigkeitswahrscheinlichkeit zu einer Kategorie wurde als Dezimalzahl angegeben. Für die Trainingsprozesse wurde für jede Kombination aus Trainingsdatensatz, MobileNet-Version und Anzahl der Trainingsschritte ein Skript mit den entsprechenden Einstellungen und Dateipfaden erstellt, sodass insgesamt 64 verschiedene Skripte entstanden sind. Im Dateinamen wurden die entsprechenden Informationen zwecks besserer Unterscheidbarkeit ebenfalls verankert. Im Anhang ist das Trainingsskript für die MobileNet Fraktionsversion 1.0 mit den normalisierten und rotierten Bilddaten und einer Anzahl von 1000 Trainingsdurchläufen aufgeführt (Skript `runTraining_mobilenet_v1.0_normalized_Rotated.sh`). Mit den beschriebenen Trainingsschritten wurde also ein Prozess des Transfer-Lernens angewandt.

Nun folgte im Sinne einer Kreuzvalidierung der Testprozess mit den zuvor herausgenommenen Testdatensätzen. Dafür wurde für jeden Testdatensatz ein Skript in Python erstellt. Durch Ausführen des jeweiligen Testskripts durchlief jeder Testdatensatz das zuvor trainierte künstliche neuronale Netzwerk. Die dem Algorithmus bis dahin unbekanntes Testdaten wurden automatisiert evaluiert und kategorisiert. Die Wahrscheinlichkeit, mit der der Algorithmus dabei die Zugehörigkeit zu den beiden Kategorien (positiv und negativ) angab, sollte als Dezimalzahl mit fünf Kommastellen ausgegeben werden. Im Anhang ist das Testskript für den Datensatz aus normalisierten und rotierten Bilddateien exemplarisch aufgeführt (labelImage_normalized_Rotated.py). Nach Abschluss des Testvorgangs wurden die Klassifikationsergebnisse in kommaseparierter Weise ausgegeben. Anschließend wurden sie in tabellarischer Form im CSV-Format (comma separated value, [englisch]) abgespeichert und in Google® Drive übertragen. Als die durch den Algorithmus ermittelte diagnostische Kategorie wurde diejenige festgelegt, deren Zugehörigkeitswahrscheinlichkeit $>0,5$ lag. In Drive wurde für jede Kombination aus Testdatensatz, künstlichem neuronalen Netzwerk und Anzahl der Durchläufe ein Tabellenblatt mit den Klassifikationsergebnissen erstellt. Außerdem wurden für die anschließende Auswertung die Ergebnisse der visuellen Beurteilung eingetragen und erfasst, ob sie mit der durch den Algorithmus getroffenen Klassifikation identisch war. So lagen schlussendlich 64 Tabellen mit einer Größe von 8x41 Zellen vor. Deren Konfiguration wird in Abbildung 18 exemplarisch demonstriert. Anschließend wurden mit Hilfe der integrierten Excel®-Funktion sowie SPSS® die entsprechenden statistischen Parameter berechnet (siehe Kapitel 3.2.6).

					ID	VisualClassification	Correct?
flipped_1604.jpg	neg	0.99994	pos	0.00006	1604	nein	ja
flipped_1615.jpg	neg	0.95248	pos	0.04752	1615	ja	nein
flipped_693.jpg	neg	1	pos	0	693	nein	ja

Abbildung 18: Exemplarische Darstellung der dateiweisen Klassifikation

Die Vorgänge aus diesem und dem vorigen Kapitel werden noch einmal in Abbildung 19 zusammengefasst. Alle erwähnten Skripte wurden im Anhang beigefügt.

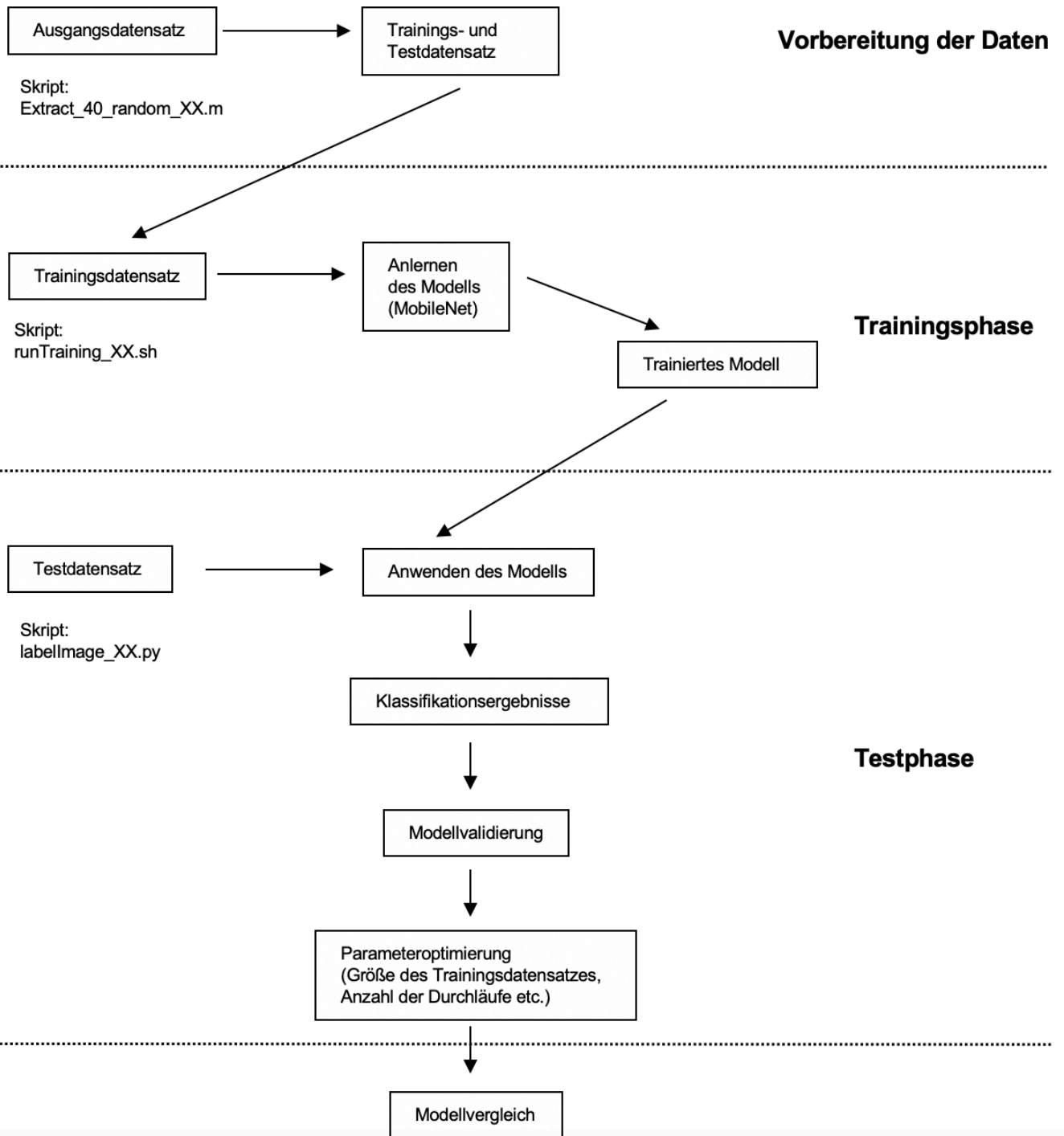


Abbildung 19: Zusammenfassung der Trainings- und Testprozesse

3.2.6. Statistische Analyse

In Google® Drive wurde für jede der Klassifikationsergebnisse enthaltenden Tabellen eine Konfusionsmatrix in Form einer Vier-Felder-Tafel mit der in Tabelle 3 dargestellten Konfiguration erstellt. RP und RN bezeichnen die Anzahl der richtig Positiven bzw. Negativen, FP und FN stehen für die Anzahl der falsch Positiven bzw. Negativen.

		Dopaminerges Defizit		Summe
		ja	nein	
Test	positiv	RP	FP	RP+FP
	negativ	FN	RN	FN+RN
Summe		RP+FN	FP+RN	RP+FN+FP+RN

Tabelle 3: Vier-Felder-Tafel

Mit Hilfe der insgesamt 64 Konfusionsmatrizen konnten nun die Parameter Genauigkeit, Sensitivität, Spezifität, positiver Prädiktionwert und negativer Prädiktionwert berechnet werden. Dies wurde innerhalb von Drive mittels einer integrierten Microsoft® Excel®-Anwendung, d. h. einem Tabellenkalkulationsprogramm, durchgeführt. Die Genauigkeit bezeichnet dabei die Richtigklassifikationsrate, d. h. die Anzahl der als positiv erkannten Positiven addiert mit den als negativ erkannten Negativen dividiert durch die Anzahl aller Bilder des jeweiligen Testdatensatzes, die 40 entsprach. Die Werte für die genannten statistischen Parameter wurden anschließend in Microsoft® Excel® eingetragen. Dort wurde für jeden Parameter eine Übersichtstabelle erstellt, die die jeweiligen Werte für die verschiedenen Netzwerke und Testdatensätze sowie für die beiden Anzahlen an Trainingsdurchläufen (500 vs. 1000) aufführt (siehe Tabelle 7- Tabelle 11 im Ergebnisteil). Für die statistische Analyse der erfassten Patientendaten wurde SPSS® 28 (Statistical Package for the Social Sciences, [englisch]) verwendet. Dabei wurde für die Analyse der Korrelation von Geschlecht und Vorliegen eines dopaminergen Defizits der Chi-Quadrat-Test angewandt, da es sich um zwei nominal skalierte Variablen handelt. Für die Bestimmung der Korrelation zwischen Alter und positivem bildmorphologischem Befund wurde der Eta-Korrelationskoeffizient ermittelt, da hier eine metrisch und eine nominal skalierte Variable vorliegen. Anschließend wurde eine univariate Varianzanalyse zur Bestimmung der Signifikanz durchgeführt.

4. Ergebnisse

4.1 Klinische Parameter

Nach manueller Durchsicht der heruntergeladenen Dateien stellten sich 442 Einzeldateien von insgesamt 442 Patient*innen (260 Männer, 182 Frauen) als geeignet für die weitere Verarbeitung und Analyse heraus. Abbildung 20 verbildlicht die Verteilung des Kollektivs nach Geschlecht sowie dem Vorliegen eines positiven Befundes im Sinne eines dopaminergen Defizits (insgesamt 192 negativ, 250 positiv). In Tabelle 4 sind die erfassten klinischen Parameter mit Lage- und Streuparametern aufgeführt.

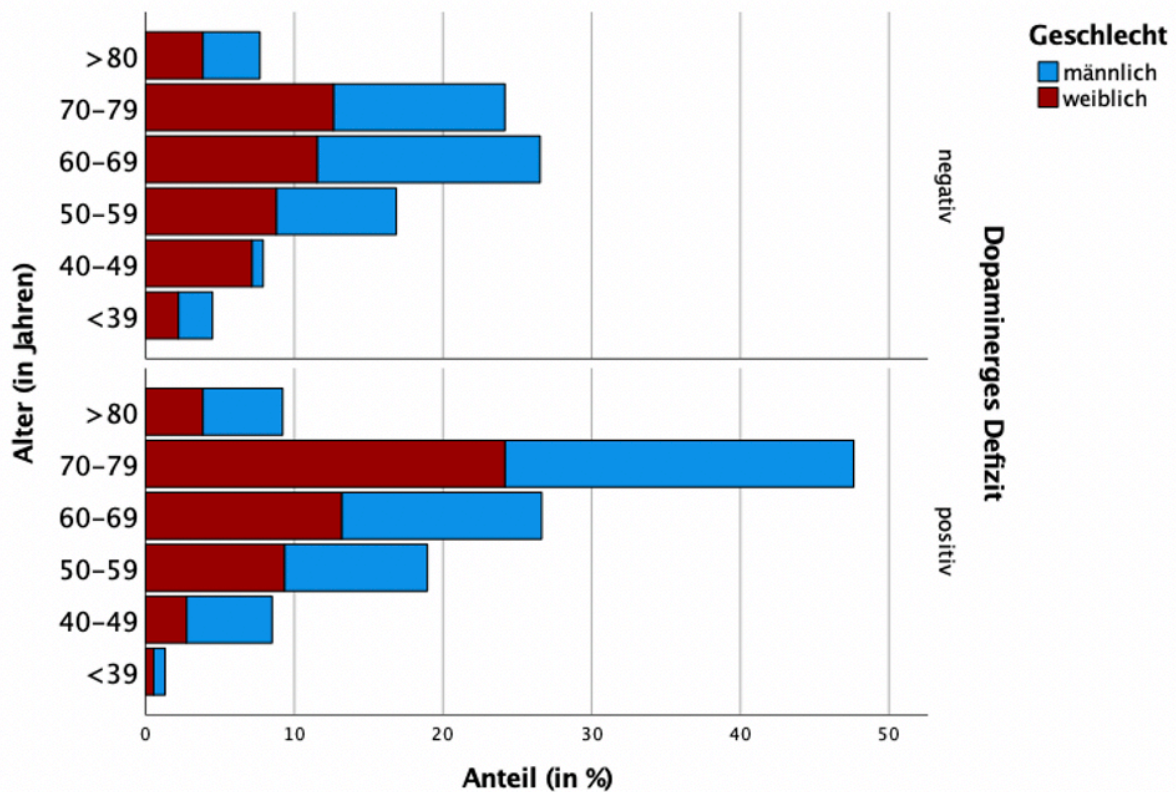


Abbildung 20: Altersabhängige Verteilung nach Geschlecht und dopaminergem Defizit

Geschlecht	m	w	Dopaminerges Defizit	neg	pos	Alter					
						Anzahl	Mittelwert	Maximum	Minimum	Standardabweichung	Varianz
				neg	pos	108	65,6	86,1	30,4	12,1	146,2
				neg	pos	152	66,8	84,0	38,6	11,2	124,9
				neg	pos	84	62,3	86,2	22,4	13,7	187,7
				neg	pos	98	67,2	83,4	35,0	10,2	105,0

Tabelle 4: Deskriptive Statistik der klinischen Parameter

Es zeigte sich eine statistisch signifikante starke positive Korrelation zwischen dem Alter der Patient*innen und dem Vorliegen eines dopaminergen Defizits in der Dopamin-Transporter-SPECT (Eta-Korrelationskoeffizient = 0.995, $p = 0.014$). Zwischen männlichen und weiblichen Personen bestand jedoch kein statistisch signifikanter Unterschied in Hinblick auf einen auffälligen Befund ($p = 0.335$).

4.2 Verarbeitung der Bilddateien

Nach dem Herunterladen der Bilddaten mussten diese in eine mit Machine Learning-Systemen kompatible Form gebracht werden. Es gelang, die heruntergeladenen dreidimensionalen Bilddateien in ein zweidimensionales Format in Form von Maximumintensitätsprojektionen (MIP) zu überführen. Im übertragenen Sinne ähnelt der Vorgang dem einer Blütenpresse, bei dem es ebenfalls zu einer Volumenreduktion des ursprünglichen Gebildes kommt, seine räumliche Struktur aber weiterhin zu erahnen ist. Abbildung 21 zeigt die dreidimensionale Struktur des Striatums sowie die exemplarische Darstellung der Bilddaten vor und nach diesem Prozess.

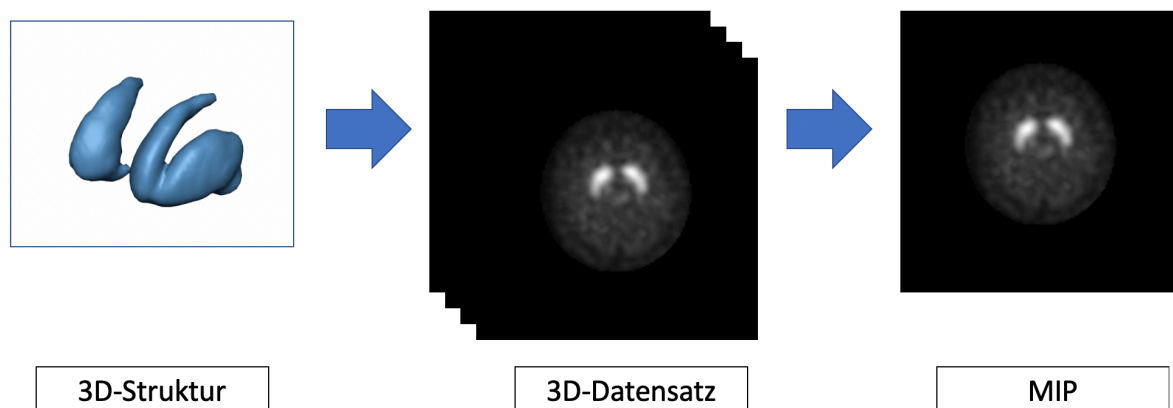


Abbildung 21: Exemplarische Erstellung der Maximumintensitätsprojektionen (MIP)

Außerdem waren die räumliche Normalisierung sowie die anschließende Augmentierung der Datensätze mittels Spiegelung und Rotation erfolgreich. Im Folgenden wird in Abbildung 22 die Überführung der ursprünglichen Bilder, d. h. der nicht normalisierten, nicht gespiegelten und nicht rotierten Bilddaten (gelb markiert) in die verschiedenen Formate exemplarisch demonstriert. Dabei wurde für die Erstellung der Abbildung ein Bilddatensatz gewählt, der in Hinblick auf das Vorliegen eines dopaminergen Defizits als positiv gewertet wurde. Durch die pathologische Asymmetrie der Putamina kommt insbesondere der Prozess der Spiegelung deutlicher zur Geltung.

