

# **Diplomarbeit**

## **Explorative Analyse zur Umsetzung eines automatischen Läsionssegmentationsverfahrens bei MS im unizentrischen Setting**

eingereicht von

**Tina Primessnig**

zur Erlangung des akademischen Grades

**Doktorin der gesamten Heilkunde  
(Dr<sup>in</sup>. med. univ.)**

an der

**Medizinischen Universität Graz**

ausgeführt an der

**Universitätsklinik für Neurologie**

unter der Anleitung von

**Univ.-Prof. Priv.-Doz. Dr.med.univ. Christian Enzinger, MBA**

und

**Assoz. Prof. Univ.-Doz. Dipl.-Ing. Dr. Stefan Ropele**

Graz, 30.06.2024

## Eidesstattliche Erklärung

*Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.*

*Graz, am 30.06.2024*

*Tina Primessnig eh*

## Danksagungen

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich bei der Erstellung dieser Diplomarbeit unterstützt haben.

Zuerst möchte ich meinem Erstbetreuer, Univ.-Prof. Priv.-Doz. Dr.med.univ. Christian Enzinger, MBA, meinen Dank für die kompetente Betreuung, die Geduld, das Engagement in der Ideenfindung, sowie den Beitrag seiner herausragenden klinischen und wissenschaftlichen Expertise aussprechen.

Zudem gilt meinem Zweitbetreuer, Assoz. Prof. Univ.-Doz. Dipl.-Ing. Dr. Stefan Ropele, Dank für seine Unterstützung und die fachlichen Anregungen aus dem Forschungsbereich des Neuroimaging. Anhand seines umfassenden Inputs ließen sich für mich anfangs komplex wirkende Aspekte aus dem Bereich der medizinischen Bildverarbeitung im Rahmen dieser Diplomarbeit überwinden und realisieren.

Frau Sen.Scientist Dipl.-Ing.in Dr.in Regina Riedl möchte ich für ihren Beitrag bei der Erarbeitung meiner statistischen Auswertung danken.

Schließlich möchte ich mich ganz besonders bei meinen Eltern, meinem Bruder Uwe, meinem Freund Michi und meinen Freund\*innen bedanken. Ihr habt mich immer motiviert und während des Studiums stets bedingungslos unterstützt. Dank eures Rückhalts ließen sich jegliche Hürden, die sich im Laufe des Studiums auftaten, meistern.

# Zusammenfassung in Deutsch

## Einleitung

Die Multiple Sklerose (MS) ist eine chronische Autoimmunerkrankung, welche das Zentralnervensystem betrifft. Mit ca. 2,3 Mio. Erkrankten weltweit und einem Hauptmanifestationsalter zwischen 20 und 40 Jahren ist sie eine der häufigsten Ursachen neurologischer Beeinträchtigung junger Erwachsener (1). Das Auftreten und die Anzahl neuer T2-Läsionen in der MRT spielen eine zentrale Rolle bei der Erkennung von Krankheitsaktivität und beeinflussen wichtige Entscheidungen im Krankheitsmanagement, wie Therapieauswahl- oder -anpassung (2,3). Der bisherige Goldstandard zur Detektion neuer Läsionen, die manuelle Läsionssegmentierung, ist aufgrund der mangelnden flächendeckenden Verfügbarkeit erfahrener Neuroradiolog\*innen, der subjektiven Beurteilung durch Menschenhand, sowie des Zeitaufwands aufgrund der großen Menge zu analysierender MRT-Daten, mit klaren Einschränkungen verbunden. Es ist zu erkennen, dass ein entscheidungsunterstützendes System, welches die Läsionssegmentierung in der klinischen Routine unterstützen könnte, von großem Interesse für das Monitoring von MS-Patient\*innen in der klinischen Praxis sein würde.

## Methodik

In dieser retrospektiven Studie wurde in Zusammenarbeit mit dem Hospital Universitari Vall d'Hebron als Teil einer internationalen, multizentrischen Studie die Anwendbarkeit eines entscheidungsunterstützenden Systems, basierend auf einem neuronalen Netzwerk, zur automatischen Läsionssegmentierung überprüft. Die Studie wurde an einer Kohorte von 67 Patient\*innen des Grazer MS-Zentrums der Universitätsklinik für Neurologie Graz durchgeführt. Pro Patient\*in wurden zwei T2-gewichtete 3D-FLAIR-Sequenz MRT-Aufnahmen, die im Abstand von 6-24 Monaten an einem 3T Forschungs-Magnetresonanztomographen akquiriert wurden, analysiert. Drei Methoden wurden hinsichtlich der detektierten Anzahl neuer T2-Läsionen in den vorhandenen Baseline- und Follow-Up-MRT-Aufnahmen der Studienpopulation verglichen: die rein manuelle Segmentierung (Methode VA, bisheriger Goldstandard), die rein automatisierte Analyse über Copilot.ms (Methode AA), sowie die durch Expert\*innen validierte automatisierte Analyse (Methode AAValid, als entscheidungsunterstützende Methode).

## **Ergebnisse**

Die automatisierte und validiert automatisierte Analyse detektierten eine höhere Anzahl neuer T2-Läsionen als die manuelle Analyse, allerdings auf Kosten einer höheren Anzahl falsch positiver Ergebnisse. Es zeigte sich jedoch, dass die Anzahl falsch positiver Ergebnisse durch den Validierungsschritt verringert, und damit die Spezifität (86,00% (95%-KI 73,26-94,18%) bei Methode AA und 88,00% (95%-KI 75,69-95,47%)) bei Methode AAValid) erhöht werden kann. Die Sensitivitäten der automatisierten und validiert automatisierten Methoden erwiesen sich mit 94,12% (95%-KI 71,31-99,85%) als gleich hoch, woraus geschlossen werden kann, dass beide Methode nur wenige neue Läsionen übersehen. Insgesamt zeigte sich anhand des negativen Vorhersagewertes (99,99% (95%-KI 99,93-100,00%)), dass sich Methode AA und Methode AAValid gut zur Erkennung stabiler Patient\*innen eignen, allerdings aufgrund des geringen positiven Vorhersagewertes (Methode AA 1,06% (95%-KI 0,53-2,10%), Methode AAValid 1,23% (95%-KI 0,58-2,60%)) hinsichtlich der zuverlässigen Detektion neuer T2-Läsionen weiterentwickelt werden müssen.

## **Diskussion**

Automatisierte neuronale Netzwerke zur Detektion neuer T2-Läsionen bei MS-Patient\*innen sind Gegenstand aktueller Forschung und haben Potenzial die Läsionssegmentierung im klinischen Alltag zu unterstützen und diese zeitlich effizienter zu gestalten. Zudem senken sie, aufgrund der hohen Sensitivität, das Risiko neue T2-Läsionen, und somit krankheitsaktive Patient\*innen, zu übersehen. Zum aktuellen Zeitpunkt ist ein Validierungsschritt durch Expert\*innen jedoch unumgänglich, da die Fähigkeit, neue T2-Läsionen zuverlässig als solche zu erkennen, durch das in dieser Studie untersuchte neuronale Netzwerk noch nicht ausreichend gegeben ist, um schließlich wichtige Therapieentscheidungen davon abhängig zu machen.

# Abstract in Englisch

## Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease which affects the central nervous system and is one of the most common causes of neurological impairment in young adults with a main age of manifestation between 20 to 40 years, affecting approximately 2.3 million people worldwide (1). The occurrence and number of new T2 lesions on MRI play a central role in the detection of disease activity and influence important decisions in disease management, such as therapy selection or adjustment (2,3). The current gold standard for detecting new lesions is manual lesion segmentation, which has apparent limitations due to it being a time-consuming task given the large amount of MRI data to be analyzed, the lack of widespread availability of experienced neuroradiologists as well as subjective human judgment, which complicates objective standardization of the results. Thus, it appears that a decision-support system that could support lesion segmentation in clinical routine would be of great interest for the monitoring of MS patients in clinical practice.

## Methods

In this retrospective study, in collaboration with the Hospital Universitari Vall d'Hebron, the feasibility of a decision support system based on a neural network, called Copilot.ms, was tested on a cohort of 67 patients of the Graz MS Center of the University Clinic of Neurology Graz as part of an international, multicenter study. Two T2-weighted 3D FLAIR sequence MRI images per patient, which were acquired on a 3T research MRI scanner 6-24 months apart, were examined. Three methods were compared regarding the detected number of new T2 lesions in the existing baseline and follow-up MRI images of the study population: the purely manual segmentation (method VA, current gold standard), the purely automated analysis via Copilot.ms (method AA), and the validated automated analysis (method AAValid, as a decision-supporting method).

## Results

The automated and validated automated analysis detected a higher number of new T2 lesions than the manual analysis, at the expense of a larger number of false positive results. However, it was shown that the number of false positive results can be reduced by the validation step, and thus the specificity (86.00% (95% CI 73.26-94.18%)) for method AA

and 88.00% (95% CI 75.69-95.47%)) for method AAValid) can be increased. The sensitivities of the automated and validated automated methods proved to be equally high at 94.12% (95% CI 71.31-99.85%), which allows the conclusion that both methods only miss a small number of new lesions. Overall, the negative predictive value (99.99% (95% CI 99.93-100.00%)) showed that method AA and AAValid are well suited for the detection of stable patients. However, they need to be further developed with regard to the reliable detection of new T2 lesions due to the low positive predictive value (method AA 1.06% (95% CI 0.53-2.10%), method AAValid 1.23% (95% CI 0.58-2.60%)).

## **Discussion**

Automated neural networks are subject of current research regarding the detection of new T2 lesions in MS patients and show potential to support lesion segmentation in everyday clinical practice, particularly in regard to time efficiency of lesion segmentation. In addition, due to its high sensitivity, the automated neural network studied reduces the risk of overlooking new T2 lesions and thus disease-active patients. At present, however, a validation step by experts is unavoidable, as the ability to reliably recognize new T2 lesions as such is not yet sufficiently given by the investigated neural network to ultimately make important decisions of disease management dependent on it.