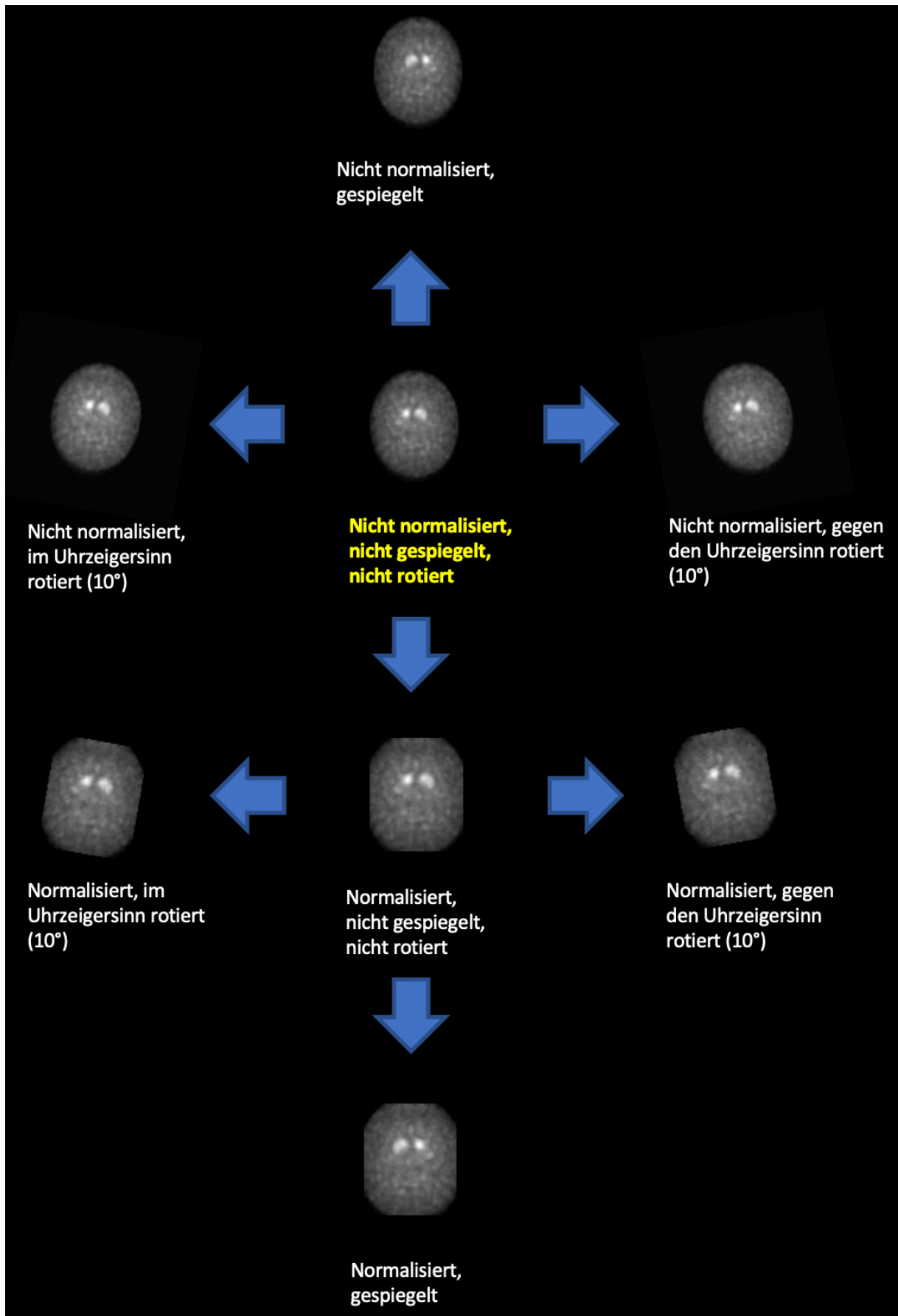


Abbildung 22: Datenaugmentierung

Die Normalisierung und Augmentierung der Daten hat zu einer drastischen Vergrößerung des initial zur Verfügung stehenden Ausgangsbilddatensatzes (nicht normalisiert, nicht gespiegelt, nicht rotiert) geführt, der die Grundlage für die anschließende Erstellung der Trainings- und Testdatensätze bildete. In Tabelle 5 ist die Anzahl der jeweiligen Bilddateien pro Datensatz aufgeführt. Es zeigt sich, dass die Datenaugmentierung zu einer Erhöhung der Bildanzahl um den Faktor acht geführt hat. Die normalisierten Datensätze umfassen gleich viele Bilder wie die nicht normalisierten, sodass für die spätere Analyse eine direkte Vergleichbarkeit gegeben ist. Die Anzahlen der Bilddateien der daraus hervorgehenden Trainingsdatensätze sind in Tabelle 6 dargestellt. Anhand des Dateiumfangs ist ersichtlich, dass die Extrahierung von jeweils 20 Einzeldateien pro Kategorie bzw. Unterordner erfolgreich umgesetzt wurde. Daraus ergibt sich der gewünschte Umfang von je 40 Bilddateien (20 Bilder mit positivem und 20 mit negativem Befund) für die Erstellung der korrespondierenden Testdatensätze, sodass die Gesamtbildanzahl der Trainingsdatensätze im Vergleich zu den Ausgangsdatsätzen um 40 reduziert ist.

Dateianzahlen der Ausgangsdatensätze			
Datensatz	Anzahl Bilder pos	Anzahl Bilder neg	Anzahl Bilder gesamt
Nicht normalisiert, nicht gespiegelt, nicht rotiert	250	192	442
Nicht normalisiert, gespiegelt	250	192	442
Nicht normalisiert, rotiert	500	384	884
Alle nicht Normalisierten	1000	768	1768
Normalisiert, nicht gespiegelt, nicht rotiert	250	192	442
Normalisiert, gespiegelt	250	192	442
Normalisiert, rotiert	500	384	884
Alle Normalisierten	1000	768	1768

Tabelle 5: Dateianzahlen der Ausgangsdatensätze

Dateianzahlen der Trainingsdatensätze			
Datensatz	Anzahl Bilder pos	Anzahl Bilder neg	Anzahl Bilder gesamt
Nicht normalisiert, nicht gespiegelt, nicht rotiert	230	172	402
Nicht normalisiert, gespiegelt	230	172	402
Nicht normalisiert, rotiert	480	364	844
Alle nicht Normalisierten	980	748	1728
Normalisiert, nicht gespiegelt, nicht rotiert	230	172	402
Normalisiert, gespiegelt	230	172	402
Normalisiert, rotiert	480	364	844
Alle Normalisierten	980	748	1728

Tabelle 6: Dateianzahlen der Trainingsdatensätze

4.3 Ergebnisse der Klassifikation

Es standen vier Ausführungen künstlicher neuronaler Netzwerke (MobileNet Version 1.0 Fraktionsversion 0.25, 0.50, 0.75 und 1.0) für das Training und das anschließende Testen mit den zuvor erstellten Test- und Trainingsdatensätzen zur Verfügung. Da jedes Netzwerk mit den acht vorhandenen Testdatensätzen (à 40 Einzeldateien) einmal nach 500 und einmal nach 1000 Trainingsdurchläufen geprüft wurde, ergaben sich die Klassifikationsergebnisse in Form von insgesamt 64 Excel-Tabellen in Google Drive. Im Folgenden ist in Tabelle 7 die Genauigkeit im Sinne der Richtigklassifikationsrate dargestellt. Tabelle 8 und Tabelle 9 führen die Sensitivität bzw. die Spezifität der Klassifikationsergebnisse auf. Schließlich sind in Tabelle 10 und Tabelle 11 die positiven bzw. negativen prädiktiven Werte ersichtlich. Jede Zelle unterhalb der jeweiligen MobileNet-Version entspricht der Durchführung eines Trainings- und anschließendem Testvorgangs eines Netzwerkes. Die Anzahl der Trainingsdurchläufe ist in Stück angegeben. Die Ergebnisse der Klassifikationen sind in Prozent aufgeführt und wurden auf ganze Zahlen gerundet.

Genauigkeit (Richtigklassifikationsrate)						
Datensatz	Anzahl Trainingsdurchläufe	Version 0.25	Version 0.50	Version 0.75	Version 1.0	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	500	88	68	75	85	
Nicht normalisiert, gespiegelt		85	80	78	85	
Nicht normalisiert, rotiert		78	85	90	70	
Alle nicht Normalisierten		83	60	73	90	
Normalisiert, nicht gespiegelt, nicht rotiert		73	90	75	83	
Normalisiert, gespiegelt		83	68	83	90	
Normalisiert, rotiert		90	85	83	75	
Alle Normalisierten		80	78	83	78	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	1000	80	88	78	70	
Nicht normalisiert, gespiegelt		83	80	73	85	
Nicht normalisiert, rotiert		70	85	85	70	
Alle nicht Normalisierten		85	90	88	80	
Normalisiert, nicht gespiegelt, nicht rotiert		75	83	83	83	
Normalisiert, gespiegelt		80	85	75	88	
Normalisiert, rotiert		83	73	78	83	
Alle Normalisierten		85	85	80	80	

Tabelle 7: Genauigkeit (Richtigklassifikationsrate) der Klassifikationsergebnisse

Sensitivität						
Datensatz	Anzahl Trainingsdurchläufe	Version 0.25	Version 0.50	Version 0.75	Version 1.0	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	500	95	95	50	100	
Nicht normalisiert, gespiegelt		95	90	85	100	
Nicht normalisiert, rotiert		80	95	80	40	
Alle nicht Normalisierten		70	100	100	100	
Normalisiert, nicht gespiegelt, nicht rotiert		50	90	55	90	
Normalisiert, gespiegelt		85	35	65	95	
Normalisiert, rotiert		80	90	85	95	
Alle Normalisierten		65	65	70	90	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	1000	95	85	100	100	
Nicht normalisiert, gespiegelt		95	75	55	95	
Nicht normalisiert, rotiert		40	95	90	100	
Alle nicht Normalisierten		100	95	100	100	
Normalisiert, nicht gespiegelt, nicht rotiert		90	90	85	90	
Normalisiert, gespiegelt		90	90	90	95	
Normalisiert, rotiert		90	95	85	95	
Alle Normalisierten		80	80	65	90	

Tabelle 8: Sensitivität der Klassifikationsergebnisse

Spezifität						
Datensatz	Anzahl Trainingsdurchläufe	Version 0.25	Version 0.50	Version 0.75	Version 1.0	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	500	80	40	100	70	
Nicht normalisiert, gespiegelt		75	70	70	70	
Nicht normalisiert, rotiert		75	75	100	100	
Alle nicht Normalisierten		95	20	45	80	
Normalisiert, nicht gespiegelt, nicht rotiert		95	90	95	75	
Normalisiert, gespiegelt		80	100	100	85	
Normalisiert, rotiert		100	80	80	55	
Alle Normalisierten		95	90	95	65	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	1000	65	90	55	40	
Nicht normalisiert, gespiegelt		70	85	90	75	
Nicht normalisiert, rotiert		100	75	80	40	
Alle nicht Normalisierten		70	85	75	60	
Normalisiert, nicht gespiegelt, nicht rotiert		60	75	80	75	
Normalisiert, gespiegelt		70	80	60	80	
Normalisiert, rotiert		75	50	70	70	
Alle Normalisierten		90	90	95	70	

Tabelle 9: Spezifität der Klassifikationsergebnisse

Positiver Prädiktionswert						
Datensatz	Anzahl Trainingsdurchläufe	Version 0.25	Version 0.50	Version 0.75	Version 1.0	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	500	83	61	100	77	
Nicht normalisiert, gespiegelt		79	75	74	77	
Nicht normalisiert, rotiert		76	79	100	100	
Alle nicht Normalisierten		93	56	65	83	
Normalisiert, nicht gespiegelt, nicht rotiert		91	90	92	78	
Normalisiert, gespiegelt		81	100	100	86	
Normalisiert, rotiert		100	82	81	68	
Alle Normalisierten		93	87	93	72	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	1000	73	89	69	63	
Nicht normalisiert, gespiegelt		76	83	85	79	
Nicht normalisiert, rotiert		100	79	82	63	
Alle nicht Normalisierten		77	86	80	71	
Normalisiert, nicht gespiegelt, nicht rotiert		69	78	81	78	
Normalisiert, gespiegelt		75	82	69	83	
Normalisiert, rotiert		78	66	74	76	
Alle Normalisierten		89	89	93	75	

Tabelle 10: Positive Prädiktionswerte der Klassifikationsergebnisse

Negativer Prädiktionswert						
Datensatz	Anzahl Trainingsdurchläufe	Version 0.25	Version 0.50	Version 0.75	Version 1.0	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	500	94	89	67	100	
Nicht normalisiert, gespiegelt		94	88	82	100	
Nicht normalisiert, rotiert		79	94	83	63	
Alle nicht Normalisierten		76	100	100	100	
Normalisiert, nicht gespiegelt, nicht rotiert		66	90	68	88	
Normalisiert, gespiegelt		84	61	74	94	
Normalisiert, rotiert		83	89	84	92	
Alle Normalisierten		73	72	76	87	
Nicht normalisiert, nicht gespiegelt, nicht rotiert	1000	93	86	100	100	
Nicht normalisiert, gespiegelt		93	77	67	94	
Nicht normalisiert, rotiert		63	94	89	100	
Alle nicht Normalisierten		100	94	100	100	
Normalisiert, nicht gespiegelt, nicht rotiert		86	88	84	88	
Normalisiert, gespiegelt		88	89	86	94	
Normalisiert, rotiert		88	91	82	93	
Alle Normalisierten		82	82	73	88	

Tabelle 11: Negative Prädiktionswerte der Klassifikationsergebnisse

Zusammenfassend lässt sich feststellen, dass sich die Genauigkeit (Richtigklassifikationsrate) im Bereich von 60-90% bewegte. Dabei wurde der höchste Wert von 90% sowohl bei niedrigeren wie auch bei den höheren Fraktionsversionen des künstlichen neuronalen Netzwerkes erreicht, bei maximaler Netzwerkausdehnung (1.0) wurde er jedoch sogar bei zwei Datensätzen erreicht. Die Anzahl der Trainingsdurchläufe wie auch die Normalisierung der Dateien schien die Genauigkeit nicht wesentlich zu beeinflussen. Die Sensitivität der Klassifikationsergebnisse zeigte starke Schwankungen zwischen den Netzwerkversion von 35-100%. Insbesondere bei Fraktionsversion 1.0 wurden beachtliche Werte erreicht. So wurden die Datensätze, die alle nicht normalisierten Dateien zusammen enthielten, sowohl bei 500 wie auch bei 1000 Trainingsdurchläufen mit maximaler Sensitivität klassifiziert. Auch die Spezifität schwankte zwischen 20-100%. Vor allem Version 0.50 zeigte bei reduzierter Durchlauftrate und nicht normalisierten Dateien unbefriedigende Ergebnisse, während Version 0.75 die normalisierten Dateien deutlich spezifischer erkannte.

Die positiven Prädiktionswerte lagen zwischen 56-100%, wobei Version 0.75 bei reduzierter Laufrate und bei Analyse nicht normalisierter Dateien überzeugen konnte. Die negativen Prädiktionswerte fielen mit Werten zwischen 63-100% aus. Hier lieferte die maximale Netzwerkausdehnung die besten Ergebnisse, vor allem bei hoher Durchlauftrate und bei nicht normalisierten Dateien.

5. Diskussion

Das dieser Arbeit zu Grunde liegende Projekt beschäftigte sich mit der automatisierten Klassifikation von Dopamin-Transporter-SPECT. Dabei wurde eine Methodik des tiefen Lernens aus dem Bereich der künstlichen Intelligenz eingesetzt. Die zu prüfende Pathologie bestand im Vorliegen einer Nuklidminderbelegung im Bereich der dopaminergen Neurone, die vorrangig in den Basalganglien zu finden sind. Beim Einsatz eines faltenden künstlichen neuronalen Netzwerkes in unterschiedlichen Konstellationen wurden teils beeindruckende Resultate erzielt.

5.1 Datengrundlage

Die erhobenen Daten zeigten in ihrer klinischen Zusammensetzung ein häufigeres Vorliegen des männlichen Geschlechts sowohl im Gesamtkollektiv wie auch in der isolierten Betrachtung der auffälligen Bildbefunde. Dieser Aspekt sowie die altersabhängige positive Korrelation pathologischer Befunde entsprechen typischen Konstellationen beim Morbus Parkinson und signalisieren eine epidemiologische Repräsentativität der hier vorliegenden Daten.⁷ Bei dem verwendeten Datensatz handelte es sich außerdem um einen reinen Bilddatensatz. Wie in den meisten Studien zur automatisierten Klassifikation medizinischer Bilddateien fand keine Korrelation zur Untersuchungsindikation oder klinischen Symptomatik der untersuchten Person statt.^{132,134,142} Eine Verknüpfung solcher Informationen mit der getroffenen Klassifikation könnte zur Stützung bzw. Infragestellung einer Diagnose beitragen. So berücksichtigten Leung et al. in ihrer Publikation zusätzlich den Wert auf der MDS-UPDRS-III-Skala sowie Faktoren wie Alter, Geschlecht und Krankheitsdauer und erzielten dabei signifikant bessere Ergebnisse als bei einer rein bildbasierten Analyse.¹⁴³ Außerdem wäre es denkbar, das bislang binäre diagnostische Spektrum der DAT-SPECT (Vorliegen bzw. Nicht-Vorliegen eines dopaminergen Defizits) mit Hilfe klinischer Informationen zu erweitern.

Die Daten wiesen eine balancierte Struktur mit ausreichend ähnlicher Verteilung der Pathologien auf, wobei eine mäßige Dominanz der auffälligen Bildbefunde vorlag. Diese lässt sich auf die Indikation der DAT-SPECT zurückführen, da sie i. d. R. bei Personen mit einer klinischen Symptomatik im Sinne eines Parkinson- oder Tremor-Syndroms durchgeführt wird.^{79,144} Demnach ist dieses Kollektiv vorselektiert und damit ein auffälliger Befund deutlich wahrscheinlicher zu erwarten als in der Normalbevölkerung. Dies deckt sich mit den Erkenntnissen der Publikation von de la Fuente-Fernández, der einen deutlichen Zusammenhang zwischen einem Parkinsonismus und einer auffälligen DAT-SPECT beschreibt.¹⁴⁵ Auch in ähnlichen Arbeiten wie der vorliegenden wurde ein weitgehend ausgeglichenes Verhältnis unauffälliger zu auffälligen Daten angestrebt, um dem Algorithmus

genug Material zum Erlernen der Differenzierung zwischen den diagnostischen Kategorien bereitzustellen. In den Publikationen von Kim et al. und Wenzel et al. wurde für einen ausgeglichenen Lerndatensatz ein Verhältnis von 1:1 erzielt, indem ein Teil der überrepräsentierten auffälligen Befunde bewusst nicht mit einbezogen wurde.^{132,133} Der von Zhang und Kagen verwendete Ausgangsdatsatz wies einen Anteil von 77% pathologischen DAT-SPECTs auf. Die genaue Zusammensetzung der Lern- und Testdaten wird in dem Artikel nicht beschrieben, allerdings ist bei einer derart ausgeprägten Überrepräsentation der auffälligen Befunde in den Ausgangsdaten eine ausgeglichene Zusammensetzung sehr unwahrscheinlich.¹³⁴ Es stellt sich insbesondere die Frage nach der Verlässlichkeit der Spezifität. Außerdem wurden in den Publikationen von Kim et al., Wenzel et al. sowie Zhang und Kagen die Testdaten auf zufällige Weise extrahiert.¹³²⁻¹³⁴ In der vorliegenden Arbeit war dies ebenso der Fall, es wurde aber ein 1:1-Verhältnis von unauffälligen zu auffälligen Befunden gewahrt, um zu ausgeprägte Verzerrungen zu vermeiden. Dies wäre bei Testdaten der Fall gewesen, die zufällig überwiegend, z.B. zu 95%, aus patho- bzw. physiologischen Bilddaten bestünden. Diese wären in Anbetracht der dann sehr unterschiedlichen Zusammensetzung zwischen Trainings- und Testdatensätzen nicht repräsentativ. Zudem wäre eine verlässliche Interpretation der statistischen Parameter deutlich eingeschränkt.

Hinsichtlich der Datenverarbeitung verlief unter anderem die Augmentierung erfolgreich, ebenso wie beispielsweise in den Veröffentlichungen von Kim et al., Chen et al. und Hussain et al. Dort wurden die Bilddateien ebenfalls rotiert.^{132,146,147} Choi und Ha et al. sowie Hussain et al. führten ebenso eine Spiegelung entlang der Vertikalen durch.^{135,147} Ebenso gelang es, ein mit dem Algorithmus kompatibles Dateiformat zu finden, sodass die Trainings- und Testprozesse überhaupt möglich wurden. Durch Verwendung der MIPs konnten alle Schichten des jeweiligen Bilddatensatzes in das tiefe Lernen einbezogen werden. In der Publikation von Zhang und Kagen wurde dagegen pro Aufnahme lediglich eine einzelne axiale Schicht für das Netzwerktraining extrahiert, sodass die Informationen der Ebenen kranial beziehungsweise kaudal dieser Schicht nicht mit in die Analyse einfließen.¹³⁴ Dies ist insofern als kritisch zu beurteilen, als dass die Bewertungsgrundlage pro Patient*in für den dort verwendeten Algorithmus deutlich geringer war als in der vorliegenden Arbeit und als es im klinischen Alltag der Fall ist. Daher ist bei solch einem Vorgehen der Vergleich der Klassifikationsleistung zu menschlichen Beurteiler*innen nur sehr eingeschränkt möglich. In der vorliegenden Arbeit wurden dagegen die kompletten DAT-SPECT-Aufnahmen verwendet wie auch in den Veröffentlichungen von Kim et al. und Wenzel et al.^{132,133} Da sich der diagnostisch relevante Bereich primär auf die Basalganglien bezog, wäre eine sogenannte Region-of-Interest-Analyse (ROI-Analyse) als weiterführende Studie denkbar. Dabei werden durch verschiedene Methoden die diagnostisch vorrangigen Areale bestimmt.¹⁴⁸ Es könnten z.B. manuell

Markierungen vorgenommen, die nuklidbelegten Bildanteile ausgeschnitten oder Grenzwerten (z.B. in Form von Graustufen) festgelegt werden. Die extrahierten Dateien könnten anschließend als Datengrundlage für den Lernprozess verwendet werden. Dies würde eine Reduktion des Rechenaufwandes durch kleinere Dateien sowie weniger Verzerrungspotential durch Konzentration auf relevante Areale bedeuten. Außerdem wäre es denkbar, das Netzwerk direkt mit den dreidimensionalen Rohdaten zu trainieren, ähnlich wie es Leung et al. in ihrer Publikation vorstellen.¹³⁷ Zu bedenken ist allerdings, dass viele Plattformen nur eingeschränkt mit volumetrischen Daten kompatibel sind, wie es auch in der vorliegenden Arbeit mit TensorFlow™ der Fall war.

Der Umfang der Datengrundlage bewegte sich im mittelgroßen Bereich. In den Arbeiten von Kim et al. sowie Ortiz et al. wurden mit insgesamt 190 respektive 269 Einzeldateien deutlich kleinere Datenmengen einbezogen.^{132,149} Zhao et al. schlossen Daten von insgesamt 1017 Personen ein. Mit diesem Datenumfang gelang sogar eine Untersuchung einer automatisierten Differentialdiagnostik des Parkinsonismus.¹³⁸ Allerdings gibt es einige Veröffentlichungen mit deutlich höheren Bildzahlen, vor allem wenn es um andere diagnostische Verfahren im Allgemeinen bzw. bildgebende Modalitäten im Speziellen geht. In der Studie von Esteva et al. z.B. wurden insgesamt 129 450 Einzeldatensätze von dermatoskopischen Aufnahmen bzw. Fotografien einbezogen. Allerdings fällt auf, dass es insgesamt 2032 verschiedene Erkrankungen einbezogen wurden, die teilweise in Kategorien zusammengefasst wurden. Dies lässt eine teilweise recht kleine Datenmenge pro Erkrankung vermuten, sodass sich hier die Frage nach einer verlässlichen Generalisierbarkeit stellt.¹⁵⁰ In den radiologischen Arbeiten von Rajpurkar et al. wurden 112 120 Röntgen-Thoraces bzw. in der Veröffentlichung von Song et al. 4581 CT-Thoraces benutzt, wobei es sich ebenfalls im Vergleich zur vorliegenden Arbeit um deutlich größere Datenmengen handelt.^{142,151} Allerdings gilt es zu bedenken, dass Verfahren wie z.B. Fotografien von Läsionen, das Röntgen oder die Computertomographie wesentlich breiter verfügbar und deutlich weniger aufwändig in der Durchführung als die DAT-SPECT sind. Dafür sind die Indikationen und damit die potentiellen Kategorien deutlich zahlreicher, sodass eine höhere Gesamtanzahl an Bilddateien nicht automatisch eine höhere Anzahl an Dateien für eine spezifische diagnostische Kategorie bedeutet. In der Publikation von Esteva et al. gab es 757 mögliche Klassifikationskategorien¹⁵⁰, in der von Rajpurkar et al. waren es immerhin 14 Stück.¹⁴² Die binäre Klassifikation der DAT-SPECT ist im Vergleich übersichtlicher und weist daher weniger mögliche Kategorien auf, zwischen denen es potentiell zu Verwechslungen kommen könnte.

In den meisten Publikationen wurde ein einzelner Datensatz bzw. jeweils nur eine Ausführung eines Datensatzes als Basis für das Netzwerktraining bzw. die -testung verwendet.^{132,134,135}

In der vorliegenden Arbeit gab es dagegen jeweils acht verschiedene Trainings- und Testdatensätze. Außerdem variierte die Größe der Datensätze, sodass Vergleiche zwischen unterschiedlichen Bilddatenumfängen und den korrespondierenden Klassifikationsleistungen ermöglicht werden konnten. Auch ein Vergleich zwischen normalisierten und nicht normalisierten Bilddateien wurde bislang nicht beschrieben. Zudem ist bislang keine Studie bekannt, in der das hier verwendete künstliche neuronale Netzwerk MobileNet Version 1.0 zur Beurteilung von SPECTs geschweige denn DAT-SPECT-Dateien eingesetzt wurde.

5.2 Klassifikationsleistung

Die Klassifizierung der recht umfangreichen Bilddatensätze war schließlich technisch möglich und erfolgreich. Hier zeigte sich die Fähigkeit von künstlichen neuronalen Netzwerken, auch große Datenmengen ohne Probleme verarbeiten zu können.¹⁰⁰ Die Klassifikationsleistung ist bei angemessen balanciertem Ausgangsdatsatz als repräsentativ zu beurteilen. Sie variierte recht stark zwischen den verschiedenen Testdatensätzen, Netzwerkfraktionen und Anzahl der Trainingsdurchläufe. Insbesondere die leistungsstärkeren Kombinationen zeigten ähnliche Ergebnisse wie bereits publizierte Arbeiten. So lag in der Veröffentlichung von Choi et al. die Genauigkeit des verwendeten Netzwerkes bei 96.0% bei der Bewertung des externen Testdatensatzes.¹³⁵ In der Arbeit von Kim et al. wurde ebenfalls ein faltendes neuronales Netzwerk trainiert und dabei eine Sensitivität von 96.3% und eine Spezifität von 66.7% erzielt.¹³² Ähnliche Resultate sind in der vorliegenden Arbeit mit der Netzwerkfraktion 0.75 und dem rotierten, nicht normalisierten Datensatz sowie der Fraktion 1.0 und allen nicht normalisierten Dateien bei einer Durchlaufrate von 500 erreicht worden. Auch bei der Kombination aus Fraktion 0.50 bei allen nicht normalisierten Dateien und einer Laufzahl von 1000 wurde eine Genauigkeit von 90% erreicht. Bei der Sensitivität wurde mehrfach ein Ergebnis von 100% erreicht, wobei dies vor allem bei den Netzwerkfraktion 0.75 und 1.0 bei nicht normalisierten Datensätzen und bei beiden Laufraten der Fall war. Auch in Hinblick auf die Spezifität wurden mehrmals 100% erreicht. Hier gab es weniger klare Tendenzen, am ehesten wurde dieser Wert bei normalisierten Daten und einer Laufrate von 500 erreicht. Es fällt ferner auf, dass in dem meisten Publikationen ein einzelnes Netzwerk bzw. ein Netzwerk in einer einzigen Ausführung verwendet wird und nicht wie hier verschiedene Varianten.^{132-135,142} Ortiz et al. verglichen zwei verschiedene faltende neuronale Netzwerke, jedoch mit recht unterschiedlichen Architekturen, sodass sich die Frage nach der Vergleichbarkeit stellt.¹⁴⁹

Im Vergleich zur menschlichen Beurteilung sind die Aspekte, die zur Entscheidung für eine diagnostische Klassifizierung beitragen, bei der Verwendung von Deep Learning-Anwendungen sehr viel schwerer nachzuvollziehen. Es handelt sich gewissermaßen um eine „Black Box“ [englisch], da die genauen Entscheidungen, die durch das künstliche neuronale

Netzwerk getroffen wurden, nicht identifizierbar sind, während ein Mensch seine Gedankengänge stets erläutern kann. Das Gebiet, das sich mit diesem Problem beschäftigt, wird als „explainable AI“ [englisch] bezeichnet.¹⁵² Ein möglicher Ansatz für Bilddateien besteht in der Erstellung sogenannter „Class Activation Maps“ [englisch], welche farbkodiert die Areale darstellen, die mehr bzw. weniger zur kategorialen Entscheidung eines Bildes beigetragen haben.¹⁵³ Solche wurden beispielsweise in der Publikation von Ha et al. erstellt.¹²⁷ Da sie dem visuellen Eindruck nach den Aufnahmen einer Wärmebildkamera ähneln, werden sie auch „Heat Maps“ [englisch] genannt. Allerdings bleibt auch dann im Unklaren, aus welchem genauem Grund manche Regionen mehr bzw. weniger für die Beurteilung relevant waren.¹⁵³ Eine genaue Nachvollziehbarkeit ergibt sich auch dann nicht, sodass hier ein Vorteil in der Diagnostik durch den Menschen liegt. Dafür sind die Klassifikationsergebnisse von künstlichen neuronalen Netzwerken objektiv und unabhängig von der Tagesform. Des Weiteren gibt der vorgestellte Algorithmus das jeweilige Klassifikationsergebnis in Form von Wahrscheinlichkeiten für die Zugehörigkeit zu einer der diagnostischen Kategorien an. So wie menschliche Beurteiler*innen sich in ihrer Diagnose mehr oder weniger sicher sein können, geben also auch die künstlichen neuronalen Netzwerke eine (prozentuale) Sicherheit für ihre Einschätzung an. Diese Sicherheit erlaubt die seitens des Netzwerkes getroffene Klassifikation quantifizieren und damit besser einschätzen zu können. Anwendungen des tiefen Lernens könnten daher ähnlich einem zweiten Augenpaar zur Überprüfung und insbesondere bei unklaren Befunden als Zweitmeinung herangezogen werden. Eine Gemeinsamkeit besteht darin, dass sowohl Algorithmen mit zunehmenden Lerndaten eine höhere diagnostische Sicherheit erlangen, wie auch Menschen Aufgaben mit zunehmender Erfahrung tendenziell besser lösen können.¹⁰⁰

5.3 Juristische Aspekte

Aus juristischer Sicht stellen sich bei Anwendungen des tiefen Lernens insbesondere aus medienprodukte-, haftungs- sowie datenschutzrechtlicher Perspektive Fragen. Bei der Einstufung als Medizinprodukt sind die Rahmenbedingungen in der Medizinprodukteverordnung („Medical Device Regulation“, [englisch], MDR) sowie im Medizinprodukte-recht-durchführungsgesetz (MPDG) festgesetzt.¹⁵⁴ Nach der MDR erfolgt eine Einordnung einer Software als Medizinprodukt, wenn sie medizinische Daten erstellt, verarbeitet, analysiert oder modifiziert.¹⁵⁵ Die in der vorliegenden Arbeit entwickelten künstlichen neuronalen Netzwerke würden also im Falle einer klinischen Implementierung unter die MDR und das MPDG fallen.^{154,155} Alle Medizinprodukte müssen ein Konformitätsbewertungsverfahren durchlaufen, welches für Software-Medizinprodukte unter anderem Wiederholbarkeit und Zuverlässigkeit fordert.^{156,157} Künstliche neuronale Netzwerke

unterliegen allerdings durch stetige Anpassungen einer gewissen Dynamik, sodass exakte Wiederholungen im Widerspruch zu ihrer Funktionsweise stehen.¹⁵⁴

Haftungsrechtlich sind sowohl der Hersteller von KI-Systemen sowie die Anwender*innen verantwortlich. Auf Seiten des Herstellers kommt das Produkthaftungsgesetz (ProdHaftG) zum Tragen, welches ihn unter anderem im Falle der Schädigung der Gesundheit einer Person infolge eines fehlerhaften Produktes schadensersatzpflichtig macht.¹⁵⁸ Bei Medizinprodukten gelten aufgrund der besonders relevanten Folgen bei etwaigen Fehlern hohe Sicherheitsanforderungen zur Minimierung von Risiken. Auch hier stellt die eingeschränkte Nachvollziehbarkeit autonomer Entscheidungsvorgänge bei künstlichen neuronalen Netzwerken eine erhebliche Hürde dar, da es z.B. bei fehlerhaften Lernprozessen kaum möglich ist, die entscheidende Fehlerquelle aufzudecken. Es ist fraglich, inwiefern bei diagnostischen Anwendungen eine falsche Diagnose mit konsekutiv fehlerbehafteter Behandlung als haftungsrechtlich relevante Gesundheitsschädigung zu werten wäre. In Bezug auf Ärzt*innen, die KI-enthaltende Medizinprodukte einsetzen, besteht wie für andere medizinische Maßnahmen eine Aufklärungs- sowie Sorgfaltspflicht.¹⁵⁹ Im Falle von tiefem Lernen wäre eine eigenständige Aufklärung vor allem dann relevant, wenn dadurch die Durchführung einer Maßnahme maßgeblich beeinflusst würde. Dies wäre z.B. bei Roboter-assistierten Operationen der Fall. Im Bereich der medizinischen Bildbeurteilung erfolgt der Einsatz von Algorithmen jedoch zur besseren Diagnosestellung, die ärztliche Indikationsstellung und therapeutische Folgerungen bleiben davon unberührt. Trotzdem wäre ein Hinweis auf den Einsatz von KI sinnvoll, um der Aufklärungspflicht vollumfänglich Rechnung zu tragen. Außerdem dürfen Behandelnde ein Medizinprodukt nach §12 des MPDG bei begründetem Verdacht auf eine Gefährdung des/der Patient*in nicht einsetzen.¹⁶⁰ Während sich z.B. die eingespeisten Patientendaten noch relativ gut auf Korrektheit kontrollieren lassen, ist die Bewertung einer autonom getroffenen Aussage deutlich schwerer zu bewerkstelligen. Es ist fraglich, wie realistisch die Sorgfaltspflicht in diesem Fall befolgt werden kann.