# Inhaltsverzeichnis

<b>Eidesstattliche Erklärung</b> .....	<b>i</b>
<b>Danksagungen</b> .....	<b>ii</b>
<b>Zusammenfassung in Deutsch</b> .....	<b>iii</b>
<b>Abstract in Englisch</b> .....	<b>v</b>
<b>Inhaltsverzeichnis</b> .....	<b>vii</b>
<b>Abkürzungen und deren Erklärung</b> .....	<b>ix</b>
<b>Abbildungsverzeichnis</b> .....	<b>xii</b>
<b>Tabellenverzeichnis</b> .....	<b>xiii</b>
<b>1 Einleitung</b> .....	<b>1</b>
1.1 Epidemiologie.....	1
1.2 Ätiopathogenese.....	2
1.3 Charakteristika der Multiplen Sklerose .....	5
1.3.1 Symptomatik .....	5
1.3.2 Klinische Phänotypen und Verlaufsformen.....	7
1.3.2.1 Radiologisch isoliertes Syndrom .....	7
1.3.2.2 Klinisch isoliertes Syndrom.....	8
1.3.2.3 Die schubförmig remittierende Multiple Sklerose (RRMS).....	9
1.3.2.4 Die sekundär progrediente Multiple Sklerose (SPMS).....	10
1.3.2.5 Die primär progrediente Multiple Sklerose (PPMS) .....	11
1.3.2.6 Aktivität und Progression .....	11
1.4 Diagnostik der Multiplen Sklerose .....	12
1.4.1 Diagnosekriterien nach McDonald.....	14
1.4.2 Die Rolle der Magnetresonanztomographie (MRT).....	15
1.5 Segmentierungstechniken zur Läsionsquantifizierung in der MRT .....	19
1.5.1 Bildvorverarbeitung.....	21
1.5.2 Läsionssegmentierung .....	22
1.6 Prognose .....	25
1.7 Therapie .....	27
1.8 Fragestellung und Zielsetzung .....	30
<b>2 Material und Methoden</b> .....	<b>32</b>
2.1 Studienaufbau .....	32
2.2 Patient*innen-Kohorte .....	32
2.3 MRT-Untersuchungen und Einschlusskriterien.....	34
2.4 MRT-Analysen und Vergleiche .....	35
2.4.1 Manuelle Segmentierung.....	35

2.4.2	Automatisierte Analyse .....	35
2.4.2.1	Deformierbare Modelle.....	36
2.4.2.2	Ablauf der automatisierten Analyse .....	37
2.4.3	Automatisierte Analyse mit Validierung.....	38
2.5	Statistische Auswertung.....	42
<b>3</b>	<b>Ergebnisse .....</b>	<b>43</b>
3.1	Absolute Zahlen neuer T2-Läsionen.....	43
3.2	Sensitivität, Spezifität, Vorhersagewerte.....	46
3.3	Test zum Vergleich der Methoden.....	50
<b>4</b>	<b>Diskussion .....</b>	<b>53</b>
4.1	Conclusio .....	63
	<b>Literaturverzeichnis .....</b>	<b>64</b>

## Abkürzungen und deren Erklärung

2D	Zweidimensional
3D	Dreidimensional
AA	Rein automatisierte Analyse mit der Plattform Copilot.ms
AAValid	Vorläufige Analyse durch die Plattform Copilot.ms, gefolgt von einer visuellen Validierung
ANNs	Artificial Neural Networks / Künstliche neuronale Netzwerke
AQP4	Aquaporin-4
BMI	Body-Mass-Index
CAD Systems	Computer-Aided Diagnosis Systems / computergestützte Diagnosesysteme
CD	Cluster Of Differentiation
CIS	Clinically Isolated Syndrome / Klinisch isoliertes Syndrom
CMSC	Consortium of Multiple Sclerosis Centers
Copilot.ms	Prototyp eines Programmes zur Läsionssegmentierung
DIS	Dissemination In Space / Örtliche Dissemination
DIT	Dissemination In Time / Zeitliche Dissemination
DMTs	Disease-Modifying Treatments / Verlaufsmodifizierende Therapie
DNA	Deoxyribonucleic Acid / Desoxyribonukleinsäure
EBV	Epstein-Barr-Virus
EDSS	Expanded Disability Status Scale
EK	Ethikkommission
FLAIR	Fluid Attenuated Inversion Recovery (MRT-Sequenz)
FN	Falsch negatives Ergebnis
FP	Falsch positives Ergebnis
HLA	Human Leukocyte Antigen
HPyV-2	Humanes Polyomavirus
ID	Identifikationsnummer
IQR	Interquartile range / Interquartilsabstand
JCV	John-Cunningham-Virus
KI	Künstliche Intelligenz
k-NN	k-Nächste-Nachbarn
MAGNIMS	Magnetic Resonance Imaging in Multiple Sclerosis

MIDAS	Medical Image Display and Analysis System
MOG	Myelin-Oligodendrozyten-Glykoprotein
MOGAD	Myelin-Oligodendrozyten-Glykoprotein-Antikörper-assoziierte Erkrankungen
MRT	Magnetresonanztomographie
MS	Multiple Sklerose
NAIMS	North American Imaging in Multiple Sclerosis Cooperative
NEDA	No Evidence Of Disease Activity / Keine Evidenz der Krankheitsaktivität
NMOSDs	Neuromyelitis-optica-Spektrum-Erkrankungen
NPV	Negative Predictive Value / Negativer Vorhersagewert
openMEDOCS	Krankenhausinformationssystem
PACS	Picture Archiving and Communication System / radiologisches Informationssystem
PML	Progressive multifokale Leukenzephalopathie
PPMS	Primary-progressive multiple sclerosis / Primär progrediente Multiple Sklerose
PPV	Positive Predictive Value / Positiver Vorhersagewert
PRMS	Progressive-relapsing multiple sclerosis / Progressiv remittierende Multiple Sklerose
RAPD	Relativ afferenter Pupillendefekt
RDA	Research, Data, Analysis (Datenbank wissenschaftsrelevanter Daten)
RIS	Radiologically Isolated Syndrome / Radiologisch isoliertes Syndrom
RN	Richtig negatives Ergebnis
RP	Richtig positives Ergebnis
RRMS	Relapsing-remitting multiple sclerosis / Schubförmig remittierende Multiple Sklerose
S1P	Sphingosin-1-Phosphat
SNPs	Single Nucleotide Polymorphism
SPMS	Secondary-progressive multiple sclerosis / Sekundär progrediente Multiple Sklerose
SPSS	Statistical Package for the Social Sciences, Statistikprogramm der Firma IBM

T	Tesla (Einheit der Magnetfeldstärke)
T1GE	Gradienten-Echo-Pulssequenz in der MRT
T1SE	Spin-Echo-Pulssequenz in der MRT
VA	Visuelle Analyse
ZNS	Zentralnervensystem

## Abbildungsverzeichnis

Abbildung 1: Beispielhafter Verlauf einer schubförmig remittierenden MS (RRMS) .....	10
Abbildung 2: Beispielhafter Verlauf einer sekundär progredienten MS (SPMS).....	11
Abbildung 3: Beispielhafter Verlauf einer primär progredienten MS (PPMS).....	11
Abbildung 4: Periventrikuläre T2-hyperintense Läsionen, FLAIR-Sequenz.....	16
Abbildung 5: Juxtakortikale T2-hyperintense Läsion, FLAIR-Sequenz.....	17
Abbildung 6: Infratentorielle T2-hyperintense Läsion, FLAIR-Sequenz .....	17
Abbildung 7: Blockschema eines typischen CAD-Systems zur Läsionssegmentierung.....	20
Abbildung 8: Kreisdiagramm zur Phänotypen-Verteilung in der Studienpopulation .....	34
Abbildung 9: Schritt 1 - Upload der MRT-Daten .....	39
Abbildung 10: Schritt 2 - Startseitenansicht von copilot.ms .....	39
Abbildung 11: Schritt 3 - Validierungsprozess .....	40
Abbildung 12: Beispiel Abschlussbericht Copilot.ms.....	41
Abbildung 13: Balkendiagramm zur Anzahl neuer T2-Läsionen nach Methode über die gesamte Studien-Kohorte (67 Teilnehmer*innen und 67 MRT-Scan-Paare) .....	46
Abbildung 14: Vorbestehende Läsion fälschlicherweise als neue Läsion erkannt .....	59
Abbildung 15: Beispiel für nicht erkannte shrinking lesion.....	60
Abbildung 16: Beispiel für nicht erkannte shrinking lesion.....	60

## Tabellenverzeichnis

Tabelle 1: Symptome der Multiplen Sklerose .....	6
Tabelle 2: McDonald-Kriterien von 2017 für die schubförmig remittierende MS .....	14
Tabelle 3: McDonald-Kriterien von 2017 für die primär progrediente MS .....	15
Tabelle 4: Merkmale mit Auswirkung auf die Prognose der MS.....	26
Tabelle 5: Deskriptive Statistik zur Studienpopulation.....	33
Tabelle 6: Deskriptive Statistik neu detektierter T2-Läsionen nach Methode über die Studienkohorte (67 Teilnehmer*innen und 67 MRT-Scan-Paare).....	45
Tabelle 7: Vergleich der Summe neuer T2-Läsionen nach Methode über die gesamte Studien-Kohorte (67 Teilnehmer*innen und 67 MRT-Scan-Paare).....	45
Tabelle 8: Kreuztabelle automatische Analyse x visuelle Analyse.....	47
Tabelle 9: Kreuztabelle validierte automatische Analyse x visuelle Analyse.....	47
Tabelle 10: Leistung der automatischen und validierten automatischen Methode in der Läsionserkennung.....	49
Tabelle 11: Null- und Alternativhypothesenübersicht des durchgeführten Vorzeichentests .....	51
Tabelle 12: Häufigkeit der errechneten Paardifferenzen mittels Vorzeichentest .....	51
Tabelle 13: Teststatistik der Vorzeichentests .....	52

# 1 Einleitung

## 1.1 Epidemiologie

Die Multiple Sklerose (MS) ist eine chronische Autoimmunerkrankung, welche das Zentralnervensystem (ZNS) betrifft und durch Demyelinisierung und axonalen Verlust eine der häufigsten Ursachen von neurologischen Beeinträchtigungen junger Erwachsener darstellt (1).

Weltweit sind ca. 2,3 Mio. Menschen von MS betroffen (4). Die mediane globale Prävalenz der MS liegt bei 33 Betroffenen pro 100 000 Einwohner\*innen, wobei sich deutliche geografische Unterschiede in der Prävalenz zeigen (1). Eine hohe Prävalenz findet sich beispielsweise in Nordamerika mit 140 pro 100 000 Personen und in Europa mit 108 pro 100 000 Personen (5). Einen der weltweiten Hotspots für die MS stellt Saskatchewan in Kanada mit einer Prävalenz von 313,6 pro 100 000 Personen dar (6). Am niedrigsten ist die Prävalenz in Afrika südlich der Sahara mit 2,1 pro 100 000 und Ostasien mit 2,2 pro 100 000 (5). Somit lässt sich festhalten, dass die Prävalenz der MS einem Nord-Süd-Gefälle unterliegt und daher mit steigendem Breitengrad zunimmt (7).

Salhofer-Polanyi et al. errechneten in einer retrospektiven Analyse die Inzidenz und Prävalenz der MS in Österreich auf Basis von Daten, welche in den Jahren 2010-2013 aus dem öffentlichen Versicherungssystem extrahiert wurden, wobei nach MS-spezifischer Medikation und MS als Entlassungsdiagnose (ICD-Code G35) selektiert wurde. Es ließ sich dadurch für Österreich eine Inzidenzrate von 19,5 pro 100 000 Personen pro Jahr und eine Prävalenz von 158,9 pro 100 000 Personen berechnen (8).

Generell lässt sich auf globaler Ebene ein Anstieg von Inzidenz und Prävalenz der MS erkennen (1). Diese Tendenz ist auch in Österreich sichtbar: frühere Studien von Baumhackl et al. zeigten geschätzte Prävalenzen von 98,5 pro 100 000 Personen im Jahr 1999 bzw. 149 pro 100 000 Personen in einer Follow-Up-Studie im Jahr 2010 (8). Mögliche Erklärungsversuche für den weltweiten Prävalenzanstieg sind erhöhtes Überleben in der Gesamtbevölkerung, Verbesserungen in der Diagnostik, sowie die Einrichtung klinischer Register und dadurch verbesserte Dokumentation der MS-Fälle (9). Epidemiologische Daten deuten außerdem auf einen weltweiten Inzidenzanstieg hin, welcher zumindest teilweise durch erhöhtes Bewusstsein der Krankheit gegenüber, bessere Gesundheitsversorgung, mehr MS-Spezialist\*innen, mehr Magnetresonanztomographie-

Geräte, sowie überarbeitete Diagnosekriterien und dadurch frühere und bessere Diagnostik erklärt werden könnte (10).

Das Hauptmanifestationsalter der MS liegt zwischen 20 und 40 Jahren, wobei Frauen zwei bis drei Mal häufiger erkranken als Männer (1). Auffällig ist, dass sich das Geschlechterverhältnis im Laufe der Zeit verändert hat, wobei der Anteil an erkrankten Frauen gestiegen ist. Dies dürfte in der ersten Hälfte des 20. Jahrhunderts darauf zurückzuführen sein, dass Männer durch ihre Berufsausübung häufiger krankenversichert waren und so besseren Zugang zu gesundheitlichen Diensten hatten, wodurch eine Diagnosestellung wahrscheinlicher war. In den Jahren danach dürfte der Anstieg unter anderem auf eine tatsächliche Vermehrung negativer Lifestyle-Faktoren wie vermehrtem Nikotinabusus oder Übergewicht in der Adoleszenz zurückzuführen sein (10).

Bereits 1971 wiesen Dean und Kurtzke auf mögliche Inzidenz-Veränderungen der MS bei Migration aus einem Hochrisikogebiet in ein Niedrigrisikogebiet hin, wobei sie eine Risikoabnahme der Erkrankung bei Migration vor dem 15. Lebensjahr postulierten (11). Mehrere Migrationsstudien unterstützen die Annahme, dass die Multiple Sklerose dem Einfluss von Umweltbedingungen unterliegt. Bei Migration aus einem Niedrigrisikogebiet in ein Hochrisikogebiet im Erwachsenenalter wird das niedrige Erkrankungsrisiko des Ursprungslandes beibehalten. Nachkommen von Einwanderern in Hochrisikogebiete hingegen erfahren ein hohes Erkrankungsrisiko (7).

## 1.2 Ätiopathogenese

Die Ätiologie der Multiplen Sklerose ist zum aktuellen Zeitpunkt noch nicht genau geklärt (1). Ein komplexes Zusammenspiel aus genetischer Suszeptibilität und Umweltfaktoren wird als ursächlich angenommen (4). Aktuell werden mehrere Umweltfaktoren mit einem erhöhten Risiko, an einer Multiplen Sklerose zu erkranken, in Verbindung gebracht (5). Dazu zählen eine Infektion mit dem Epstein-Barr-Virus (EBV), Rauchen, niedrige Vitamin-D-Spiegel, sowie ein erhöhter Body-Mass-Index (BMI) während der Adoleszenz (1).

Belbasis et al. haben vorhandene Reviews und Metaanalysen, die sich mit insgesamt 44 fraglichen Umweltfaktoren beschäftigten, welche die Entstehung der MS beeinflussen könnten, zusammengefasst und konnten die stärkste Assoziation zu der Entstehung einer MS bei einem positiven EBV-Biomarker (anti-EBNA IgG Seropositivität), einer symptomatischen EBV-Infektion (infektiöse Mononukleose) und Rauchen zeigen (5).

Für eine ursächliche Rolle des Epstein-Barr-Virus spricht, dass das MS-Risiko nach einer infektiösen Mononukleose und bei positiven EBV-Serum-Antikörpertitern erhöht ist. Außerdem konnten manche pathologische Studien EBV in MS-Läsionen aufzeigen. Trotzdem konnte eine Kausalität bisher noch nicht nachgewiesen werden, denn dies würde bedeuten, dass manche der Betroffenen, welche nach einer EBV-Infektion eine MS entwickelt haben, nicht an einer MS erkrankt wären, wäre nie eine EBV-Infektion aufgetreten. Um dieser Fragestellung weiter auf den Grund zu gehen, führten Bjornevik et al. eine retrospektive Analyse an einer Kohorte von über 10 Millionen US-Soldat\*innen durch, wo gezeigt werden konnte, dass das MS-Risiko nach einer EBV-Infektion 32-fach erhöht ist. Darüber hinaus war auffällig, dass einzig bei einem MS-Fall keine EBV-Infektion vorausgegangen war. Alle anderen 800 MS-Fälle waren EBV-positiv. Gestützt wurden diese Ergebnisse durch Nachweis erhöhter Serumkonzentrationen von Neurofilament-Leichtketten, einem Biomarker für neuro-axonale Degeneration, nach einer EBV-Serokonversion. Bjornevik et al. postulieren anhand genannter Ergebnisse, dass EBV eine Hauptursache der MS darstellt (12).

Der zugrundeliegende pathogene Mechanismus des EBV in der Entstehung einer MS ist ungeklärt. Diskutiert werden molekulares Mimikry, EBV-induzierte B-Zell-Immortalisierung und -Transformation (7). EBV kann allerdings nicht allein für die Entstehung der MS verantwortlich gemacht werden, da ein Großteil der Gesamtbevölkerung infiziert ist, aber nur ein kleiner Anteil eine MS entwickelt. Zusammenfassend kann gesagt werden, dass EBV für die Entstehung einer MS vermutlich notwendig ist, allerdings als alleiniger Trigger nicht ausreicht (13).

Genetische Faktoren stellen einen weiteren wichtigen Eckpfeiler in der MS-Entstehung dar. Ungefähr einer von acht Betroffenen hat eine positive Familienanamnese für MS (7). Zwillingsstudien aus Kanada und den USA zeigen bei eineiigen Zwillingen erhöhte klinische Konkordanzraten von 25% im Vergleich zu zweieiigen Zwillingen mit 5% (14). Das genetische Hauptrisiko, welches mit einer MS assoziiert ist, scheint von dem „human leukocyte antigen“ (HLA)-Gencluster auszugehen, genauer dem HLA-DRB1\*15, welches eine wichtige Rolle in der Antigenpräsentation der Immunabwehr spielt (7,13,15). Genomweite Assoziationsstudien konnten außerdem mehr als 150 mit einer MS-assozierte SNPs (single nucleotide polymorphisms) identifizieren, welche aber allesamt eine relativ geringe Odds Ratio von ca. 1,1-1,2 aufweisen (7).

Auf histopathologischer Ebene lassen sich in den frühen Phasen der Erkrankung perivenuläre, inflammatorische Läsionen finden, die in weiterer Folge zur Entstehung von demyelinisierenden Plaques in der weißen und grauen Substanz führen, welche das Hauptkennzeichen der MS darstellen (16). Die entzündlichen Infiltrate bestehen hauptsächlich aus CD8<sup>+</sup> T-Zellen und in kleineren Mengen auch aus B-Zellen, Plasmazellen und Autoantikörpern (16,17). Diese Gegebenheit lässt einen T-Zell-medierten Entzündungsprozess als krankheitsbestimmend annehmen, der durch Oligodendrozyten-Zerstörung zu Demyelinisierung führt. In fortgeschrittenen Krankheitsstadien kommt es schließlich auch zu irreversibler, axonaler Schädigung (7).

Die aktuell akzeptierteste Annahme ist, dass es sich bei den inflammatorischen Prozessen, welche eine MS antreiben, um Autoimmunkaskaden handelt. Autoreaktive T-Zellen (vorwiegend vom Th17-Phänotyp) überqueren vermehrt die Blut-Hirn-Schranke und richten sich vermutlich infolge von molekularer Mimikry gegen Myelin-Antigene (14,16). Das ursprüngliche, die Autoreaktivität auslösende Agens ist jedoch nach wie vor ungeklärt. Nichtsdestotrotz kann die Autoimmunhypothese unter anderem dadurch gestützt werden, dass immunmodulatorische Therapien Wirkung zeigen. Es spielen auch humorale Immunmechanismen eine wichtige Rolle. Bei Patient\*innen mit MS lässt sich eine erhöhte Zahl an B-Zellen im ZNS finden, wobei diese auch im Liquor persistieren und deren Anzahl mit der intrathekalen IgG-Synthese (oligoklonale Antikörper), welche für die Diagnostik einer MS relevant ist, und der aktiven Entzündung korreliert (16). Die Wirksamkeit B-Zell-depletierender Medikamente legt eine pathogenetische Beteiligung von B-Zellen ebenfalls nahe (18).