Die datenschutzrechtlichen Rahmenbedingungen sind vor allem in der Datenschutzgrundverordnung (DS-GVO) hinterlegt. Sie gilt für Algorithmen, wenn personenbezogene Informationen verwendet werden.¹⁶¹ Dazu gehören Grundsätze wie Transparenz, Zweckgebundenheit und Datenminimierung, welche aber bei KI-Algorithmen nur eingeschränkt kontrolliert werden können.¹⁶² In einem Positionspapier des Bundesverbandes Informationswirtschaft, Telekommunikation und neue Medien (Bitkom e.V.) und dem Deutschen Forschungszentrum für Künstliche Intelligenz wird die datenschutzkonforme Verwendung von KI-Produkten als umsetzbar und rechtliche Anpassungen der DS-GVO als

notwendig angesehen.¹⁶³ So darf eine Person nicht einer automatisiert getroffenen Entscheidung unterworfen werden, es sei denn, diese unterliegt einer menschlichen Kontrolle.¹⁶⁴ Dies war in der vorliegenden Arbeit zutreffend, da die Entscheidung des Algorithmus anhand der ärztlichen Befunde sowie mittels eigener visueller Beurteilung überprüft werden konnte. Insgesamt zeigt sich, dass die Anwendung von KI-Anwendungen zu medizinischen Zwecken komplexen rechtlichen Regularien unterliegt, die sich nur eingeschränkt auf solche Systeme auslegen lassen.

5.4 Ethische Aspekte

Die Anwendung von KI-Anwendungen stellt auch im ethischen Kontext eine Herausforderung dar. Da solche Systeme bereits jetzt ähnliche Leistungen wie medizinische Expert*innen zeigen bzw. sie teilweise sogar übertreffen¹⁶⁵, stellt sich nun schon die Frage, inwiefern es vertretbar ist, diese noch nicht flächendeckend einzusetzen. Dem gegenüber stehen die teils unklaren rechtlichen Rahmenbedingungen sowie die Frage nach der Akzeptanz auf Seite der Bevölkerung. Besonders problematisch ist die Tatsache, dass ein Algorithmus vor allem die Daten gut klassifizieren kann, mit denen er trainiert wurde. Sind also bestimmte Gruppen (z.B. nach Geschlecht oder Alter) im Lerndatensatz unterrepräsentiert, könnten sie durch häufigere Fehler bei den Klassifikationsergebnissen benachteiligt werden. Ein äußerst erschreckendes Beispiel aus dem Alltag für solche eine Fehlerkennung zeigte sich 2015 bei der App Google Photos. Dieses Programm erstellt mit Hilfe eines Algorithmus Begriffe (sogenannte „Tags“) z.B. für das Kategorisieren und Teilen von Fotos in sozialen Netzwerken. Dabei kam es dazu, dass Menschen mit dunklen Hauttypen nicht als Menschen, sondern als Gorillas klassifiziert wurden.^{166,167} Anscheinend wurde der Algorithmus vor allem mit Fotos von Menschen hellerer Hauttypen trainiert, sodass es einen Bias zu hellerer Haut gab.

Eine solche Diskriminierung von Personengruppen ist ethisch völlig inakzeptabel. In Bezug auf den medizinischen Kontext würden Fehlklassifikationen mit einer Gesundheitsgefährdung einhergehen, sodass hier die Bedeutung eines heterogenen und repräsentativen Lerndatensatzes deutlich wird.¹⁶⁷ Lerndatensätze müssten standardisiert sein und von unabhängiger Stelle z.B. im Rahmen eines Zertifizierungsprozesses überprüft werden. Ansonsten könnte es dazu kommen, dass z.B. wirtschaftliche Interessen zur Manipulation von Lerndaten führen. So könnte ein Algorithmus dahingehend trainiert werden, Bilddaten eher als pathologisch und damit mehr Patient*innen als erkrankt zu klassifizieren. Dies könnte eine erhöhte Nachfrage nach der entsprechenden Therapie bedingen und das Interesse zur Einflussnahme z.B. pharmazeutischer Unternehmen wecken. Allerdings gilt zu bedenken, dass zahlreiche Erkrankungen bestimmte Personengruppen häufiger betreffen als andere. Es gilt also einen Mittelweg zwischen balancierten und aussagekräftigen Daten und

angemessener Berücksichtigung von prädisponierenden Faktoren wie Alter oder sozialem Status zu finden. Auch stellt sich in diesem Kontext die Frage, inwiefern medizinische KI-Anwendungen zu kommerziellen Zwecken eingesetzt werden dürfen. So könnten z.B. Versicherungen ihre Aufnahmebedingungen an eine automatisierte Klassifikation im Sinne einer Risikostratifizierung knüpfen. So würden Personen, die eine Bereitstellung ihrer Gesundheitsdaten im Sinne der informationellen Selbstbestimmung ablehnen, gegebenenfalls benachteiligt.

Zudem ist zu klären, wie die Ergebnisse einer automatisierten Klassifikation kommuniziert würden. Insbesondere bei Patient*innen, die sich Transparenz und eine Erläuterung z.B. ihres Bildbefundes wünschen, wäre bei einer alleinigen Nennung der prozentualen Sicherheit der Klassifikation nur eine geringe Akzeptanz und eher Verunsicherung zu erwarten. Hier sind vor allem die menschliche Zuwendung und das Einordnen des Klassifikationsergebnisse in den Gesamtkontext entscheidende Faktoren. Gerade bei auffälligen Befunden stellt sich die Frage nach dem weiteren Behandlungsverlauf und therapeutischen Möglichkeiten. All diese Fragen kann ein Algorithmus, der meist nur auf eine spezifische Fragestellung ausgerichtet ist, nicht beantworten, von der fehlenden Fähigkeit zur Empathie ganz zu schweigen. Allerdings geht es auch nicht darum, menschliches Personal gänzlich zu ersetzen, sondern mit Anwendungen des tiefen Lernens unterstützende, effiziente und schnelle Hilfsmittel zur Verfügung zu stellen. Insbesondere in Anbetracht des Personal- und Zeitmangels in vielen medizinischen Einrichtungen liegt die Sinnhaftigkeit einer computergestützten Zweitmeinung nahe. Kontrolliert eingesetzte KI-Anwendungen könnten so zur schnelleren Diagnostik und Erhöhung der Patientensicherheit sowie zur Entlastung des Personals im Gesundheitssystem beitragen.

5.5 Limitationen

In Hinblick auf die Datengrundlage ist deren Monozentrität als limitierender Faktor zu sehen, weshalb eine Übertragbarkeit auf andere Institutionen bei unterschiedlichen klinikinternen Untersuchungsprotokollen und technischen Ausstattungen nur eingeschränkt möglich ist. Auch kann der Einfluss verschiedener Herangehensweisen sowie der allgemeinen klinikinternen Expertise in der diagnostischen Beurteilung der DAT-SPECTs beim monozentrischen Ansatz nur eingeschränkt eingeordnet werden. In dieser Hinsicht wäre daher das Einbeziehen multizentrischer Daten sinnvoll. In diesem Kontext ist auch der Faktor der Multinationalität von Interesse, unter anderem aufgrund der unterschiedlichen Zulassungen für Tracer bzw. technischen Unterschieden, z.B. durch unterschiedliche Gamma-Kamera-Modelle. Dank der weiten Verbreitung des DICOM-Formates in den PACS-Umgebungen (Picture Archiving and Communication System, [englisch]) auch auf internationaler Ebene ist die Erstellung eines länderübergreifenden Datensatzes aus technischer Sicht recht

unkompliziert.¹⁶⁸ In der Veröffentlichung von Choi und Ha et al. wurde dies bewerkstelligt, indem ein Datensatz des Seoul National University Hospitals sowie Daten der Parkinson's Progression Markers Initiative-Studie (PPMI-Studie, siehe unten) verwendet wurden. Auf interindividuelle Unterschiede hinsichtlich der Bilddatenakquirierung zwischen den verschiedenen Zentren wird allerdings nicht eingegangen, da der Fokus in der Publikation auf der Klassifikationsleistung liegt.¹³⁵ Auch Wenzel et al. wählten als Datengrundlage einen Datensatz aus der PPMI-Kohorte sowie klinikeigene Daten.¹³³ Im Gegensatz zu anderen bildgebenden Verfahren ist die Varianz zwischen verschiedenen Institutionen auf nationaler Ebene bei der Dopamin-Transporter-SPECT geringer, da für diese Modalität nur ein Tracer in Deutschland erhältlich ist.⁶⁶ Tracer-spezifische Unterschiede wie z.B. unterschiedliche Zeitprofile in der Nuklidbelegung sind daher bei der Verwendung in Deutschland akquirierter Datensätze auszuschließen.

Einschränkenden Einfluss hat ebenso die hier verwendete diagnostische Referenz in Form der ärztlichen semiquantitativen Bildbeurteilung. Die Beurteilung der DAT-SPECTs erfolgte durch mehrere Ärzt*innen, sodass individuell unterschiedliche Ansichtsweisen sowie Erfahrungsumfänge einen möglichen verzerrenden Einfluss auf die diagnostische Grundwahrheit der Daten darstellen. Allerdings gilt die Begutachtung durch Expert*innen nach wie vor als Goldstandard, sodass eine gewisse Subjektivität und inter- sowie intraindividuelle Fehleranfälligkeit der Diagnose zum aktuellen Zeitpunkt als modalitätsimmanent zu beurteilen sind.⁶⁶ Auch wäre die Beurteilung aller Bilddateien durch eine einzelne Person zur Reduktion der Interrater-Variabilität im klinischen Alltag aus Kapazitätsgründen nicht realistisch umsetzbar. Ein Lösungsansatz ergibt sich durch die Sicherung der Diagnose, die allerdings erst postmortal mittels histologischer Analyse möglich ist.¹⁶⁹ Dafür würde sich beispielsweise die Datenbank der Parkinson's Progression Markers Initiative-Studie (PPMI-Studie) eignen. Diese longitudinale Beobachtungsstudie startete 2010 und ist zudem multizentrisch und multinational. Sie dient der Identifizierung von Biomarkern beim Morbus Parkinson unter anderem mittels klinischer, bildgebender, laborchemischer und histologischer Verfahren.¹⁷⁰ Mittels Einspeisung von DAT-SPECT-Daten sowie histologischen Informationen ließe sich die die Grundwahrheit maximal objektivieren. Allerdings beinhaltet auch die Betrachtung der histologischen Schnitte eine (wenn auch geringere) subjektive Komponente. Choi und Ha et al. sowie Wenzel et al. setzten bereits erfolgreich die DAT-SPECT-Dateien des PPMI-Datensatzes zur Entwicklung ihres Algorithmus ein.^{133,135} Außerdem wäre das Hinzufügen von weiteren Bilddateien aus der PPMI-Datenbank im Sinne einer weiteren Augmentierung der Daten (siehe oben) denkbar. Allerdings liegt der Fokus der PPMI-Studie auf den pathologischen Befunden, was sich auch in der Zusammensetzung der Studienkohorte widerspiegelt. Bis Oktober 2023 schloss sie 1758 Personen ein (Stand 10.10.2023). Allerdings

handelte es sich nur bei 237 Personen um gesunde Kontrollen, was lediglich 13% aller Teilnehmenden entspricht.¹⁷¹ Diejenigen mit pathologischen Auffälligkeiten sind deutlich überrepräsentiert, was den Einsatz dieser Daten sehr beschränkt.

Zum anderen ist der Datenumfang eine verbesserungsfähige Komponente. Dabei ist zu beachten, dass die über 400 in dieser Arbeit verwendeten Datensätze aus einem über sechs Jahre umfassenden Zeitraum stammen. Dateianzahlen im fünfstelligen Bereich oder höher wie es z.B. in der Arbeit von Esteva et al. der Fall war, wären demnach nur mittels multizentrischen Ansatzes und/oder einem größeren Zeitfenster zu erzielen.¹⁵⁰ Außerdem stammten in dieser Arbeit die unauffälligen Befunde von Personen, die eine Indikation zur DAT-SPECT besaßen und nicht von gesunden Kontrollen. Da zahlreiche Erkrankungen wie z.B. der essentielle Tremor mit einer gänzlich unauffälligen DAT-SPECT einhergehen können, stellt sich allerdings die Frage nach dem Mehrwert einer gesunden Vergleichsgruppe. Außerdem wäre eine wissentliche Strahlenexposition von Gesunden mit potentiell unklarem Erkenntnisgewinn ethisch fragwürdig. Die PPMI-Datenbank eignet sich aufgrund ihres Fokus auf erkrankte Personen nur eingeschränkt zur Vergrößerung der Datenbasis. Der Anteil der auffälligen Befunde in der vorliegenden Studienkohorte betrug 57%, für die Testdatensätze wurde aber ein exaktes Verhältnis von 1:1 von physiologischen zu pathologischen Befunden gewählt. Dies erleichterte die Erkennung einer eventuellen Tendenz zur bevorzugten Erkennung auffälliger bzw. unauffälliger Befunde. Ein gewisses Überwiegen auffälliger Befunde wäre unter Umständen repräsentativer gewesen, birgt aber das Risiko der Verzerrung. Zudem wurden Bilddateien in gefilterter Rückprojektion, artefaktbehaftete, sowie jene in frontaler und sagittaler Ebene ausgeschlossen. Die Verwendung solcher Daten müsste in weiteren Arbeiten untersucht werden.

Bezüglich des methodischen Vorgehens gilt zu bedenken, dass zweidimensionale prozessierte Daten verwendet wurden. Dies entspricht zwar dem, was menschliche Beurteiler*innen ebenfalls als Basis ihrer Diagnostik benutzen. Allerdings ist es denkbar, dass in den dreidimensionalen Rohdaten zusätzliche (technische) Faktoren wie z.B. minimal unterschiedliche Graustufen oder räumliche Verteilungsdiskrepanzen vorliegen, die durch das menschliche Auge nicht erfasst werden können, durch eine Maschine aber eindeutig identifizierbar sind. Hinsichtlich der Klassifikationsleistung sind die recht starken Unterschiede zwischen den verschiedenen Kombinationen aus Testdatensätzen, Durchlaufrate und Netzwerkfraktion auffällig. Eine klare Tendenz in eine Richtung ist nicht zu erkennen, wengleich Faktoren wie eine höhere Durchlaufrate und Netzwerkauslastung einen positiven Einfluss zu haben scheinen. Die maximal erreichten Werte für Sensitivität, Spezifität und die positiven und negativen Prädiktionswerte lagen beim Optimum von 100%, bei der

Richtigklassifikationsrate wurden allerdings maximal 90% erreicht. Im Vergleich zu anderen erwähnten Veröffentlichungen wird hier der Optimierungsbedarf bis zur klinischen Einsetzbarkeit offensichtlich. Schließlich ist die methodenimmanente eingeschränkte Nachvollziehbarkeit der diagnostischen Kategorisierung durch den Algorithmus einschränkend. Es erfolgte zudem keine Korrelation der Bildmorphologie mit anderen z.B. klinischen Befunden, sodass das Einbeziehen weiterer Aspekte zur Untermauerung der Klassifizierung weiter untersucht werden sollte.

5.6 Zusammenfassung und Ausblick

Es konnte in der vorliegenden Arbeit gezeigt werden, dass die automatisierte Klassifikation von DAT-SPECT-Dateien nach entsprechender Datenpräparation durch eine Verwendung eines künstlichen neuronalen Netzwerks unter TensorFlow™ möglich ist. Es wurden überzeugende Ergebnisse mit Richtigklassifikationsraten von bis zu 90 % erreicht, wobei mit der ärztlichen Bildbeurteilung der aktuelle Goldstandard als Referenz diente. Dabei wurden bei höheren Netzwerkfraktionen und Anzahl an Trainingsdurchläufen tendenziell bessere Ergebnisse erzielt. Allerdings gab es ausgeprägte Unterschiede zwischen den Ergebnissen verschiedener Konstellationen, sodass die weitere Untersuchung der leistungsfähigeren Kombinationen aussichtsreicher erscheint.