Neben Inflammation, Demyelinisierung und neuro-axonaler Degeneration kommt es bei der MS auch zu Remyelinisierung (14). Bei der Remyelinisierung werden axonale Myelinscheiden teilweise wiederhergestellt, was zu verbesserter Nervenleitgeschwindigkeit und Stoffwechselfersorgung der Neuronen führt. Die genaue Rolle der Remyelinisierung ist noch unklar, allerdings gibt es Hinweise darauf, dass Remyelinisierung neuroprotektiv wirken kann. Typisch für remyelinisierte Plaques sind dünnere Myelinscheiden und kürzere internodale Strecken als in normal erscheinender weißer Substanz (19). Remyelinisierte Bereiche werden als „shadow plaques“ bezeichnet und fallen als Bereiche reduzierter Myelin-Dichte auf (17).

## 1.3 Charakteristika der Multiplen Sklerose

### 1.3.1 Symptomatik

Die Multiple Sklerose kann weite Teile des ZNS betreffen, wobei motorische, sensorische, visuelle und autonome Systeme im Erkrankungsprozess beteiligt sein können, was eine große Bandbreite an möglichen neurologischen Symptomen ergibt (4,20).

Nur sehr wenige klinische Zeichen gelten als MS-spezifisch. Dazu zählen das sogenannte „Uhthoff-Phänomen“, welches eine vorübergehende Verschlechterung bereits vorhandener neurologischer Symptomatik bei erhöhter Körpertemperatur beschreibt, was auf eine temperaturabhängig veränderte elektrische Leitfähigkeit entlang demyelinisierter Axone zurückzuführen ist, und das „Lhermitte-Zeichen“ (20). Letzteres ist gekennzeichnet durch das Auftreten von elektrischen Missempfindungen entlang der Wirbelsäule und Extremitäten bei Nackenflexion (14).

Zu den häufigsten Erstmanifestationen der MS zählt eine Optikusneuritis, bei der es zu entzündlichen Veränderungen des Sehnerven kommt. Typische Anzeichen einer Optikusneuritis sind eine subakute Sehverschlechterung, welche mit Gesichtsfeldausfällen, charakteristischerweise einem Zentralskotom, einhergehen kann, sowie mit Schmerzen bei Augenbewegung, verändertem Farbsehen, sowie einem relativ afferenten Pupillendefekt (RAPD), wodurch es bei schneller und wiederholter abwechselnder Beleuchtung beider Augen zu einer paradoxen Pupillenerweiterung am betroffenen Auge kommt. In 90% der Fälle ist vorwiegend ein Auge betroffen (20).

Weitere relevante Frühmanifestationen der MS sind fokale Entzündungen des Rückenmarks (Myelitis), sowie Läsionen im Hirnstamm oder Kleinhirn. Klinische Zeichen der Myelitis sind motorische oder sensible Ausfälle unterhalb des betroffenen Rückenmarksniveaus, wobei zu Beginn häufig sensorische Symptome, wie Parästhesien oder Dysästhesien im Vordergrund stehen (18,20). Im weiteren Verlauf kann es zur Entstehung von schlaffen Lähmungen und Muskelschwäche kommen und zuletzt ist die Ausbildung einer Spastik möglich. Eine Multiple Sklerose äußert sich meist als partielle Myelitis, was bedeutet, dass eine Symptomatik entsprechend eines kompletten Querschnittssyndroms sehr selten ist. Bei Beteiligung des Rückenmarks lassen sich auch Blasenstörungen und Kontinenzprobleme beobachten. Typische Anzeichen einer Kleinhirnbeteiligung sind Ataxie, Schwindel, Dysmetrie beim Finger-Nase-Versuch und Auffälligkeiten beim Strichgang. Störungen der Augenbewegung, Doppelbilder,

Nystagmus, Oszillopsien (Zucken des Gesichtsfeldes) sowie eine internukleäre Ophthalmoplegie sind hinweisend für Läsionen im Hirnstamm. Bei der internukleären Ophthalmoplegie kann ein Auge nicht adduziert werden, während das abduzierende Auge einen Nystagmus aufweist. Ausgeprägte kortikale Demyelinisierung und globale Hirnatrophie führen zu kognitiver Dysfunktion und chronischer Fatigue (20).

Folgende Tabelle soll eine Übersicht über mögliche Symptome der Multiplen Sklerose geben:

**Tabelle 1: Symptome der Multiplen Sklerose**

Topografie	Symptome	Klinische Zeichen
<b>Großhirn</b>	Kognitive Dysfunktion	<b>früh:</b> Defizite der Aufmerksamkeit, des logischen Denkens und der exekutiven Funktionen <b>spät:</b> ausgeprägte kognitive Dysfunktion in mehreren Domänen
	Depression	
	Fatigue	
	halbseitige motorische und sensible Ausfälle	Pyramidenbahnzeichen
<b>Kleinhirn</b>	Schwindel	
	Gangataxie, Dysmetrie	
	Intentionstremor	
	Dysarthrie	
<b>Nervus Opticus</b>	einseitiger, schmerzhafter Sehverlust	Zentralskotom
	Farbentsättigung	eingeschränktes Farbsehen
		eingeschränkte Sehschärfe
	relativ afferenter Pupillendefekt (RAPD)	
<b>Hirnstamm</b>	Diplopie, Oszillopsien	Nystagmus
		Internukleäre Ophthalmoplegie
	Schwindel	
<b>Rückenmark</b>	Schluck- und Sprechstörungen	
	Hemi-/Mono-/Paraparese	
	Spastik	Pyramidenbahnzeichen
	Schwäche	
	Hypästhesien, Parästhesien	
	Blasendysfunktion	
	Erektile Dysfunktion	
	Dysfunktion des Analsphinkters	

Quelle: in Anlehnung an Compston et al., 2008, S. 1503 (14) und Oh et al., 2018, S. 753

(1)

### 1.3.2 Klinische Phänotypen und Verlaufsformen

Eine Differenzierung der Multiplen Sklerose in verschiedene klinische Phänotypen wurde erstmals 1996 durch das „US National Multiple Sclerosis Society Advisory Committee on Clinical Trials in Multiple Sclerosis“ durchgeführt. Eine genauere Beschreibung des klinischen Verlaufes war vonnöten, um die Kommunikation im klinischen Setting, die Prognoseerstellung, Planung und Rekrutierung im Zuge von klinischen Studien, sowie die Entscheidungsfindung in der Behandlung der MS zu standardisieren (21). Bestimmend für den klinischen Subtyp sind der zeitliche Verlauf des Auftretens der Symptome, sowie die Entwicklung der Symptome, was in weiterer Folge für die Wahl der richtigen Therapie entscheidend ist (20).

Initial wurde die MS in vier verschiedene klinische Phänotypen eingeteilt: die schubförmig remittierende MS (RRMS – relapsing-remitting MS), die sekundär progrediente MS (SPMS – secondary-progressive MS), die primär progrediente MS (PPMS – primary-progressive MS) und die progressiv remittierende MS (PRMS – progressive-relapsing MS) (1). Letztere Verlaufsform wurde jedoch verworfen, da die Bezeichnung als vage empfunden wurde und sich viele Überlappungen zu anderen klinischen Phänotypen ergaben (21). Mit der Wissenszunahme zur Pathologie der MS wurde deutlich, dass eine rein klinische Einteilung in unterschiedliche Phänotypen, ohne Berücksichtigung bildgebender Verfahren oder biologischer Korrelate, nicht ausreichend den zugrundeliegenden pathophysiologischen Prozessen gerecht wurde, weshalb 2014 Revisionen zur Definition der klinischen Verläufe der MS veröffentlicht wurden. Die ursprüngliche Einteilung in RRMS, SPMS und PPMS wurde beibehalten, allerdings wurden jene Phänotypen um die Deskriptoren „Aktivität“ und „Progression“ erweitert und zwei neue klinische Verlaufsformen, das „klinisch isolierte Syndrom“ (CIS – clinically isolated syndrome) und das „radiologisch isolierte Syndrom“ (RIS – radiologically isolated syndrome) definiert (22).

#### 1.3.2.1 Radiologisch isoliertes Syndrom

Der Begriff „radiologisch isoliertes Syndrom (RIS)“ wurde erstmals 2009 von Okuda et. al eingeführt und beschreibt in der Magnetresonanztomographie (MRT) zufällig detektierte, mit einer demyelinisierenden Erkrankung vereinbare Abnormalitäten, welche jedoch ohne klinische Symptome einer Multiplen Sklerose auftreten (23). Das RIS ist aktuell allerdings nicht als eigenständiger Phänotyp der Multiplen Sklerose akzeptiert, da klinische Beweise für eine Multiple Sklerose fehlen und alleinige Veränderungen in der MRT unspezifisch

sein können (21). Okuda et al. konnten nichtsdestotrotz nachweisen, dass ein RIS insbesondere bei bestimmten Konstellationen mit einem erhöhten Risiko für das Auftreten klinischer Symptomatik und/oder einer radiologischen Progression verbunden sein kann und somit mit dem Auftreten einer klinisch manifesten MS (21,23). In einer 2014 veröffentlichten, retrospektiven Multicenter-Studie wurden das 5-Jahres Risiko und Risikofaktoren für das Auftreten eines klinischen Ereignisses in 451 asymptomatischen Personen mit RIS untersucht. 34% der Personen entwickelten innerhalb von 5 Jahren ab der ersten mit dem Vorliegen eines RIS vereinbaren MRT-Untersuchung des Gehirns klinische Ereignisse. Alter < 37 Jahre, männliches Geschlecht und Beteiligung des Rückenmarks stellen die wichtigsten unabhängigen Prädiktoren für das Auftreten von Symptomen dar (24).