Es erscheint sinnvoll, ärztliche und computerwissenschaftliche Kompetenzen zu kombinieren, da beide Gebiete in sich zu komplex sind, um sie gleichermaßen zu beherrschen. Für die bessere Nachvollziehbarkeit von Vorgängen des tiefen Lernens könnten im weiterführenden Kontext Methoden der „explainable AI“ [englisch] angewandt werden. Nichtsdestotrotz ist die Kontrolle der automatisierten Ergebnisse durch menschliche Expertise unerlässlich, da der Algorithmus seine Entscheidung rein bildmorphologisch ohne klinische Korrelation trifft und sie nicht auf Plausibilität prüft. In Anbetracht der raschen Entwicklung immer leistungsfähigerer Systeme erscheint ein flächendeckender Einsatz in mehreren Jahren sinnvoll und realistisch, was vor allem auf ethischer und juristischer Ebene Herausforderungen mit sich bringt. Die bislang recht allgemein gehaltenen Regularien und Normen werden den umfangreichen Möglichkeiten von selbstlernenden Anwendungen nur bedingt gerecht, sodass die Entwicklung angepasster Vorgaben für einen verantwortungsvollen Einsatz solcher Systeme notwendig erscheint. Von einer klinischen Einsatzfähigkeit des hier vorgestellten Algorithmus ist derzeit noch nicht auszugehen, wenngleich die Ergebnisse das immense Potential künstlicher neuronaler Netzwerke zur Bilddatenanalyse aufzeigen. Insbesondere bildgebende Modalitäten mit einer übersichtlichen Anzahl an diagnostischen Kategorien wie die DAT-SPECT könnten von solchen Entwicklungen zuerst profitieren.

Es ist zu erwarten, dass sich die Arbeitsabläufe in der Medizin durch den Einsatz von Anwendungen des tiefen Lernens verändern werden. Algorithmen könnten als objektive Zweitmeinung dienen und Gesundheitspersonal in seiner Arbeit unterstützen. Dabei kommt dem ärztlichen Personal vor allem in der Radiologie und Nuklearmedizin eine Schlüsselrolle in der Einordnung automatisiert erstellter Ergebnisse zu. Hintergrundwissen aus dem Bereich der Computerwissenschaften wird eine zunehmende Rolle spielen, während es bei Routineaufgaben zu kürzerem zeitlichem Aufwand kommen könnte. Derzeit sind aber noch viele Fragestellungen offen, ehe es zu Änderungen der bewährten Abläufe kommen kann, wenngleich ein positiver Einfluss der Methode automatisierter Klassifikationen auf das Gesundheitssystem zu erwarten ist.

6. Literaturverzeichnis

1. Deutsche Parkinson-Gesellschaft e.V. Hintergrundinformationen Parkinson-Krankheit. 2020. Berlin: Deutsche Gesellschaft für Parkinson und Bewegungsstörungen e.V. <https://www.parkinson-gesellschaft.de/die-dpg/morbus-parkinson.html> (Zuletzt abgerufen am 10.10.2023)
2. Riedel O, Bitters D, Amann U, Garbe E, Langner I. Estimating the prevalence of Parkinson's disease (PD) and proportions of patients with associated dementia and depression among the older adults based on secondary claims data. *Int J Geriatr Psychiatry* 2016; **31**(8): 938-43.
3. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord* 2014; **29**(13): 1583-90.
4. Dorsey ER, Elbaz A, Nichols E, et al. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* 2018; **17**(11): 939-53.
5. Gustavsson A, Svensson M, Jacobi F, et al. Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* 2011; **21**(10): 718-79.
6. Savica R, Grossardt BR, Bower JH, Ahlskog JE, Rocca WA. Time Trends in the Incidence of Parkinson Disease. *JAMA Neurology* 2016; **73**(8): 981-9.
7. Deutsche Gesellschaft für Neurologie. Leitlinien für Diagnostik und Therapie in der Neurologie. Idiopathisches Parkinson-Syndrom. 2016. Berlin: Deutsche Gesellschaft für Neurologie. <https://dgn.org/leitlinie/130> (Zuletzt abgerufen am 10.10.2023)
8. Max-Planck-Institut für Psychiatrie. Parkinson-Krankheit. 2020. München: Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. <https://www.psych.mpg.de/847793/parkinson> (Zuletzt abgerufen am 10.10.2023)
9. Vila M. Neuromelanin, aging, and neuronal vulnerability in Parkinson's disease. *Mov Disord* 2019; **34**(10): 1440-51.
10. 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; **60**(3): 337-41.
11. Lesage S, Brice A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* 2009; **18**(1): 48-59.
12. Lin X, Parisiadou L, Gu X-L, et al. Leucine-Rich Repeat Kinase 2 Regulates the Progression of Neuropathology Induced by Parkinson's-Disease-Related Mutant α -synuclein. *Neuron* 2009; **64**(6): 807-27.
13. Baylor College of Medicine. Atypical Parkinsonism. 2020. Houston (TX): Baylor College of Medicine. <https://www.bcm.edu/healthcare/specialties/neurology/parkinsons-and-movement-disorders/conditions/atypical-parkinsonism> (Zuletzt abgerufen am 10.10.2023)
14. Deuschländer AB, Ross OA, Dickson DW, Wszolek ZK. Atypical parkinsonian syndromes: a general neurologist's perspective. *Eur J Neurol* 2018; **25**(1): 41-58.
15. Boxer AL, Yu J-T, Golbe LI, Litvan I, Lang AE, Höglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. *The Lancet Neurology* 2017; **16**(7): 552-63.
16. Mahapatra RK, Edwards MJ, Schott JM, Bhatia KP. Corticobasal degeneration. *The Lancet Neurology* 2004; **3**(12): 736-43.
17. Ozawa T, Paviour D, Quinn NP, et al. The spectrum of pathological involvement of the striatonigral and olivopontocerebellar systems in multiple system atrophy: clinicopathological correlations. *Brain* 2004; **127**(12): 2657-71.
18. Kim WS, Kågedal K, Halliday GM. Alpha-synuclein biology in Lewy body diseases. *Alzheimers Res Ther* 2014; **6**(5): 73-.
19. Gilbert R. Understanding Atypical Parkinsonism. 2018. New York City: American Parkinson Disease Association. <https://www.apdaparkinson.org/article/atypical-parkinsonism/> (Zuletzt abgerufen am 10.10.2023)

20. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967; **17**(5): 427-42.
21. Takahashi H, Wakabayashi K. The cellular pathology of Parkinson's disease. *Neuropathology* 2001; **21**(4): 315-22.
22. Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R, Goedert M. α -Synuclein in Lewy bodies. *Nature* 1997; **388**(6645): 839-40.
23. Sandmann-Keil D, Braak H. Zur postmortalen Diagnose des idiopathischen Morbus Parkinson. *Der Pathologe* 2005; **26**(3): 214-20.
24. Gao X, Carroni M, Nussbaum-Krammer C, et al. Human Hsp70 Disaggregase Reverses Parkinson's-Linked α -Synuclein Amyloid Fibrils. *Mol Cell* 2015; **59**(5): 781-93.
25. Braak H, Del Tredici K. Invited Article: Nervous system pathology in sporadic Parkinson disease. *Neurology* 2008; **70**(20): 1916-25.
26. Braak H, Del Tredici K, Bratzke H, Hamm-Clement J, Sandmann-Keil D, Rüb U. Staging of the intracerebral inclusion body pathology associated with idiopathic Parkinson's disease (preclinical and clinical stages). *J Neurol* 2002; **249** [Suppl 3]: : III/1-III/5.
27. Braak H, Tredici KD, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* 2003; **24**(2): 197-211.
28. Braak H, Rüb U, Gai WP, Del Tredici K. Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *J Neural Transm* 2003; **110**(5): 517-36.
29. Del Tredici K, Rüb U, de Vos RAI, Bohl JRE, Braak H. Where Does Parkinson Disease Pathology Begin in the Brain? *J Neuropathol Exp Neurol* 2002; **61**(5): 413-26.
30. Hawkes CH, Del Tredici K, Braak H. Parkinson's disease: a dual-hit hypothesis. *Neuropathol Appl Neurobiol* 2007; **33**(6): 599-614.
31. Klingelhofer L, Reichmann H. Pathogenesis of Parkinson disease—the gut–brain axis and environmental factors. *Nature Reviews Neurology* 2015; **11**(11): 625-36.
32. Mello LEAM, Villares J. Neuroanatomy of the basal ganglia. *Psychiatr Clin North Am* 1997; **20**(4): 691-704.
33. Haber S, Gdowski M. The Basal Ganglia. In: Paxinos G, Mai J, eds. *The Human Nervous System*. 3rd ed. San Diego: Academic Press/El Sevier; 2012: 678-82.
34. Báez-Mendoza R, Schultz W. The role of the striatum in social behavior. *Front Neurosci* 2013; **7**(233): 1-14.
35. Knierim J. Chapter 4: Basal Ganglia. 2020. Houston (TX): Department of Neurobiology and Anatomy, McGovern Medical School at UHealth, The University of Texas. <https://nba.uth.tmc.edu/neuroscience/m/s3/chapter04.html> (Zuletzt abgerufen am 10.10.2023)
36. Lanciego JL, Luquin N, Obeso JA. Functional neuroanatomy of the basal ganglia. *Cold Spring Harb Perspect Med* 2012; **2**(12): 1-20.
37. Nagano-Saito A, Martinu K, Monchi O. Function of basal ganglia in bridging cognitive and motor modules to perform an action. *Front Neurosci* 2014; **8**(187): 1-12.
38. Courtemanche R, Cammalleri A. Basal Ganglia: Striosomes and the Link between Motivation and Action. *Curr Biol* 2019; **29**(2): 62-5.
39. Probst CC, van Eimeren T. The Functional Anatomy of Impulse Control Disorders. *Curr Neurol Neurosci Rep* 2013; **13**(10): 386.
40. Redgrave P, Rodriguez M, Smith Y, et al. Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nature Reviews Neuroscience* 2010; **11**(11): 760-72.
41. Gremel CM, Lovinger DM. Associative and sensorimotor cortico-basal ganglia circuit roles in effects of abused drugs. *Genes Brain Behav* 2017; **16**(1): 71-85.
42. Dittmann A. Rigor. 2016. Berlin: Pschyrembel online. Walter de Gruyter GmbH. <https://www.pschyrembel.de/Rigor/K0JXP> (Zuletzt abgerufen am 10.10.2023)
43. Askenasy JJ, Yahr MD. Parkinsonian tremor loses its alternating aspect during non-REM sleep and is inhibited by REM sleep. *Journal of neurology, neurosurgery, and psychiatry* 1990; **53**(9): 749-53.
44. Lamberti P, Armenise S, Castaldo V, et al. Freezing Gait in Parkinson's Disease. *Eur Neurol* 1997; **38**(4): 297-301.

45. Lücke C, Pape F. Parkinson-Syndrom. 2020. Berlin: Pschyrembel online. Walter de Gruyter GmbH. <https://www.pschyrembel.de/Parkinson-Syndrom/K0GCW> (Zuletzt abgerufen am 10.10.2023)
46. Matsumoto JY. Parkinsonism: Causes and coping strategies. 2019. Rochester, Minnesota: Mayo Foundation for Medical Education and Research (MFMER). <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/expert-answers/parkinsonism/faq-20058490> (Zuletzt abgerufen am 10.10.2023)
47. Chaudhuri KR, Healy DG, Schapira AH. Non-motor symptoms of Parkinson's disease: diagnosis and management. *Lancet Neurol* 2006; **5**(3): 235-45.
48. Poewe W. Non-motor symptoms in Parkinson's disease. *Eur J Neurol* 2008; **15**(s1): 14-20.
49. D'Costa DF, Sheehan LJ, Phillips PA, Moore-Smith B. The levodopa test in Parkinson's disease. *Age Ageing* 1995; **24**(3): 210-2.
50. Mahlkecht P, Hotter A, Hussl A, Esterhammer R, Schocke M, Seppi K. Significance of MRI in diagnosis and differential diagnosis of Parkinson's disease. *Neurodegener Dis* 2010; **7**(5): 300-18.
51. Pagano G, Niccolini F, Politis M. Imaging in Parkinson's disease. *Clin Med* 2016; **16**(4): 371-5.
52. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov Disord* 2008; **23**(15): 2129-70.
53. Brooks DJ. Dopamine agonists: their role in the treatment of Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 2000; **68**(6): 685-9.
54. Blandini F, Armentero M-T. Dopamine receptor agonists for Parkinson's disease. *Expert Opinion on Investigational Drugs* 2014; **23**(3): 387-410.
55. Pahwa R, Lyons KE. Levodopa-related wearing-off in Parkinson's disease: identification and management. *Curr Med Res Opin* 2009; **25**(4): 841-9.
56. Foley P, Gerlach M, Youdim MBH, Riederer P. MAO-B inhibitors: multiple roles in the therapy of neurodegenerative disorders? *Parkinsonism Relat Disord* 2000; **6**(1): 25-47.
57. Honig H, Antonini A, Martinez-Martin P, et al. Intrajejunal levodopa infusion in Parkinson's disease: A pilot multicenter study of effects on nonmotor symptoms and quality of life. *Mov Disord* 2009; **24**(10): 1468-74.
58. Fasano A, Ricciardi L, Lena F, Bentivoglio AR, Modugno N. Intrajejunal levodopa infusion in advanced Parkinson's disease: long-term effects on motor and non-motor symptoms and impact on patient's and caregiver's quality of life. *Eur Rev Med Pharmacol Sci* 2012; **16**(1): 79-89.
59. Lang AE, Siderowf AD, Macklin EA, et al. Trial of Cinpanemab in Early Parkinson's Disease. *N Engl J Med* 2022; **387**(5): 408-20.
60. Pagano G, Taylor KI, Anzures-Cabrera J, et al. Trial of Prasinezumab in Early-Stage Parkinson's Disease. *N Engl J Med* 2022; **387**(5): 421-32.
61. de Lau LML, Schipper CMA, Hofman A, Koudstaal PJ, Breteler MMB. Prognosis of Parkinson Disease: Risk of Dementia and Mortality: The Rotterdam Study. *Arch Neurol* 2005; **62**(8): 1265-9.
62. Sato K, Hatano T, Yamashiro K, et al. Prognosis of Parkinson's disease: Time to stage III, IV, V, and to motor fluctuations. *Mov Disord* 2006; **21**(9): 1384-95.
63. Anger HO. Scintillation Camera. *Rev Sci Instrum* 1958; **29**(1): 27-33.
64. Hutton BF. The origins of SPECT and SPECT/CT. *Eur J Nucl Med Mol Imaging* 2014; **41**(1): 3-16.
65. Graham RNJ, Perriss RW, Scarsbrook AF. DICOM demystified: A review of digital file formats and their use in radiological practice. *Clin Radiol* 2005; **60**(11): 1133-40.
66. Tatsch K, Buchert R, Bartenstein P, et al. DGN-Handlungsempfehlung (S1-Leitlinie). SPECT-Untersuchungen mit dem 123I-markierten Dopamintransporter-Liganden FP-CIT (DaTSCANTM). 2017. Göttingen: Deutsche Gesellschaft für Nuklearmedizin. <https://www.nuklearmedizin.de/leistungen/leitlinien/docs/031-037.pdf> (Zuletzt abgerufen am 10.10.2023)