### 1.3.2.2 Klinisch isoliertes Syndrom

Unter dem „klinisch isolierten Syndrom“ (CIS) versteht man das erste akute klinische Ereignis, welches Charakteristika einer entzündlichen Demyelinisierung wie bei der Multiplen Sklerose aufweist, jedoch die Kriterien der Dissemination in der Zeit, welche für die Diagnose MS maßgeblich sind, noch nicht erfüllt (21). Häufig manifestiert sich das CIS als Optikus-Neuritis, inkomplette Myelitis oder Hirnstamm-Syndrom (25). Ein derartiges Ereignis dauert analog zur Definition eines MS-Krankheitsschubes mindestens 24 Stunden an und tritt unabhängig von Fieber oder Infektion auf. Ein CIS ist im weiteren Verlauf mit einem erhöhten Risiko des Auftretens einer klinisch manifesten MS verbunden (22). Verschiedene Forschungsgruppen untersuchten die Konversionsraten von CIS zu klinisch manifester MS, wobei Fisniku et al. eine Kohorte von CIS-Patient\*innen über 20 Jahre hinweg beobachteten und eine Konversionsrate von CIS zu klinisch manifester MS von 63% ermitteln konnten (26). Die MRT ist der beste Prädiktor, um eine solche Konversion festzustellen. Studien konnten zeigen, dass CIS-Patient\*innen mit demyelinisierenden Läsionen in der MRT ein 60-80%iges Langzeit-Risiko für die Entwicklung einer klinisch manifesten MS aufweisen, verglichen mit einem 20%igen Langzeit-Risiko bei CIS-Patient\*innen, welche keine Läsionen in der MRT aufweisen (25). Gewisse demografische und bildgebende Charakteristika haben einen Einfluss auf die Konversionsrate von CIS zur klinisch manifesten MS. Die am häufigsten beschriebenen Prädiktoren in diesem Zusammenhang sind das Vorhandensein und die Anzahl von T2-Läsionen in der weißen Substanz in der MRT, sowie ein erhöhter IgG-Index oder das Vorkommen von oligoklonalen Banden im Liquor (22).

### 1.3.2.3 Die schubförmig remittierende Multiple Sklerose (RRMS)

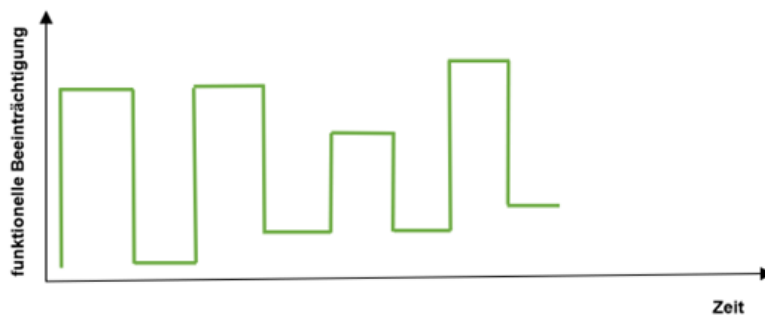
Die schubförmig remittierende Multiple Sklerose präsentiert sich in 85% der Patient\*innen und ist somit der häufigste Phänotyp der MS (25). Sie äußert sich durch wiederkehrende, sogenannte Schübe, welche Phasen neurologischer Symptome darstellen, auf welche Phasen relativer klinischer Stabilität folgen, auch als Remission bezeichnet, in denen sich die neurologischen Defizite vollständig oder unvollständig zurückbilden können (20,22). In den Remissionsphasen kommt es allerdings zu keiner Verschlechterung der neurologischen Funktionsstörungen, die Folgen der Schübe können lediglich stabil bleibend fortbestehen (27).

Ein Schub ist definiert als Episode neuer oder wiederkehrender neurologischer Symptome, welche von Patient\*innen angegeben werden und idealerweise auch auf objektiven Befunden von Untersucher\*innen basieren. Diese Symptome und/oder Zeichen sollten unabhängig von Fieber oder Infektion auftreten und mindestens 24 Stunden andauern (25,28). Im Schnitt erlitten MS-Patient\*innen in historischen Kohorten 1,5 Schübe pro Jahr (14). Die Schubfrequenz ist heutzutage allerdings deutlich geringer, sowohl in „natürlichen“ als auch in behandelten Kohorten (29). Die Häufigkeit der Schübe, sowie das Ausmaß der entzündlichen Prozesse nehmen zudem mit zunehmendem Alter ab. Auch während der Schwangerschaft lässt sich eine verringerte Schubfrequenz beobachten, was vermutlich mit dem natürlich gegebenen immuntoleranten Effekt einer Schwangerschaft zusammenhängt. Zu beachten ist, dass in den ersten 3 Monaten postpartal mit einer erhöhten Schubfrequenz zu rechnen ist (22).

Zu Beginn der Erkrankung erfolgt die Erholung von Schüben meist vollständig. Mit der Zeit kommt es allerdings zur unvollständigen Erholung und zu einer Häufung bleibender Defizite, da Schübe bei fast der Hälfte der Episoden Restdefizite herbeiführen (4,22). Schließlich kann eine RRMS in eine sekundär progrediente Multiple Sklerose übergehen, wobei diese Progression bei bis zu 80% der unbehandelten (oder mit zum damaligen Studienzeitpunkt existierenden, schwach wirksamen Präparaten behandelten) Patient\*innen beschrieben wird und im Median nach 10 – 15 Jahren ab Erstdiagnose stattfindet (4,20).

„Pseudoschübe“, auch als „Uhthoff-Phänomen“ bekannt, sind als wichtige Differentialdiagnose von Schüben abzugrenzen, da diese keine tatsächliche inflammatorische Aktivität anzeigen, sondern lediglich durch erhöhte Körpertemperatur

bedingt sind und keiner Therapie bedürfen, da sie bei Abkühlung selbstlimitierend verlaufen (20).

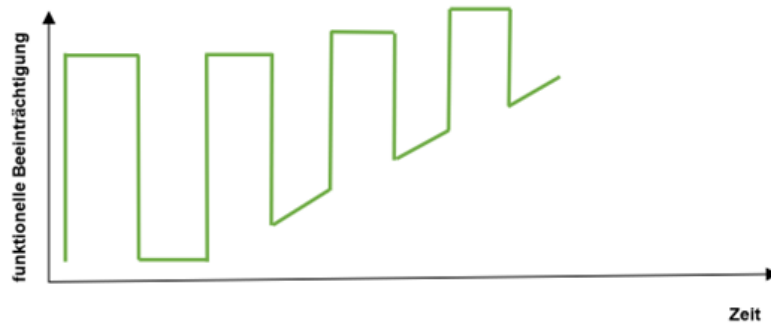


**Abbildung 1: Beispielhafter Verlauf einer schubförmig remittierenden MS (RRMS)**

Quelle: eigene Abbildung in Anlehnung an Klineova et al., 2018, S. 5 (22)

#### 1.3.2.4 Die sekundär progrediente Multiple Sklerose (SPMS)

Die sekundär progrediente Multiple Sklerose entsteht aus der RRMS und ist charakterisiert durch eine stetige Zunahme bleibender neurologischer Defizite. Der Begriff „Progression“ bezeichnet die kontinuierliche Verschlechterung der neurologischen Beeinträchtigung über mindestens 6-12 Monate. Vom Beginn der schubförmigen MS bis zur sekundären Progression vergehen im Mittel bis zu 19 Jahre (27). Die Diagnose SPMS wird meist anhand der Anamnese einer graduellen Verschlechterung retrospektiv gestellt, wobei dies oft erst Jahre nach dem tatsächlichen Beginn der Progression erfolgt (21,22). Bis dato gibt es keine eindeutigen klinischen, bildgebenden, immunologischen oder pathologischen Kriterien, um den Übergangspunkt (die Transition) von RRMS in SPMS festzumachen (21). Allerdings ließen sich einige Prädiktoren identifizieren, welche mit einer früheren Progression von RRMS zu SPMS in Zusammenhang stehen. Dazu zählen höheres Alter bei Krankheitsbeginn, männliches Geschlecht, rückenmarks-assoziierte Symptome und inkomplette Erholung von einem Schub (27). Aktuell gibt es keinen allgemeinen Konsens, was die Mechanismen, welche einer Progression zugrunde liegen, betrifft. Epidemiologische und bildgebende Studien deuten darauf hin, dass in einer progressiven Phase neurodegenerative Prozesse überwiegen und inflammatorische Mechanismen in den Hintergrund treten (27).

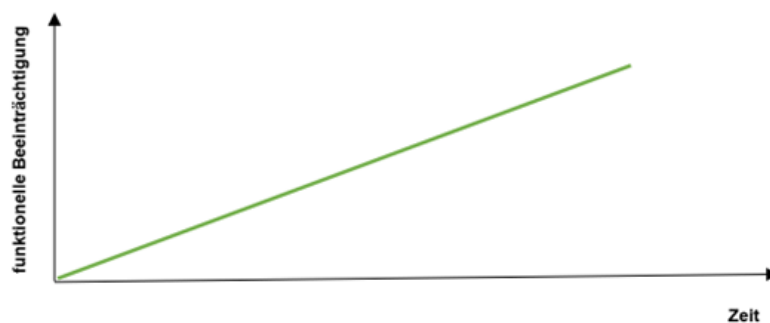


**Abbildung 2: Beispielhafter Verlauf einer sekundär progredienten MS (SPMS)**

Quelle: eigene Abbildung in Anlehnung an Klineova et al., 2018, S. 7 (22)

### 1.3.2.5 Die primär progrediente Multiple Sklerose (PPMS)

Bei ca. 15% der MS-Patient\*innen manifestiert sich die Erkrankung als primär progrediente MS, welche sich durch das Fehlen eines initialen schubförmigen Verlaufs und von Beginn an fortschreitender Zunahme neurologischer Beeinträchtigungen äußert (4,22). Der Verlauf der primären Progression ist interindividuell verschieden und kann auch selten überlagerte Schübe, sowie Phasen relativer Krankheitsstabilität umfassen (22). In der Regel präsentiert sich die PPMS als progressive Myelopathie und daraus resultierend einer spastischen Paraparese (4,25). Typisch für die PPMS ist eine Erstmanifestation im höheren Alter, häufig erst in der fünften bis sechsten Lebensdekade, sowie ein größerer Anteil an männlichen Betroffenen (20,25).