67. Buchert R, Buhmann C, Apostolova I, Meyer PT, Gallinat J. Nuklearmedizinische Diagnostik bei Parkinson-Syndromen. *Dtsch Arztebl International* 2019; **116**(44): 747-54.
68. Pirker W. Update: SPECT in der Diagnostik von Parkinson-Syndromen. *Journal für Neurologie, Neurochirurgie und Psychiatrie* 2015; **16**(2): 60-71.
69. Piccini P, Whone A. Functional brain imaging in the differential diagnosis of Parkinson's disease. *The Lancet Neurology* 2004; **3**(5): 284-90.
70. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015; **30**(12): 1591-601.
71. Marek K, Seibyl J, Eberly S, et al. Longitudinal follow-up of SWEDD subjects in the PRECEPT Study. *Neurology* 2014; **82**(20): 1791-7.
72. Ibrahim N, Kusmirek J, Struck AF, et al. The sensitivity and specificity of F-DOPA PET in a movement disorder clinic. *Am J Nucl Med Mol Imaging* 2016; **6**(1): 102-9.
73. Pavese N, Brooks DJ. Imaging neurodegeneration in Parkinson's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 2009; **1792**(7): 722-9.
74. Turjanski N, Bhatia K, Burn DJ, Sawle GV, Marsden CD, Brooks DJ. Comparison of striatal 18F-dopa uptake in adult-onset dystonia-parkinsonism, Parkinson's disease, and dopa-responsive dystonia. *Neurology* 1993; **43**(8): 1563-.
75. Rinne JO. Positron emission tomography in the differential diagnosis of parkinsonism. *Journal of movement disorders* 2009; **2**(2): 53-7.
76. Darcourt J, Schiavza A, Sapin N, et al. 18F-FDOPA PET for the diagnosis of parkinsonian syndromes. *Q J Nucl Med Mol Imaging* 2014; **58**(4): 355-65.
77. Egerton A, Demjaha A, McGuire P, Mehta MA, Howes OD. The test-retest reliability of 18F-DOPA PET in assessing striatal and extrastriatal presynaptic dopaminergic function. *Neuroimage* 2010; **50**(2): 524-31.
78. Nanni C, Fanti S, Rubello D. 18F-DOPA PET and PET/CT. *J Nucl Med* 2007; **48**(10): 1577-9.
79. Tatsch K, Bartenstein P, Brust P, et al. Leitlinie für SPECT-Untersuchungen mit dem 123I -markierten Dopamin-D2-Rezeptor Liganden IBZM. o.J. Göttingen: Deutsche Gesellschaft für Nuklearmedizin e.V. https://www.nuklearmedizin.de/leistungen/leitlinien/html/spect_rez.php?navId=53 (Zuletzt abgerufen am 10.10.2023)
80. Bernardo M, Parellada E, Lomeña F, et al. Double-blind olanzapine vs. haloperidol D2 dopamine receptor blockade in schizophrenic patients: a baseline-endpoint. *Psychiatry Res* 2001; **107**(2): 87-97.
81. Briani C, Cagnin A, Chierichetti F, Tiberio M, Battistin L, Pizzolato G. Thiethylperazine-induced parkinsonism: in vivo demonstration of dopamine D2 receptors blockade. *Eur J Neurol* 2004; **11**(10): 709-10.
82. Kluge G, Sahlmann C, Schmidt M, Franzius C, Meller J. DGN-Handlungsempfehlung (S1-Leitlinie). Tumor-Szintigraphie mit 123Iod- (131I-) meta-Iodbenzylguanidin (mIBG). 2015. Göttingen: Deutsche Gesellschaft für Nuklearmedizin e.V. https://www.nuklearmedizin.de/leistungen/leitlinien/docs/031-050I_S1_Tumor_Szintigraphie_123Iod_meta_Iodbenzylguanidin_2015-06.pdf (Zuletzt abgerufen am 10.10.2023)
83. Yang T, Wang L, Li Y, et al. (131I)-MIBG myocardial scintigraphy for differentiation of Parkinson's disease from multiple system atrophy or essential tremor in Chinese population. *J Neurol Sci* 2017; **373**: 48-51.
84. Skowronek C, Zange L, Lipp A. Cardiac 123I-MIBG Scintigraphy in Neurodegenerative Parkinson Syndromes: Performance and Pitfalls in Clinical Practice. *Front Neurol* 2019; **10**: 152.
85. Jost W, Landvogt C. MIBG-Szintigrafie: wertvoll in der Differenzialdiagnostik. *Der Neurologe & Psychiater* 2010; **11**(8): 40-2.
86. Pirker W, Wenning G, Scherfler C, et al. Leitlinie Diagnostik des Morbus Parkinson. 2014. Wien: Österreichische Parkinson Gesellschaft. https://www.parkinson.at/files/CONTENT/OEPG/publikationen/P-aktuell_3_14.pdf (Zuletzt abgerufen am 10.10.2023)

87. Mehnert S, Reuter I, Schepp K, Maaser P, Stolz E, Kaps M. Transcranial sonography for diagnosis of Parkinson's disease. *BMC Neurol* 2010; **10**(1): 9.
88. Smajlovic D, Ibrahimagic OC. Transcranial Brain Sonography in Parkinson's Disease and Other Parkinsonian Disorders: a Hospital Study from Tuzla, Bosnia and Herzegovina. *Medical archives (Sarajevo, Bosnia and Herzegovina)* 2017; **71**(4): 261-4.
89. Sommer U, Hummel T, Cormann K, et al. Detection of presymptomatic Parkinson's disease: Combining smell tests, transcranial sonography, and SPECT. *Mov Disord* 2004; **19**(10): 1196-202.
90. Berg D, Godau J, Walter U. Transcranial sonography in movement disorders. *The Lancet Neurology* 2008; **7**(11): 1044-55.
91. Kok JN, Boers EJW, Kusters WA, Van der Putten P, Poel M. ARTIFICIAL INTELLIGENCE: DEFINITION, TRENDS, TECHNIQUES, AND CASES. 2009. Paris: Encyclopedia of Life Support Systems (EOLSS) UNESCO-EOLSS Joint Committee. <https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf> (Zuletzt abgerufen am 10.10.2023)
92. o.V. Artificial intelligence. 2023. Springfield (MA): Merriam-Webster.com Dictionary. <https://www.merriam-webster.com/dictionary/artificial%20intelligence> (Zuletzt abgerufen am 10.10.2023)
93. Copeland BJ. Artificial intelligence. 2023. Chicago (IL): Encyclopædia Britannica. <https://www.britannica.com/technology/artificial-intelligence#ref219078> (Zuletzt abgerufen am 10.10.2023)
94. o.V. artificial intelligence, n. 2023. Oxford: Oxford University Press. <https://www.oed.com/viewdictionaryentry/Entry/271625> (Zuletzt abgerufen am 10.10.2023)
95. Raschka S, Mirjalili V. Machine Learning mit Python und Scikit-Learn und TensorFlow: Das umfassende Praxis-Handbuch für Data Science, Deep Learning und Predictive Analytics. 2nd ed. Frechen: mitp Verlags GmbH & Co. KG; 2018.
96. Chollet F. Deep Learning with Python. 1st ed. Shelter Island, New York: Manning Publications Co.; 2018.
97. Degens PO, Federkiel H. Clusteranalyse in der Medizin. In: Ehlers CT, Klar R, eds. Informationsverarbeitung in der Medizin Medizinische Informatik und Statistik. Berlin, Heidelberg: Springer Verlag; 1979: 459-60.
98. Chouard T, Venema L. Machine intelligence. *Nature* 2015; **521**(7553): 435-.
99. Jones N. Computer science: The learning machines. *Nature* 2014; **505**(7482): 146-8.
100. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**(7553): 436-44.
101. Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol* 2008; **26**(9): 1011-3.
102. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci Biobehav Rev* 2017; **74**(Part A): 58-75.
103. Deru M, Ndiaye A. Deep Learning mit TensorFlow, Keras und TensorFlow.js. 1 ed. Bonn: Rheinwerk Computing Verlag; 2019.
104. Jin C, Jang S, Sun X, Li J, Christenson R. Damage detection of a highway bridge under severe temperature changes using extended Kalman filter trained neural network. *Journal of Civil Structural Health Monitoring* 2016; **6**(3): 545-60.
105. o.V. TensorFlow For Poets. 2016. Mountain View (CA): Google LLC. <https://kiosk-dot-codelabs-site.appspot.com/codelabs/tensorflow-for-poets/#0> (Zuletzt abgerufen am 29.06.2023)
106. Goodfellow I, Bengio Y, Courville A. Deep Learning. 1st ed. Cambridge, MA: MIT Press; 2016.
107. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 2009; **22**(10): 1345-59.
108. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *Journal of Big Data* 2016; **3**(1): 9.
109. Sassman L. Add structured data to your web pages. 2022. Mountain View (CA): Google LLC. <https://codelabs.developers.google.com/codelabs/structured-data#0> (Zuletzt abgerufen am 10.10.2023)

110. Zhao K, Matsukawa T, Suzuki E. Retraining: A Simple Way to Improve the Ensemble Accuracy of Deep Neural Networks for Image Classification. 2018 24th International Conference on Pattern Recognition (ICPR); 2018 20-24 Aug. 2018; 2018. p. 860-7.
111. Microsoft Corporation. The Microsoft Cognitive Toolkit. 2022. Redmond (WA): Microsoft Corporation. <https://docs.microsoft.com/en-us/cognitive-toolkit/> (Zuletzt abgerufen am 10.10.2023)
112. Stott L. Deep Learning with Microsoft Cognitive Toolkit CNTK. 2019. Redmond (WA): Microsoft Corporation. <https://techcommunity.microsoft.com/t5/educator-developer-blog/deep-learning-with-microsoft-cognitive-toolkit-cntk/ba-p/378280> (Zuletzt abgerufen am 10.10.2023)
113. Dean J, Corrado G, Monga R, et al. Large scale distributed deep networks. *Adv Neural Inf Process Syst*; 2012; Lake Tahoe, Nevada: Curran Associates Inc.; 2012. p. 1223-31.
114. Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI '16); 2016; Savannah, Georgia: USENIX: The Advanced Computing Systems Association; 2016. p. 265-83.
115. Wojna Z, Gorban AN, Lee D-S, et al. Attention-based extraction of structured information from street view imagery. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR); 2017; Kyoto: IEEE; 2017. p. 844-50.
116. Choi H, Jin KH. Fast and robust segmentation of the striatum using deep convolutional neural networks. *J Neurosci Methods* 2016; **274**: 146-53.
117. Shao W, Rowe SP, Du Y. SPECTnet: a deep learning neural network for SPECT image reconstruction. *Ann Transl Med* 2021; **9**(9): 819.
118. Cheng J-Z, Ni D, Chou Y-H, et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep* 2016; **6**(1): 24454.
119. Shaffie A, Soliman A, Fraiwan L, et al. A Generalized Deep Learning-Based Diagnostic System for Early Diagnosis of Various Types of Pulmonary Nodules. *Technol Cancer Res Treat* 2018; **17**: 1533033818798800.
120. Nam JG, Park S, Hwang EJ, et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology* 2019; **290**(1): 218-28.
121. Liu B, Chi W, Li X, et al. Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect. *J Cancer Res Clin Oncol* 2020; **146**(1): 153-85.
122. Cui S, Ming S, Lin Y, et al. Development and clinical application of deep learning model for lung nodules screening on CT images. *Sci Rep* 2020; **10**(1): 13657.
123. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018; **29**(8): 1836-42.
124. Mori Y, Kudo S-E, Berzin TM, Misawa M, Takeda K. Computer-aided diagnosis for colonoscopy. *Endoscopy* 2017; **49**(8): 813-9.
125. Urban G, Tripathi P, Alkayali T, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* 2018; **155**(4): 1069-78.
126. Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nature Biomedical Engineering* 2018; **2**(10): 741-8.
127. Ha R, Chang P, Karcich J, et al. Convolutional Neural Network Based Breast Cancer Risk Stratification Using a Mammographic Dataset. *Acad Radiol* 2019; **26**(4): 544-9.
128. Yates EJ, Yates LC, Harvey H. Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. *Clin Radiol* 2018; **73**(9): 827-31.
129. Coy H, Hsieh K, Wu W, et al. Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT. *Abdominal Radiology* 2019; **44**(6): 2009-20.

130. Choi H, Kim YK, Yoon EJ, Lee J-Y, Lee DS, for the Alzheimer's Disease Neuroimaging I. Cognitive signature of brain FDG PET based on deep learning: domain transfer from Alzheimer's disease to Parkinson's disease. *Eur J Nucl Med Mol Imaging* 2020; **47**(2): 403-12.
131. Qiu S, Miller MI, Joshi PS, et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nature Communications* 2022; **13**(1): 3404.
132. Kim DH, Wit H, Thurston M. Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning. *Nucl Med Commun* 2018; **39**(10): 887-93.
133. Wenzel M, Milletari F, Krüger J, et al. Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics. *Eur J Nucl Med Mol Imaging* 2019; **46**(13): 2800-11.
134. Zhang YC, Kagen AC. Machine Learning Interface for Medical Image Analysis. *J Digit Imaging* 2017; **30**(5): 615-21.
135. Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical* 2017; **16**: 586-94.
136. Magesh PR, Myloth RD, Tom RJ. An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery. *Comput Biol Med* 2020; **126**: 104041.
137. Leung K, Shao W, Solnes L, Rowe S, Pomper M, Du Y. A deep learning-based approach for disease detection in the projection space of DAT-SPECT images of patients with Parkinson's disease. *J Nucl Med* 2020; (61): 509.
138. Zhao Y, Wu P, Wu J, et al. Decoding the dopamine transporter imaging for the differential diagnosis of parkinsonism using deep learning. *Eur J Nucl Med Mol Imaging* 2022; **49**(8): 2798-811.
139. Loh HW, Hong W, Ooi CP, et al. Application of Deep Learning Models for Automated Identification of Parkinson's Disease: A Review (2011-2021). *Sensors (Basel)* 2021; **21**(21).
140. Howard A, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv [Preprint]* 2017; **arXiv:1704.04861**.
141. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision* 2015; **115**(3): 211-52.
142. Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:171105225* 2017.
143. Leung KH, Rowe SP, Pomper MG, Du Y. A three-stage, deep learning, ensemble approach for prognosis in patients with Parkinson's disease. *EJNMMI Res* 2021; **11**(1): 52.
144. O'Brien JT, Oertel WH, McKeith IG, et al. Is ioflupane I123 injection diagnostically effective in patients with movement disorders and dementia? Pooled analysis of four clinical trials. *BMJ Open* 2014; **4**(7): e005122.
145. de la Fuente-Fernández R. Role of DaTSCAN and clinical diagnosis in Parkinson disease. *Neurology* 2012; **78**(10): 696.
146. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage: Clinical* 2017; **15**: 633-43.
147. Hussain Z, Gimenez F, Yi D, Rubin D. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. *AMIA Annu Symp Proc* 2017; **2017**: 979-84.
148. Alexandre-Santos L, Trevisan AC, Pitella FA, et al. Assessment of different regions of interest-based methods for [99mTc]Tc DAT-SPECT quantification using an anthropomorphic striatal phantom. *EJNMMI Physics* 2022; **9**(1): 91.
149. Ortiz A, Munilla J, Martínez-Ibañez M, Górriz JM, Ramírez J, Salas-Gonzalez D. Parkinson's Disease Detection Using Isosurfaces-Based Features and Convolutional Neural Networks. *Front Neuroinform* 2019; **13**: 48.
150. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**(7639): 115-8.
151. Song Q, Zhao L, Luo X, Dou X. Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. *J Healthc Eng* 2017; **2017**: 8314740.
152. Nazari M, Kluge A, Apostolova I, et al. Explainable AI to improve acceptance of convolutional neural networks for automatic classification of dopamine transporter SPECT in

the diagnosis of clinically uncertain parkinsonian syndromes. *Eur J Nucl Med Mol Imaging* 2022; **49**(4): 1176-86.

153. Samek W, Wiegand T, Müller K-R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:170808296* 2017.

154. Helle K. Intelligente Medizinprodukte: Ist der geltende Rechtsrahmen noch aktuell? *Medizinrecht* 2020; **38**(12): 993-1000.

155. MDCG 2019-11

Guidance on Qualification and Classification

of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. Medical Device Coordination Group (MDCG); 2019. p. 6 f.

156. o.V. EU-Medizinprodukteverordnung (MDR). Kapitel II - Bereitstellung auf dem Markt und Inbetriebnahme von Produkten, Pflichten der Wirtschaftsakteure, Aufbereitung, CE-Kennzeichnung, freier Verkehr. Artikel 10 - Allgemeine Pflichten der Hersteller. 2017. Brüssel: Europäisches Parlament, Rat der Europäischen Union. https://lexparency.de/eu/MDR/ART_10/ (Zuletzt abgerufen am 10.10.2023)

157. o.V. EU-Medizinprodukteverordnung (MDR). Anhang I - Grundlegende Sicherheits- und Leistungsanforderungen. Kapitel II Anforderungen an Auslegung und Herstellung. Abschn. 17.1. 2017. Brüssel: Europäisches Parlament, Rat der Europäischen Union. https://lexparency.de/eu/MDR/ANX_I/ (Zuletzt abgerufen am 10.10.2023)

158. Bundesministerium der Justiz. Gesetz über die Haftung für fehlerhafte Produkte (Produkthaftungsgesetz - ProdHaftG)

§ 1 Haftung. 2023. Berlin: Bundesministerium der Justiz. https://www.gesetze-im-internet.de/prodhaftg/_1.html (Zuletzt abgerufen am 10.10.2023)

159. Bundesministerium der Justiz. Bürgerliches Gesetzbuch (BGB) §630e Aufklärungspflichten. 2023. Berlin: Bundesministerium der Justiz. https://www.gesetze-im-internet.de/bgb/_630e.html (Zuletzt abgerufen am 10.10.2023)

160. Bundesministerium der Justiz. Gesetz zur Durchführung unionsrechtlicher Vorschriften betreffend Medizinprodukte (Medizinprodukte-Durchführungsgesetz - MPDG)

§ 12 Verbote zum Schutz von Patienten, Anwendern und Dritten. 2023. Berlin: Bundesministerium der Justiz. https://www.gesetze-im-internet.de/mpdg/_12.html (Zuletzt abgerufen am 10.10.2023)

161. o.V. Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016

zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) Art. 2 Sachlicher Anwendungsbereich Abs. 1. 2016. Brüssel: Europäisches Parlament, Rat der Europäischen Union. <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:02016R0679-20160504> (Zuletzt abgerufen am 10.10.2023)

162. o.V. Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016

zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) Art. 5 Grundsätze für die Verarbeitung personenbezogener Daten. 2016. Brüssel: Europäisches Parlament, Rat der Europäischen Union. <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:02016R0679-20160504> (Zuletzt abgerufen am 10.10.2023)

163. Bundesverband Informationswirtschaft Telekommunikation und neue Medien e.V. (Bitkom e.V.), Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI). Künstliche Intelligenz - Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. 2017. Berlin, Kaiserslautern: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., Deutsches Forschungszentrum für Künstliche Intelligenz GmbH. https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf (Zuletzt abgerufen am 10.10.2023)

164. o.V. Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling. 2016. Brüssel: Europäisches Parlament, Rat der Europäischen Union. <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:02016R0679-20160504> (Zuletzt abgerufen am 10.10.2023)
165. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health* 2019; **1**(6): e271-e97.
166. Kasperkevic J. Google says sorry for racist auto-tag in photo app. 2015. London: The Guardian. <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app> (Zuletzt abgerufen am 10.10.2023)
167. Jannes M, Friele M, Jannes C, Woopen C. Algorithmen in der digitalen Gesundheitsversorgung. Eine interdisziplinäre Analyse. 2019. Gütersloh: Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_Studie_Algorithmen.pdf (Zuletzt abgerufen am 10.10.2023)
168. Fridell K, Edgren L, Lindsköld L, Aspelin P, Lundberg N. The impact of PACS on radiologists' work practice. *J Digit Imaging* 2007; **20**(4): 411-21.
169. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. *The Lancet Neurology* 2006; **5**(1): 75-86.
170. Marek K, Jennings D, Lasch S, et al. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol* 2011; **95**(4): 629-35.
171. Parkinson's Progression Markers Initiative. Data at a Glance. 2023. o.A.: Parkinson's Progression Markers Initiative (PPMI). <https://www.ppmi-info.org/access-data-specimens/data> (Zuletzt abgerufen am 10.10.2023)

7. Anhang

7.1 Abbildungsverzeichnis

Abbildung 1: Darstellung des Striatums	16
Abbildung 2: Aufbau einer Gamma-Kamera	21
Abbildung 3: Iterative Rekonstruktion und gefilterte Rückprojektion im Vergleich	22
Abbildung 4: Unauffällige und auffällige DAT-SPECT im Vergleich.....	24
Abbildung 5: Varianten des maschinellen Lernens	30
Abbildung 6: Tiefes Lernen im Gefüge der künstlichen Intelligenz nach Chollet (2018).....	31
Abbildung 7: Aufbau eines Knotens nach Deru und Ndiaye (2019).....	33
Abbildung 8: Aufbau eines tiefen neuronalen Netzwerkes.....	34
Abbildung 9: Methodisches Vorgehen	44
Abbildung 10: Verarbeitung der Rohdaten.....	45
Abbildung 11: Exemplarische Darstellung ausgeschlossener Bilddateien.....	47
Abbildung 12: Exemplarische Darstellung eingeschlossener Bilddateien.....	47
Abbildung 13: Exemplarische Darstellung der Erfassung geeigneter Dateien und klinischer Informationen	48
Abbildung 14: Initiale Datenstruktur	49
Abbildung 15: Verwendetes Template in axialer Ebene	50
Abbildung 16: Vorbereitung der Bilddateien.....	52
Abbildung 17: Datenstruktur der Trainings- und Testdatensätze.....	53
Abbildung 18: Exemplarische Darstellung der dateiweisen Klassifikation	55
Abbildung 19: Zusammenfassung der Trainings- und Testprozesse.....	56
Abbildung 20: Altersabhängige Verteilung nach Geschlecht und dopaminergem Defizit	58
Abbildung 21: Exemplarische Erstellung der Maximumintensitätsprojektionen (MIP)	59
Abbildung 22: Datenaugmentierung.....	60

7.2 Tabellenverzeichnis

Tabelle 1: Verwendete Hardware.....	43
Tabelle 2: Verwendete Software	43
Tabelle 3: Vier-Felder-Tafel	57
Tabelle 4: Deskriptive Statistik der klinischen Parameter	58
Tabelle 5: Dateianzahlen der Ausgangsdatensätze.....	62
Tabelle 6: Dateianzahlen der Trainingsdatensätze.....	63
Tabelle 7: Genauigkeit (Richtigklassifikationsrate) der Klassifikationsergebnisse.....	65
Tabelle 8: Sensitivität der Klassifikationsergebnisse	66
Tabelle 9: Spezifität der Klassifikationsergebnisse	67

Tabelle 10: Positive Prädiktionswerte der Klassifikationsergebnisse.....	68
Tabelle 11: Negative Prädiktionswerte der Klassifikationsergebnisse	69

7.3 Skripts

Skript 1: CreatePlaneSums.m

```

1. CreatePlaneSums.m
2.
3.
4.
5.
6. pathImages = '/DATA/hammesj/DatscanAI/Rohdaten/Datscan_Meier_allFiles/';
7. %pathOutputImagesSum = '/DATA/hammesj/DatscanAI/outputImages_MIP_8bit_2012-2015/';
8. pathOutputImagesMIP = '/DATA/hammesj/DatscanAI/outputImages_MIP_8bit_allFiles/';
9.
10.
11.
12.
13. %Find all Files with in Input Folder
14. inputFiles = dir(pathImages);
15. inputFiles(1:2) = [];
16.
17.
18.
19.
20. testInfo = dicominfo([pathImages inputFiles(1).name]);
21. testDicom = dicomread([pathImages inputFiles(1).name]);
22.
23.
24.
25.
26.
27.
28.
29.
30. for i = 1:length(inputFiles)
31.     inputFilepath = [pathImages inputFiles(i).name];
32.
33.     tempDicom = dicomread(inputFilepath);
34.     tempDicomInfo = dicominfo(inputFilepath);
35.
36.     tempDicomInfo.PatientName;
37.
38.     temp3D = tempDicom(:,:,1,:);
39.     volumes = squeeze(temp3D);
40.
41.     %planeSum = sum(volumes{i},3);
42.     %planeSumplaneSumReduced = planeSum / max(planeSum(:)) * 256;
43.     %imwrite(planeSumReduced, [pathOutputImagesSum 'planeSum_' num2str(i) '.gif']);
44.
45.     planeMIP = max(volumes{i}, [], 3);
46.     planeMIP = double(max(volumes{i}, [], 3));
47.     planeMIPplaneMIPReduced = planeMIP / max(planeMIP(:)) * 256;
48.
49.     outputFilepath = [pathOutputImagesMIP 'planeMIP_' num2str(i) '_' tempDicomInfo.P
atientBirthDate '_' tempDicomInfo.PatientName.FamilyName '_' tempDicomInfo.PatientNa
me.GivenName '_' tempDicomInfo.AcquisitionDate '.gif']
50.     imwrite(planeMIPReduced, outputFilepath);
51.
52. end

```

Skript 2: CreateNiftiFromDicom.m

```
1. CreateNiftiFromDicom.m
2.
3.
4. pathImages = '/DATA/hammesj/DatscanAI/Rohdaten/Datscan_Meier_allFiles/';
5. pathOutputNiftis = '/DATA/hammesj/DatscanAI/output_Niftis_FBP/';
6. pathFilterList = '/DATA/hammesj/DatscanAI/outputImagesForTensorFlow/Datscan_AI_FileL
   ist_TF_new.csv';
7.
8.
9. %Find all Files in Input Folder
10. inputFiles = dir(pathImages);
11. inputFiles(1:2) = [];
12.
13. %filterList = importfile('/DATA/hammesj/DatscanAI/outputImagesForTensorFlow/Datscan_
   AI_FileList_TF.csv')
14. [ID,Auswertung,Rekonstruktion,DopaminergesDefizit] = importfile2(pathFilterList);
15.
16. %Filter
17. filteredIDs = ID((Auswertung == 'ja') & (Rekonstruktion == 'iterativ'));
18. filteredIDs = ID((Auswertung == 'ja') & (Rekonstruktion == 'FBP'));
19.
20.
21. for i = 1:length(inputFiles)
22.
23.     currentID = i;
24.
25.     % Check
26.     if ismember(currentID,filteredIDs)
27.
28.         inputFilepath = [pathImages inputFiles(i).name];
29.
30.         tempDicom = dicomread(inputFilepath);
31.         tempDicomInfo = dicominfo(inputFilepath);
32.
33.         tempDicomInfo.PatientName;
34.
35.         temp3D = tempDicom(:,:,1,:);
36.         currentVolume = squeeze(temp3D);
37.
38.         %Rotate by 270 (i.e.3*90) degrees slice-wise
39.         for j = 1:size(currentVolume,3)
40.
41.             currentVolume(:,:,j) = rot90(squeeze(currentVolume(:,:,j)));
42.             currentVolume(:,:,j) = rot90(squeeze(currentVolume(:,:,j)));
43.             currentVolume(:,:,j) = rot90(squeeze(currentVolume(:,:,j)));
44.
45.         end
46.
47.         currentVolume = flip(currentVolume, 3);
48.
49.         if size(currentVolume,3) < 30
50.             zDim = 4;
51.         else
52.             zDim = 2;
53.
54.         end
55.
56.         currentNifti = make_nii(currentVolume, [2 2 zDim]);
57.
58.         %Save nifti
```

```

59.         outputFilepath = [pathOutputNiftis 'Nifti_' num2str(i) '_' tempDicomInfo.Pat
    ientBirthDate '_' tempDicomInfo.PatientName.FamilyName '_' tempDicomInfo.PatientName
    .GivenName '_' tempDicomInfo.AcquisitionDate '.nii'];
60.         save_nii(currentNifti, outputFilepath);
61.
62.         disp(i);
63.
64.
65.     end
66.
67. end

```

Skript 3: SetOriginInNiftis.m

```

1. SetOriginInNiftis.m
2.
3.
4. pathImages = '/DATA/hammesj/DatscanAI/output_Niftis/';
5. pathOutputNiftis = '/DATA/hammesj/DatscanAI/outputNiftis_OriginSet/';
6.
7.
8. %Find all Files in Input Folder
9. inputFiles = dir(pathImages);
10. inputFiles(1:2) = [];
11.
12. for i = 1:length(inputFiles)
13.
14.     currentNifti = load_nii([pathImages inputFiles(i).name]);
15.
16.     currentNifti.hdr.hist.originator(1) = 64;
17.     currentNifti.hdr.hist.originator(2) = 78;
18.     currentNifti.hdr.hist.originator(3) = currentNifti.hdr.dime.dim(4) * 0.3;
19.
20.     disp(i);
21.     save_nii(currentNifti, [pathOutputNiftis 'OS' inputFiles(i).name])
22.
23.
24. end

```

Skript 4: RemoveAccents.m

```

1. RemoveAccents.m
2.
3.
4. pathImagesOS = '/DATA/meierm/DatScan_AI/DatScanNiftiOS/';
5. pathOutputImagesNoAccents = '/DATA/meierm/DatScan_AI/DatScanNiftiOS_NoAccents/';
6.
7. %Find all DatScanNiftiOS files, find, remove and replace accents
8. inputImages = dir(pathImagesOS);
9. inputImages(1:2) = [];
10.
11. Umlautvektor = {'Ä', 'ä', 'Ö', 'ö', 'Ü', 'ü', 'é', 'è', 'ê'}
12. Ersatzvektor = {'A', 'a', 'O', 'o', 'U', 'u', 'e', 'e', 'e'}
13.
14. for i = 1:length(inputImages)
15.     current_filename = inputImages(i).name;
16.
17.     for j = 1:length(Umlautvektor)
18.         current_filename = strrep(current_filename, Umlautvektor{j}, Ersatzvektor{j}
    );

```



```

19.     end
20.     copyfile([pathImagesOS inputImages(i).name], [pathOutputImagesNoAccents curr
ent_Filename])
21.     %Accents = 'á,à,â,é,è,ê,í,ì,î,ó,ò,ô,ú,ù,û';
22.     %inputImagesAccInfo = ismember(Accents,inputFilepath);
23.     %OutputImagesNoAccents = replace(inputFilepath,'á,à,â,é,è,ê,í,ì,î,ó,ò,ô,ú,ù,û','
a,e,i,o,u');
24.     %OutputFilepath = [pathOutputImagesNoAccents OutputImagesNoAccents(i).name(1:end
-4) '.gif'];
25.     %imwrite(inputImages,OutputFilepath);
26. end

```

Skript 5: BatchNormalizeDatScan.m

```

1. BatchNormalizeDatScan.m
2.
3. % Normalisiert DatScans zu 123I-FP-CIT.nii template
4. % anatomical information (Basal Ganglia!)
5.
6.
7. dirDatScans = '/DATA/meierm/DatScan_AI/DatScanNiftiOS_NoAccents/';
8. dirToScripts = '/DATA/meierm/MatlabScripts/Normalize_DatScans/';
9.
10. spectFilesDatScans = dir(fullfile(dirDatScans,'OS*.nii'))
11.
12.
13.
14.
15. for i = 1:numel(spectFilesDatScans)
16.     spm_jobman('initcfg');
17.
18.     currentDatScanFile = spectFilesDatScans(i).name;
19.
20.     fin = fopen(strcat(dirToScripts,'NormalisierungDatScanTemplate.m'),'r');
21.     fout = fopen(strcat(dirToScripts,'NormalisierungDatScanIntermed.m'),'w');
22.
23.
24.     findstr1 = 'REPLACE_DATSCAN_IMAGE';
25.     replacestr1 = strcat(dirDatScans, currentDatScanFile);
26.
27.
28.     while ~feof(fin)
29.         s = fgetl(fin);
30.         s = strrep(s, findstr1, replacestr1);
31.         fprintf(fout, '%s\n', s)
32.     end
33.
34.
35.     fclose(fin)
36.     fclose(fout)
37.
38.
39.     disp(currentDatScanFile);
40.     disp(i)
41.
42.
43.     %CODE FOR MATLAB BATCH PROCESS
44.     % List of open inputs
45.     nrun = 1; % enter the number of runs here
46.
47.     jobfile = {strcat(dirToScripts,'NormalisierungDatScanIntermed.m')};
48.
49.
50.     spm('defaults', 'PET');

```

```

51.         spm_jobman('run', jobfile);
52.     %END CODE FOR MATLAB BATCH PROCESS
53.
54. end

```

Skript 6: MIPConvertGifToJpg.m

```

1.  MIPConvertGifToJpg.m
2.
3.  pathImagesMIP = '/DATA/meierm/DatScan_AI/DatScanNifti_MIPs_NoAccents/';
4.  pathOutputImagesJPG = '/DATA/meierm/DatScan_AI/DatScanNifti_MIPs_JPG/';
5.
6.  %Find all DatScan MIPs (gif-Format)
7.  inputMIPs = dir(pathImagesMIP);
8.  inputMIPs(1:2) = [];
9.
10. for i = 1:length(inputMIPs)
11.     current_Gif = inputMIPs(i).name;
12.
13.     I = imread([pathImagesMIP current_Gif]);
14.     imwrite(I, [pathOutputImagesJPG current_Gif(1:end-3) 'jpg']);
15. end

```

Skript 7: dataAugmentation_flipImages.m

```

1.  dataAugmentation_flipImages.m
2.
3.
4.  clear all
5.
6.  %inputPath =
    '/DATA/hammesj/DatscanAI/DataAugmentation/original_unaugmented_Files/datscans_native
    Images/';
7.  inputPath =
    '/DATA/hammesj/DatscanAI/DataAugmentation/original_unaugmented_Files/datscans_normal
    izedImages/';
8.
9.  %outputPath =
    '/DATA/hammesj/DatscanAI/DataAugmentation/flippedImages/flipped_datscans_nativeImage
    s/';
10. outputPath =
    '/DATA/hammesj/DatscanAI/DataAugmentation/flippedImages/flipped_datscans_normalizedI
    mages/';
11.
12.
13. % find subfolders in parent directory, these folders will be kept also in
14. % the augmented images
15. subfolders = dir(inputPath);
16. subfolders(1:2) = [];
17. subfolders = subfolders([subfolders(:).isdir])
18.
19. for i = 1:size(subfolders,1)
20.     mkdir(outputPath, subfolders(i).name);
21.
22.     filesInCurrentSubfolder = dir([inputPath subfolders(i).name]);
23.     filesInCurrentSubfolder(1:2) = [];
24.
25.     %Read images, flip and save to output path
26.     for j = 1:size(filesInCurrentSubfolder,1)

```

```

27.         currentImage = imread([inputPath subfolders(i).name '/'
filesInCurrentSubfolder(j).name]);
28.         flippedImage = flipdim(currentImage, 2);
29.         imwrite(flippedImage, [outputPath subfolders(i).name '/' 'flipped_'
filesInCurrentSubfolder(j).name]);
30.     end
31.
32. end
33.
34.
35. display('done.');
```