**Abbildung 3: Beispielhafter Verlauf einer primär progredienten MS (PPMS)**

Quelle: eigene Abbildung in Anlehnung an Klineova et al., 2018, S. 8 (22)

### 1.3.2.6 Aktivität und Progression

Bis zur 2014 veröffentlichten Revision zur Definition der klinischen Verläufe der MS wurden MS-Phänotypen alleinig anhand des aktuellen neurologischen Status und der Anamnese in zwei Kategorien unterteilt: schubförmig und progredient. Diese Einteilung ist

allerdings eher statisch und gibt keinerlei zeitliche Informationen über den laufenden Krankheitsprozess bzw. die Krankheitsdynamik. Deshalb schlugen Lublin et al. vor, die Phänotypen der MS um die beiden Deskriptoren „Aktivität“ und „Progression“ zu erweitern (21,22). Hinweise auf Krankheitsaktivität und klinische Progression, welche nach heutigem Verständnis laufende entzündliche oder neurodegenerative Prozesse widerspiegeln, können sich auf Prognose, therapeutische Entscheidungen und die Gestaltung und Ergebnisse klinischer Studien auswirken. Unter „Aktivität“ versteht man das Auftreten klinischer Schübe oder Aktivität in der MRT (Kontrastmittel-aufnehmende Läsionen oder neue/sich vergrößernde T2-Läsionen), „Progression“ beschreibt eine von Schüben unabhängige, klinisch nachgewiesene Zunahme der Beeinträchtigung, bezogen auf einen definierten Zeitraum (21).

## 1.4 Diagnostik der Multiplen Sklerose

Trotz jüngster Fortschritte in der Diagnostik, besonders hinsichtlich unterstützender radiologischer Modalitäten, ist es wichtig zu betonen, dass die Multiple Sklerose primär klinisch determiniert wird (25). Die Diagnose wird anhand einer Zusammenschau aus Anamnese, Klinik, Befunden aus bildgebenden Verfahren und Laboruntersuchungen gestellt, wobei die 2017 revidierten McDonald Kriterien standardmäßig für die Diagnosestellung herangezogen werden (28,30). Es wird klinische Expertise benötigt, um eine für die Diagnose notwendige örtliche und zeitliche Dissemination nachzuweisen und andere Erkrankungen differentialdiagnostisch auszuschließen (31).

Unter örtlicher Dissemination (DIS – dissemination in space) versteht man das Vorkommen von mehr als einer Läsion, welche für die MS charakteristisch ist, in zwei oder mehr von vier ZNS-Regionen (periventrikulär, kortikal oder juxtakortikal, infratentoriell, spinal) (28). Diese Läsionen können entweder durch mehrere klinische Ereignisse, welche auf verschiedene ZNS-Regionen zurückzuführen sind, oder durch mehrere T2-hyperintense Läsionen in der MRT oder beides zugleich nachgewiesen werden (18). Die zeitliche Dissemination (DIT- dissemination in time) wird als Entwicklung von MS-Läsionen im zeitlichen Verlauf definiert (28). Die zeitliche Dissemination kann mittels MRT durch gleichzeitiges Vorhandensein von Gadolinium-anreichernden (akuten) und nicht-anreichernden (chronischen) Läsionen oder durch das Auftreten neuer T2-Läsionen in einer Folge-MRT-Untersuchung nachgewiesen werden. Alternativ lässt sich die DIT anhand multipler klinischer Attacken definieren (18). Zusätzlich kann der Nachweis von oligoklonalen Banden (Immunglobulinen) im Liquor bei Patient\*innen mit klinisch

isoliertem Syndrom ebenfalls das Kriterium der DIT erfüllen, da dies bei gleichzeitigem Fehlen derartiger Immunglobuline im Serum eine intrathekale Antikörper-Bildung nahelegt (18,30).

Sind die 2017 revidierten McDonald-Kriterien erfüllt und es lässt sich keine bessere Erklärung für die Symptomatik finden, so kann die Diagnose Multiple Sklerose gestellt werden (28). Dementsprechend ist es wichtig differentialdiagnostisch an Erkrankungen zu denken, welche die Multiple Sklerose nachahmen können - sogenannte „Mimicker“ (20). Eine atypische Präsentation der MS, sowie das Vorliegen von Red Flags, kombiniert mit unspezifischen oder atypischen Veränderungen der weißen Substanz, führen oft zu Falschdiagnosen durch „Mimicker“. Zu den Red Flags zählen u.a. das Auftreten von systemischen Symptomen, sowie neurologische Zeichen, welche außerhalb des ZNS auftreten, wie beispielsweise eine periphere Neuropathie (30). Differentialdiagnostisch besonders nennenswert sind die Neuromyelitis-optica-Spektrum-Erkrankungen (NMOSDs), da diese mit der MS verwechselt werden können und gewisse verlaufsmodifizierende Therapien der MS bei NMOSDs kontraindiziert und mit schlechterem Behandlungsergebnis assoziiert sind (28). Hierbei handelt es sich um relativ seltene, entzündliche, demyelinisierende Erkrankungen des ZNS, wobei Patient\*innen klassischerweise eine akute Optikusneuritis und eine längsverlaufende, mehr als drei Wirbelsegmente umfassende, transversale Myelitis aufweisen. Für die Diagnose der NMOSDs hochspezifisch sind Autoantikörper, welche auf Aquaporin-4 abzielen (AQP4-IgG) (25). Analoge Überlegungen treffen auf Myelin-Oligodendrozyten-Glykoprotein (MOG)-Antikörper-assoziierte Erkrankungen (MOGAD= MOG antibody-associated diseases) zu. Dabei handelt es sich um Erkrankungen mit ZNS-Demyelinisierung und dem Vorhandensein von Serumantikörpern gegen MOG (32).

Die von Kurtzke entwickelte EDSS (Expanded Disability Status Scale) spielt in der Diagnostik und in der Verlaufskontrolle der MS als Instrument zur klinischen Beurteilung weiterhin eine bedeutende Rolle. Die EDSS stellt das am weitesten verbreitete Scoring-Tool dar und ermöglicht studienübergreifende Vergleiche durch seine breite Anwendung als primärer Endpunkt in klinischen Studien (33). Mittels EDSS kann der Schweregrad der Behinderung bei MS anhand der Gehstrecke und der Beurteilung von acht Funktionssystemen ermittelt werden. Die acht Funktionssysteme umfassen Pyramidenbahn, Kleinhirn, Hirnstamm, Sensorium, zerebrale Funktion, Sehstörungen, Blasen- und Mastdarmstörungen, sowie sonstige Funktionen, wobei jedes Funktionssystem

getrennt zu bewerten ist. Es handelt sich um eine Ordinalskala mit Werten von 0 (keine neurologischen Defizite) bis 10 (Tod durch MS) (34).

#### 1.4.1 Diagnosekriterien nach McDonald

Die aktuellen McDonald-Kriterien stammen aus dem Jahr 2017. Die McDonald-Kriterien von 2010 wurden revidiert, da neue Forschungsergebnisse berücksichtigt werden mussten und besonders Fortschritte im Bereich der bildgebenden Diagnostik eine frühere, sensitivere und spezifischere Diagnostik ermöglichen. Die revidierten Kriterien von 2017 heben zudem die diagnostische Bedeutung Liquor-spezifischer oligoklonaler Banden in Bezug auf die zeitliche Dissemination hervor und beziehen auch kortikale und symptomatische Läsionen in die Erfüllung der örtlichen Dissemination ein (28). Zu unterscheiden sind Diagnosekriterien für eine schubförmig remittierende MS und eine primär-progrediente MS (18).

**Tabelle 2: McDonald-Kriterien von 2017 für die schubförmig remittierende MS**

<b>McDonald-Kriterien für die schubförmig remittierende MS</b>		
<b>Schübe (Klinik)</b>	<b>Anzahl der klinisch objektivierbaren Läsionen</b>	<b>Zusätzlich erforderliche Befunde</b>
≥ 2	≥ 2	keine
	1 (sowie eindeutige historische Hinweise in der Anamnese auf einen früheren Schub mit einer Läsion an einer der definierten anatomischen Lokalisationen)	keine
	1	<u>örtliche Dissemination</u> nachgewiesen anhand <ul style="list-style-type: none"> <li>• der MRT</li> <li><b>ODER</b></li> <li>• eines zusätzlichen Schubes eine andere ZNS-Lokalisation betreffend</li> </ul>
1	≥ 2	<u>zeitliche Dissemination</u> nachgewiesen anhand <ul style="list-style-type: none"> <li>• der MRT</li> <li><b>ODER</b></li> <li>• Liquor-spezifischer oligoklonaler Banden</li> <li><b>ODER</b></li> <li>• eines zusätzlichen Schubes</li> </ul>
1	1	<u>örtliche UND zeitliche Dissemination</u> nachgewiesen

Quelle: Thompson et al., 2018, S. 167 (28)

**Tabelle 3: McDonald-Kriterien von 2017 für die primär progrediente MS**

<b>McDonald-Kriterien für die primär progrediente MS</b>
<ul style="list-style-type: none"><li>• <math>\geq 1</math> Jahr Fortschreiten der klinischen Funktionseinschränkung/Behinderung, unabhängig von Schüben</li></ul> <p><b>PLUS</b> zwei der folgenden drei Kriterien</p> <ul style="list-style-type: none"><li>• <math>\geq 1</math> T2-hyperintense Läsion(en), für die MS charakteristisch, in mind. einem der folgenden Areale<ul style="list-style-type: none"><li>○ periventrikulär</li><li>○ kortikal/ juxtakortikal</li><li>○ infratentoriell</li></ul></li><li>• <math>\geq 2</math> spinale T2-hyperintense Läsionen</li><li>• Vorliegen Liquor-spezifischer oligoklonaler Banden</li></ul>

Quelle: Thompson et al., 2018, S. 169 (28)

#### 1.4.2 Die Rolle der Magnetresonanztomographie (MRT)

Die Einführung der Magnetresonanztomographie (MRT) in den frühen 1980er Jahren revolutionierte Diagnose und Management der MS durch erstmalige in vivo Visualisierung der Läsionsaktivität bzw. -last. Durch stetige technologische Weiterentwicklungen wurde die MRT über die letzten Jahrzehnte hinweg rasch zum wichtigsten paraklinischen Diagnose- und Monitoringwerkzeug der MS (35). Bei Patient\*innen mit CIS spielt die Berücksichtigung der MRT in der diagnostischen Abklärung und Einleitung einer passenden Therapie eine wichtige Rolle, bei bereits diagnostizierter MS wird die MRT für die Überwachung der Krankheitsentwicklung herangezogen und in klinischen Studien liefert die MRT anerkannte Maßstäbe für die Wirksamkeit von Behandlungen (36).

Die meisten, wenn nicht alle Patient\*innen benötigen eine MRT-gestützte Diagnostik, um die Diagnose „Multiple Sklerose“ anhand der McDonald-Kriterien stellen zu können (20). Durch die MRT wird eine Diagnose auch deutlich früher möglich, wobei eine kürzere Zeit von CIS zu MS-Diagnose von Tintore et al. beobachtet wurde und der Anteil von Patient\*innen mit CIS wesentlich zugunsten der Patient\*innen mit MS nach McDonald-Kriterien geringer wird (37). Die standardmäßig weit verbreiteten MRT-Protokolle zur Diagnose der MS inkludieren vor allem T2-gewichtete 2D-FLAIR-Sequenzen mit axialer Schichtung und einer Schichtdicke von 3 mm, sowie T1-gewichtete Kontrastpulssequenzen mit Gadolinium, welche bei Magnetfeldstärken von 1,5 T (Tesla) sowohl von Gehirn als auch Rückenmark akquiriert werden. Fortschrittlichere MRT mit Feldstärken von 3T und 7T ermöglichen ein besseres Signal-Rausch-Verhältnis und eine verbesserte räumliche

Auflösung, allerdings erfolgt dies bei einer Feldstärke von 7T auf Kosten vermehrter Artefakte, mangelnder Standardisierung und höherer Kosten (3,35).

Typische MRT-Befunde der MS sind Läsionen in Regionen weißer Substanz, welche in T2-gewichteten FLAIR-Sequenzen hyperintens und T1-gewichteten Bildern partiell hypointens erscheinen. Es handelt sich hierbei um Areale der Demyelinisierung und Gliose, sowie Entzündung und damit gekoppelter Wassereinlagerung (20). T2-Läsionen weisen typischerweise eine ovale/ovoide Form auf, messen häufig mehr als 5mm im Durchmesser und neigen dazu, sich um zentripetale parenchymale Venen und Venolen zu bilden, wodurch sie gehäuft in bestimmten Arealen des ZNS auftreten. Diese Beobachtung spiegelt sich in den aktuellen Diagnosekriterien der MS wider, welche vier Bereiche des ZNS als Prädilektionsstellen für das Auftreten von T2-hyperintensiven MS-Läsionen beschreiben – die periventrikuläre weiße Substanz, kortikale/juxtakortikale und infratentorielle Hirnregionen, sowie das Rückenmark (28,35).

Folgende Abbildungen von Patient\*innen des MS-Zentrums Graz, untersucht am 3T-MRT-Forschungsgerät an der Abteilung für Neuroradiologie des Universitätsklinikums Graz, sollen Läsionen in drei der vier genannten Hirnregionen beispielhaft darstellen:

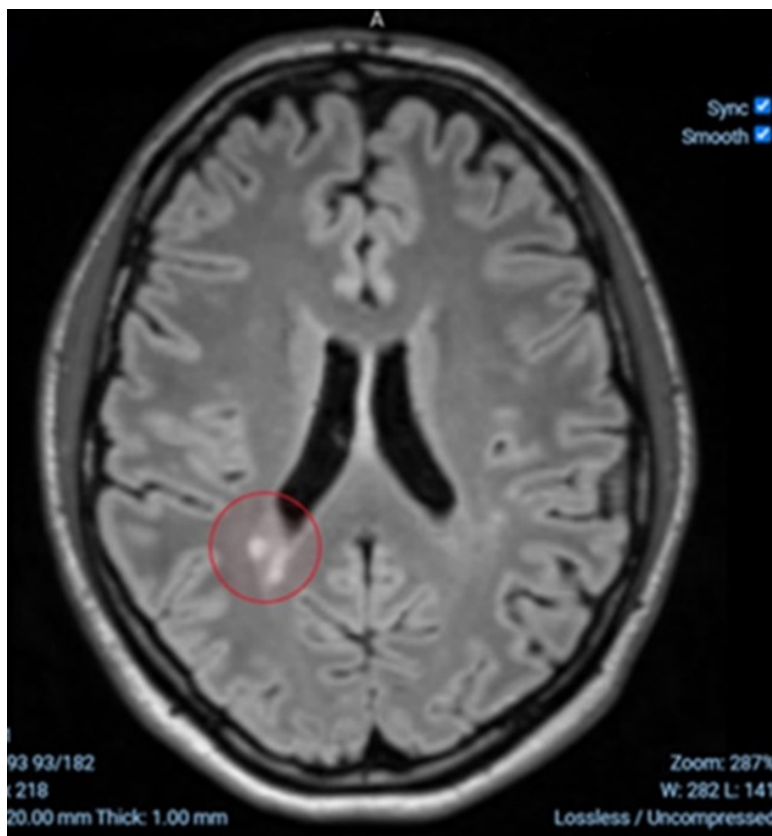


Abbildung 4: Periventrikuläre T2-hyperintense Läsionen, FLAIR-Sequenz

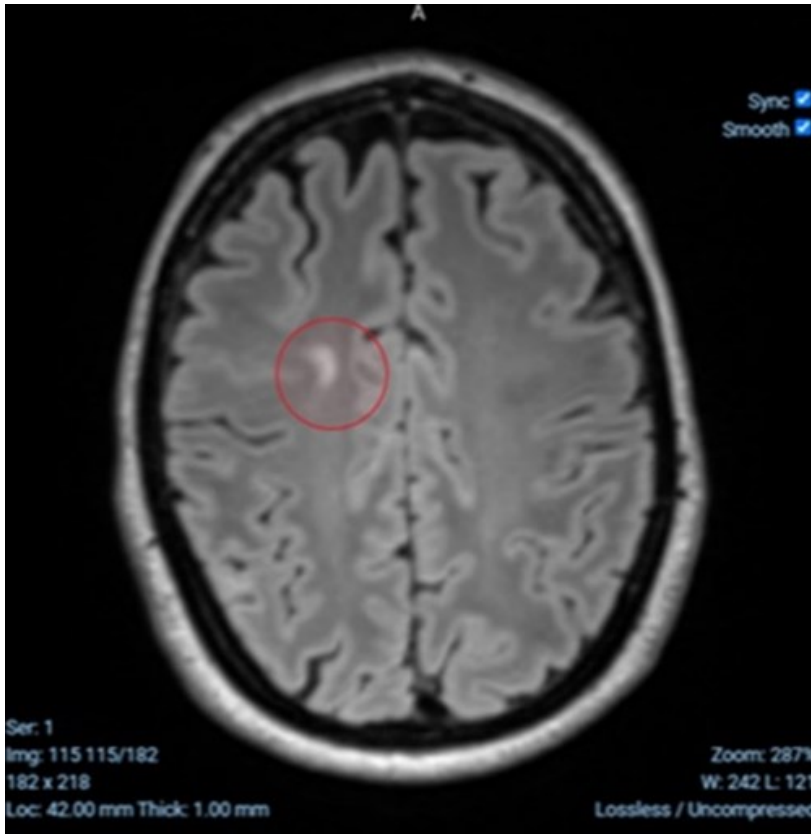


Abbildung 5: Juxtakortikale T2-hyperintense Läsion, FLAIR-Sequenz



Abbildung 6: Infratentorielle T2-hyperintense Läsion, FLAIR-Sequenz

In der Routineuntersuchung der MS häufig verwendete T1-gewichtete Pulssequenzen sind das Spin-Echo (T1SE) und das Gradienten-Echo (T1GE). T1-gewichtete Pulssequenzen bieten eine ausgezeichnete Strukturdefinition, wie beispielsweise den Kontrast zwischen fettreichen Strukturen (Myelin), welche hell erscheinen, und wasserdominierten Strukturen (Cortex), welche dunkel erscheinen. Demyelinisierung und Axonverlust vermindern den Fettgehalt der axonalen Strukturen und erhöhen deren Wassergehalt, weshalb MS-Läsionen in T1-Sequenzen zum Teil hypointens erscheinen (35). Bereiche in einer T1-gewichteten Aufnahme, welche im Vergleich zur weißen Substanz hypointens erscheinen, werden als „black holes“ bezeichnet (38,39). Sie stimmen mit hyperintensiven T2-Läsionen überein, wobei nur etwa 20-30% der akuten hyperintensiven T2-Läsionen auch in T1-gewichteten Spin-Echo-Aufnahmen darstellbar sind (40).

T1SE und T1GE werden herangezogen, um kontrastmittelanreichernde Läsionen nach Gadolinium-Verabreichung darzustellen. In einer akut entzündlichen Phase der MS kommt es zu lokalisierter ZNS-Inflammation mit Zusammenbruch der Blut-Hirn-Schranke, weshalb Gadolinium-Kontrastmittel in das umliegende Parenchym eindringt. Somit ist das Vorhandensein Gadolinium-anreichernder T1-Läsionen ein Marker für akute Krankheitsaktivität und diese Läsionen sind oftmals mit einem Schub assoziiert, obwohl kontrastmittelanreichernde Läsionen fünf bis zehn Mal häufiger auftreten als klinische Schübe, oft klinisch stumm ablaufen und nur schwach mit Funktionseinschränkungen korrelieren (35).

2021 wurden die internationalen MAGNIMS-CMSC-NAIMS-Konsensusempfehlungen, welche aus den MAGNIMS (Magnetic Resonance Imaging in Multiple Sclerosis) Guidelines von 2015, den CMSC (Consortium of Multiple Sclerosis Centers) Guidelines von 2016 und Empfehlungen der NAIMS (North American Imaging in Multiple Sclerosis Cooperative) entstanden sind, veröffentlicht. Diese Empfehlungen enthalten aktualisierte Leitlinien dazu, wie und wann die MRT zur Diagnose, Prognose und Therapieüberwachung der MS eingesetzt werden sollte, wobei der Schwerpunkt auf der Verwendung standardisierter MRT-Protokolle, dem vernünftigen Einsatz gadoliniumhaltiger Kontrastmittel und einer standardisierten Bewertung liegt (3).