Skript 8: dataAugmentation_rotateImages.m

```

1. dataAugmentation_rotateImages.m
2.
3.
4.
5.
6.
7.
8.
9.
10. clear all
11.
12.
13.
14.
15. %inputPath =
'/DATA/hammesj/DatscanAI/DataAugmentation/original_unaugmented_Files/datscans_native
Images/';
16. inputPath =
'/DATA/hammesj/DatscanAI/DataAugmentation/original_unaugmented_Files/datscans_normal
izedImages/';
17.
18.
19.
20.
21. %outputPath =
'/DATA/hammesj/DatscanAI/DataAugmentation/flippedImages/flipped_datscans_nativeImage
s/';
22. outputPath = '/DATA/meierm/DatScan_AI/Data_Augmentation/Normalized_rotated/'
23.
24.
25.
26.
27.
28.
29.
30.
31. % find subfolders in parent directory, these folders will be kept also in
32. % the augmented images
33. subfolders = dir(inputPath);
34. subfolders(1:2) = [];
35. subfolders = subfolders([subfolders(:).isdir])
36.
37.
38.
39.
40. for i = 1:size(subfolders,1)
41.     mkdir(outputPath, subfolders(i).name);
42.
43.     filesInCurrentSubfolder = dir([inputPath subfolders(i).name]);
44.     filesInCurrentSubfolder(1:2) = [];
```

```

45.
46.     %Read images, flip and save to output path
47.     for j = 1:size(filesInCurrentSubfolder,1)
48.         currentImage = imread([inputPath subfolders(i).name '/'
filesInCurrentSubfolder(j).name]);
49.         rotatedImageCounter = imrotate(currentImage, 10);
50.         rotatedImageClock = imrotate(currentImage, -10);
51.         imwrite(rotatedImageCounter, [outputPath subfolders(i).name '/' 'rotatedCC_'
filesInCurrentSubfolder(j).name]);
52.         imwrite(rotatedImageClock, [outputPath subfolders(i).name '/' 'rotatedCl_'
filesInCurrentSubfolder(j).name]);
53.     end
54.
55. end
56.
57.
58.
59.
60.
61.
62.
63.
64. display('done.');
```

Skript 9: Extract_40_random_Normalized_Rotated.m

```

1. Extract_40_random_Normalized_Rotated.m
2.
3. %Extract 20 random positive and 20 random negative files (40 random files
4. %in total)%
5. Dest_Pos = '/home/meierm/tensorflow/tensorflow-for-poets-
2/tf_files/Evaluation_Testdatensatze/Testdatensatz_Normalized_Rotated/Pos/';
6. Folder_Pos = '/home/meierm/tensorflow/tensorflow-for-poets-
2/tf_files/Evaluation_Testdatensatze/Trainingsdatensatz_Normalized_Rotated/Pos/'
7. FileList_Pos = dir(fullfile(Folder_Pos, '*.jpg'));
8. Index1 = randperm(numel(FileList_Pos), 20);
9. for k = 1:20
10. Source_Pos = fullfile(Folder_Pos, FileList_Pos(Index1(k)).name);
11. movefile(Source_Pos, Dest_Pos);
12. end
13.
14. Dest_Neg = '/home/meierm/tensorflow/tensorflow-for-poets-
2/tf_files/Evaluation_Testdatensatze/Testdatensatz_Normalized_Rotated/Neg/';
15. Folder_Neg = '/home/meierm/tensorflow/tensorflow-for-poets-
2/tf_files/Evaluation_Testdatensatze/Trainingsdatensatz_Normalized_Rotated/Neg/'
16. FileList_Neg = dir(fullfile(Folder_Neg, '*.jpg'));
17. Index2 = randperm(numel(FileList_Neg), 20);
18. for l = 1:20
19. Source_Neg = fullfile(Folder_Neg, FileList_Neg(Index2(l)).name);
20. movefile(Source_Neg, Dest_Neg);
21. end
```

Skript 10: runTraining_mobilenet_v1.0_normalized_Rotated.sh

```

1. runTraining_mobilenet_v1.0_normalized_Rotated.sh
2.
3. IMAGE_SIZE=128
4. ARCHITECTURE="mobilenet_1.0_${IMAGE_SIZE}"
5. python -m scripts.retrain \
6. --bottleneck_dir=tf_files/bottlenecks \
7. --how_many_training_steps=1000 \
8. --model_dir=tf_files/models/ \
```

```

9.  --summaries_dir=tf_files/training_summaries/"${ARCHITECTURE}" \
10. --output_graph=tf_files/retrained_graph.pb \
11. --output_labels=tf_files/retrained_labels.txt \
12. --architecture="${ARCHITECTURE}" \
13.  --
    image_dir=tf_files/Evaluation_Testdatensaetze/Trainingsdatensatz_normalized_Rotated/

```

Skript 11: label_image_normalized_Rotated.py

```

1.  label_image_normalized_Rotated.py
2.
3.
4.
5.
6.  # Copyright 2017 The TensorFlow Authors. All Rights Reserved.
7.  #
8.  # Licensed under the Apache License, Version 2.0 (the "License");
9.  # you may not use this file except in compliance with the License.
10. # You may obtain a copy of the License at
11. #
12. #   http://www.apache.org/licenses/LICENSE-2.0
13. #
14. # Unless required by applicable law or agreed to in writing, software
15. # distributed under the License is distributed on an "AS IS" BASIS,
16. # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
17. # See the License for the specific language governing permissions and
18. # limitations under the License.
19. # =====
20.
21.
22.
23.
24. from __future__ import absolute_import
25. from __future__ import division
26. from __future__ import print_function
27.
28.
29.
30.
31. import argparse
32. import sys
33. import time
34. import os
35.
36.
37.
38.
39. import numpy as np
40. import tensorflow as tf
41.
42.
43.
44.
45. def load_graph(model_file):
46.     graph = tf.Graph()
47.     graph_def = tf.GraphDef()
48.
49.
50.
51.
52.     with open(model_file, "rb") as f:
53.         graph_def.ParseFromString(f.read())
54.     with graph.as_default():

```

```

55.     tf.import_graph_def(graph_def)
56.
57.
58.
59.
60.     return graph
61.
62.
63.
64.
65. def read_tensor_from_image_file(file_name, input_height=299, input_width=299,
66.                               input_mean=0, input_std=255):
67.     input_name = "file_reader"
68.     output_name = "normalized"
69.     file_reader = tf.read_file(file_name, input_name)
70.     if file_name.endswith(".png"):
71.         image_reader = tf.image.decode_png(file_reader, channels = 3,
72.                                           name='png_reader')
73.     elif file_name.endswith(".gif"):
74.         image_reader = tf.squeeze(tf.image.decode_gif(file_reader,
75.                                                      name='gif_reader'))
76.     elif file_name.endswith(".bmp"):
77.         image_reader = tf.image.decode_bmp(file_reader, name='bmp_reader')
78.     else:
79.         image_reader = tf.image.decode_jpeg(file_reader, channels = 3,
80.                                           name='jpeg_reader')
81.     float_caster = tf.cast(image_reader, tf.float32)
82.     dims_expander = tf.expand_dims(float_caster, 0);
83.     resized = tf.image.resize_bilinear(dims_expander, [input_height, input_width])
84.     normalized = tf.divide(tf.subtract(resized, [input_mean]), [input_std])
85.     sess = tf.Session()
86.     result = sess.run(normalized)
87.
88.
89.
90.
91.     return result
92.
93.
94.
95.
96. def load_labels(label_file):
97.     label = []
98.     proto_as_ascii_lines = tf.gfile.GFile(label_file).readlines()
99.     for l in proto_as_ascii_lines:
100.         label.append(l.rstrip())
101.     return label
102.
103.
104.
105.
106.     if __name__ == "__main__":
107.         file_name = "/home/meierm/tensorflow/tensorflow-for-poets-
108. 2/tf_files/Evaluation_Testdatensaetze/Trainingsdatensatz_normalized_Rotated/rota
109. tedCl_wOSNifti_212_XXXXXXXXX_XXXXXX_XXXXX_20120831.jpg"
110.         model_file = "/home/meierm/tensorflow/tensorflow-for-poets-
111. 2/tf_files//retrained_graph.pb"
112.         label_file = "/home/meierm/tensorflow/tensorflow-for-poets-
113. 2/tf_files/retrained_labels.txt"
114.         input_height = 128
115.         input_width = 128
116.         input_mean = 128
117.         input_std = 128
118.         input_layer = "input"
119.         output_layer = "final_result"

```

```

118.
119.
120.     parser = argparse.ArgumentParser()
121.     parser.add_argument("--image", help="image to be processed")
122.     parser.add_argument("--graph", help="graph/model to be executed")
123.     parser.add_argument("--labels", help="name of file containing labels")
124.     parser.add_argument("--input_height", type=int, help="input height")
125.     parser.add_argument("--input_width", type=int, help="input width")
126.     parser.add_argument("--input_mean", type=int, help="input mean")
127.     parser.add_argument("--input_std", type=int, help="input std")
128.     parser.add_argument("--input_layer", help="name of input layer")
129.     parser.add_argument("--output_layer", help="name of output layer")
130.     args = parser.parse_args()
131.
132.
133.
134.
135.     if args.graph:
136.         model_file = args.graph
137.     if args.image:
138.         file_name = args.image
139.     if args.labels:
140.         label_file = args.labels
141.     if args.input_height:
142.         input_height = args.input_height
143.     if args.input_width:
144.         input_width = args.input_width
145.     if args.input_mean:
146.         input_mean = args.input_mean
147.     if args.input_std:
148.         input_std = args.input_std
149.     if args.input_layer:
150.         input_layer = args.input_layer
151.     if args.output_layer:
152.         output_layer = args.output_layer
153.
154.
155.
156.
157.     graph = load_graph(model_file)
158.
159.     #pathToFiles = "/home/hammesj/tensorflow/tensorflow-for-poets-
2/tf_files/Testdatensatz_Normalized/"
160.     pathToFiles = "/home/meierm/tensorflow/tensorflow-for-poets-
2/tf_files/Evaluation_Testdatensaeetze/Testdatensatz_normalized_Rotated/"
161.     myFileList = os.listdir(pathToFiles)
162.
163.
164.
165.
166.     for filename in myFileList:
167.
168.
169.         #t = read_tensor_from_image_file(file_name,
170.         #print(pathToFiles + filename)
171.         t = read_tensor_from_image_file(pathToFiles + filename,
172.                                         input_height=input_height,
173.                                         input_width=input_width,
174.                                         input_mean=input_mean,
175.                                         input_std=input_std)
176.
177.
178.
179.
180.         input_name = "import/" + input_layer
181.         output_name = "import/" + output_layer
182.         input_operation = graph.get_operation_by_name(input_name);

```

```

183.         output_operation = graph.get_operation_by_name(output_name);
184.
185.
186.
187.
188.         with tf.Session(graph=graph) as sess:
189.             start = time.time()
190.             results = sess.run(output_operation.outputs[0],
191.                               {input_operation.outputs[0]: t})
192.             end=time.time()
193.             results = np.squeeze(results)
194.
195.
196.
197.
198.
199.
200.             top_k = results.argsort()[-5:][::-1]
201.             labels = load_labels(label_file)
202.
203.
204.
205.
206.             # print(top_k)
207.             # print(results)
208.             template = "{}; {:.5f}; "
209.             sys.stdout.write(filename + "; ")
210.             sys.stdout.write(template.format(labels[0], results[0]))
211.             sys.stdout.write(template.format(labels[1], results[1]))
212.             print(" ")
213.
214.
215.
216.
217.             """print('\nEvaluation time (1-image): {:.3f}s\n'.format(end-start))
218.             template = "{} (score={:.5f})"
219.             for i in top_k:
220.                 print(template.format(labels[i], results[i]))"""

```


8. Vorabveröffentlichungen von Ergebnissen

Erste Ergebnisse der vorliegenden Arbeit wurden auf den folgenden wissenschaftlichen Veranstaltungen in Form eines Posters vorgestellt. Das Poster ist auf der Folgeseite aufgeführt.

Hammes J, Meier M, Drzezga A, Dratsch T, Kobe C, Pinto dos Santos D, van Eimeren T. Using Artificial Intelligence on Dopamine Transporter SPECT. 2019. Düsseldorf: Deutscher Kongress für Parkinson und Bewegungsstörungen (DPG), 07.03.-09.03.2019.

Hammes J, Meier M, Drzezga A, Dratsch T, Kobe C, Pinto dos Santos D, van Eimeren T. Using Artificial Intelligence on Dopamine Transporter SPECT. 2019. Lissabon, Portugal: The 14th International Conference on Alzheimer's and Parkinson's Diseases (ADPD), 26.03.-31.03.2019.

Hammes J, Meier M, Drzezga A, Dratsch T, Kobe C, Pinto dos Santos D, van Eimeren T. Using Artificial Intelligence on Dopamine Transporter SPECT. 2019. Bonn: 1st Bonn-Köln Neuroscience Symposium, 03.07.2019.

Hammes J, Meier M, Drzezga A, Dratsch T, Kobe C, Pinto dos Santos D, van Eimeren T. Using Artificial Intelligence on Dopamine Transporter SPECT. 2019. Nizza, Frankreich: International Congress of Parkinson's Disease and Movement Disorders (MDS), 22.09.-26.09.2019.



Using Artificial Intelligence on Dopamine Transporter SPECT

Hammes J¹, Meier M¹, Drzezga A¹, Dratsch T², Kobe C¹, Pinto dos Santos D², van Eimeren T¹

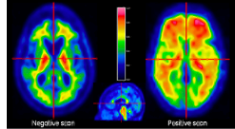
1) University Hospital Cologne, Department of Nuclear Medicine, Cologne, Germany

2) University Hospital Cologne, Department of Radiology, Cologne, Germany

Artificial Intelligence and Machine Learning

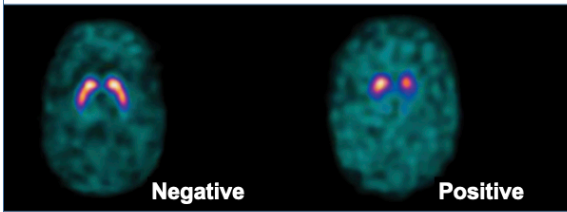
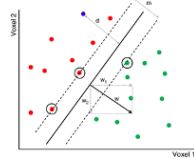


Artificial Intelligence is used for autonomous driving



Artificial Intelligence can also be used for classification in medical imaging

Vandenberghe et al. NeuroImage. 2013



Positive Visual Reads of Dopamine Transporter Imaging using FP-CIT SPECT reliably detect degenerative Parkinsonian disorders.

Winogrodzka et al. J Neurol Transm. 2001

Aim: Train a virtual neural network on Dopamine Transporter SPECT to predict expert visual reads

Patients and Methods

Training data:

- 442 FP-CIT SPECT data sets were acquired
- Reconstruction with an OSEM algorithm
- Binarized expert visual reads regarding pathology (positive / negative)
- Data augmentation for more and more heterogenous training data

Composition of the patient data

Expert Visual Read	Positive	Negative
Number of patients	250	192
Percentage (%)	56,6	43,4

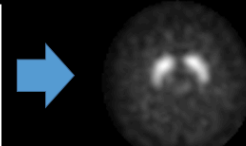
Image pre-processing

- Calculation of cranio-caudal Maximum Intensity Projections (MIPs) -> two-dimensional input images
- Spatial normalization using a 123I-FP-CIT template

Network Training

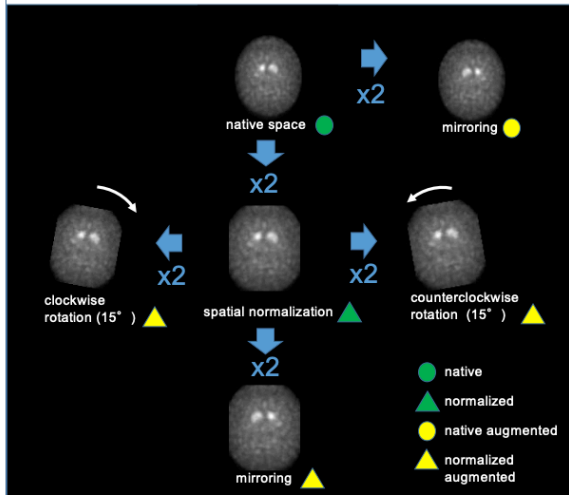
- Set-up of a TensorFlow® AI environment
- Training of Inception Net after removal of the last training layer
- Training of various different convolutional neural networks (MobileNets)
- Cross validation of classification performance by leave-n-out approach

Creation of axial Maximum Intensity Projections (MIPs)



Transfer from three-dimensional to two-dimensional images

Data Augmentation: rotation and flipping



Validation strategy: Division of the data set into two parts -> 85% of the data for learning, 15% for testing

Results

MobileNet Version / Data Set used	correctly classified (%)	incorrectly classified (%)
Version 0.5 / ●	77.50	22.50
Version 0.75 / ●	80.00	20.00
Version 1.0 / ●	77.50	22.50
Version 0.5 / ▲	77.50	22.50
Version 0.75 / ▲	65.00	35.00
Version 1.0 / ▲	80.00	20.00
Version 0.5 / ●	77.50	22.50
Version 0.5 / ▲	77.50	22.50

Conclusions

- AI-based image processing and analysis of FP-CIT images is possible
- Classification performance of images requires further improvement
- Data augmentation, threshold analysis and Region of Interest (ROI) analysis are options to enhance classification performance