Die standardisierten MRT-Protokolle aus den Konsensusempfehlungen von 2021 enthalten eine Reihe wichtiger Parameter, welche berücksichtigt werden sollten, um vergleichbare MRT-Bilder bei MS-Patient\*innen weltweit erzeugen zu können. Dazu zählen verwendete

Feldstärke, Schichtdicke, Bildauflösung, der durch die MRT abzudeckende Bereich des ZNS, sowie die axiale Scanausrichtung. Außerdem werden für Gehirn und Rückenmark je nach Fragestellung passende Pulssequenzen empfohlen, welche dem jeweiligen Protokoll aus den Konsensusempfehlungen von Wattjes et al. zu entnehmen sind (3).

Die Empfehlungen zur Verwendung der MRT zur Diagnosestellung der MS sehen bei CIS-Patient\*innen alle 6-12 Monate eine Schädel-MRT vor, um die zeitliche Dissemination und damit Konversion der Diagnose in eine Multiple Sklerose erkennen zu können. Die MRT ist allerdings nicht nur von diagnostischer Relevanz. Mittels der MRT kann auch die Wirksamkeit der Therapie und damit verbunden die Krankheitsaktivität überprüft werden. Laut MAGNIMS-CMSC-NAIMS-Konsensusempfehlungen soll vor Beginn einer verlaufsmodifizierenden Therapie eine Baseline-Schädel-MRT erstellt werden. 3-6 Monate nach Behandlungsbeginn soll eine erneute „Baseline“-Schädel-MRT durchgeführt werden. Dies ist insofern sinnvoll, als dass verlaufsmodifizierende Therapeutika einen verzögerten Wirkeintritt zeigen und dadurch die Fehlinterpretation von neuen Läsionen, die vor Wirkungseintritt entstanden sind, im Sinne einer „echten“ Progression unter Therapie (d.h. Indiz für Therapieversagen) vermieden wird. Die erneute Baseline-MRT soll ohne Gadolinium akquiriert werden, da die Krankheitsaktivität anhand neuer T2-Läsionen eruiert werden kann. Bei der prä-therapeutischen Baseline-MRT wird eine Gadolinium-Applikation empfohlen. Bei Patient\*innen unter verlaufsmodifizierender Therapie soll in weiterer Folge jährlich eine Schädel-MRT durchgeführt werden. Die MRTs sollen unter Befolgung der standardisierten Protokolle erstellt werden, um eine identische Positionierung der Schichten, Impulssequenzen, Magnetfeldstärken und räumliche Auflösung zu gewährleisten. Üblicherweise sind abgekürzte Protokolle mit T2-gewichteten 2D- oder 3D-FLAIR-Sequenzen ausreichend, da man anhand dieser neue fokale T2-Läsionen sehr gut erkennen kann und neue oder sich vergrößernde T2-Läsionen ein zuverlässiger Marker von aktiver entzündlicher Erkrankung sind, wenn eine technisch vergleichbare frühere und aktuelle (d.h. innerhalb von  $\leq 1$  Jahr durchgeführte) MRT-Untersuchung verfügbar ist (3).

## 1.5 Segmentierungstechniken zur Läsionsquantifizierung in der MRT

Unter Segmentierung versteht man die Identifizierung und Abgrenzung ausgewählter Merkmale eines Bildes. Die älteste und nach wie vor gebräuchlichste Methode zur

Läsionssegmentierung in einem MRT-Bild bei Multipler Sklerose ist die manuelle Verfolgung und Einzeichnung von Läsionsumrissen mit einer Maus zur Steuerung eines Cursors auf dem Computerbildschirm (41). Der Bediener muss die Grenze zwischen normalem Gewebe und Läsion anhand von Signalintensitätsunterschieden erkennen (42). Dies ermöglicht die Schätzung der Gesamtläsionslast (43). Aufgrund der großen Menge der zu analysierenden MRT-Daten ist die manuelle Segmentierung von Läsionen durch klinische Expert\*innen jedoch ein sehr zeitaufwändiger Prozess, welcher große Expertise erfordert und zudem subjektiv und folglich anfällig für menschliche Fehler ist (43–45). Zudem nimmt der Arbeitsaufwand mit der technischen Weiterentwicklung, durch welche immer besserer Kontrast und Auflösung, sowie dünnere Schichtdicken möglich werden, stetig zu (3,35,43–45). In mehreren Studien wurde die große Variabilität bei der manuellen Segmentierung durch geschultes Personal untersucht, wobei Volumenunterschiede der segmentierten Läsionen zwischen 10% und 68% festgestellt wurden (43). Infolgedessen steht die Entwicklung von halbautomatischen oder automatischen Methoden zur Läsionssegmentierung seit vielen Jahren im Fokus bildgebender Forschung (41,43). Ziel ist es, Methoden zu entwickeln, welche im klinischen Alltag reproduzierbar, zuverlässig und zeitsparend eingesetzt werden können (42,43).

Computergestützte Diagnosesysteme (CAD systems = Computer-Aided Diagnosis systems), wozu auch halbautomatische und automatische Segmentierungsmethoden zählen, folgen einem grundlegenden Prinzip hinsichtlich der Bildverarbeitung, welches im folgenden Blockschema beschrieben wird:

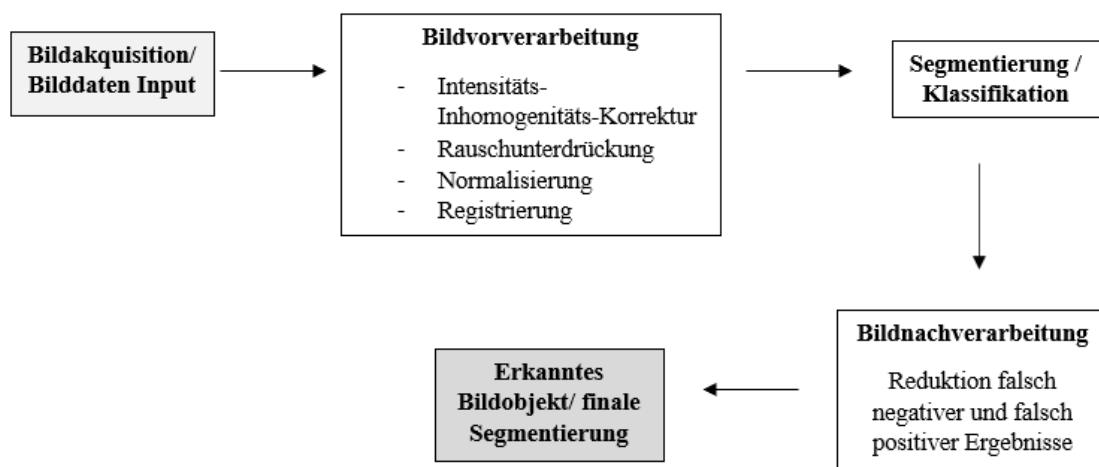


Abbildung 7: Blockschema eines typischen CAD-Systems zur Läsionssegmentierung

Quellen: eigene Abbildung in Anlehnung an Danelakis et al., 2018, S. 88 (44), García-Lorenzo et al., 2013, S. 5 (43), Handels, 2009, S. 50 (46), Mortazavi et al., 2012, S. 303 (45)

García-Lorenzo et al. bezeichnen die Gesamtheit der Bildverarbeitungsschritte, welche erforderlich sind, um Läsionen aus rohen Bilddaten zu segmentieren, auch als Segmentierungspipeline. Es lässt sich erkennen, dass diese Pipeline in zwei große Arbeitsschritte unterteilbar ist: Vorverarbeitung und Segmentierungsmethode selbst (43).

### 1.5.1 Bildvorverarbeitung

Ziel der Vorverarbeitung ist die Bilddatenqualität für die darauffolgende Segmentierung zu verbessern (46). Die Vorverarbeitung ist ein wichtiger und umfangreicher Prozess, wobei folgende Unterpunkte beispielhaft erwähnt werden sollen:

- **Registrierung:** Goshtasby et al. beschreiben die Bild-Registrierung als Prozess, bei dem die Übereinstimmung zwischen allen Punkten in zwei oder mehr Bildern derselben Szene ermittelt wird (47). Die Registrierung hat den Sinn, Bilddaten, welche beispielweise zu unterschiedlichen Zeitpunkten, anhand verschiedener bildgebender Geräte oder Messmethoden akquiriert wurden, anzupassen und in einem Koordinatensystem auszurichten. So können die Bilddaten unterschiedlicher Modalitäten in einem gemeinsamen Koordinatensystem abgebildet und fusioniert werden (46). Durch die Registrierung von Bildern verschiedener Schichten in der MRT können außerdem Fehlansichtungen, welche durch Bewegung der Patient\*innen während der Bildaufnahme entstanden sind, ausgeglichen werden (45).
- **Intensitäts-Inhomogenitäts-Korrektur:** Die Intensität desselben Gewebes kann aufgrund der Kopfspulen, welche beim Signalempfang nicht in allen Bereichen gleich empfindlich sind, über das Bild hinweg variieren (48). Dieses Phänomen kann bei intensitäts-basierten Segmentierungstechniken zu Fehlklassifikationen führen, weshalb Korrekturmethode angewandt werden müssen, um diese Intensitätsschwankungen zu reduzieren (45).
- **Rauschunterdrückung:** Der Aufnahmeprozess kann ein gewisses Bildrauschen erzeugen (43). Unter Bildrauschen versteht man die lokale Variation der Bildfunktionswerte. Um die Bildfunktionswerte zu vereinheitlichen und so das Bildrauschen zu minimieren, können Glättungsfiler angewandt werden (46). Die

korrekte Auswahl eines Filters ist hier entscheidend. Tiefpassfilter können Bildrauschen verringern, führen aber zugleich zu einem unschärferen und verschwommenen Bildergebnis, was als Filtermethode für medizinische Bilddaten nicht akzeptabel ist (49). Anisotrope 3D Diffusionsfilter stellen die bessere Wahl dar, da sie Bildrauschen verringern, ohne morphologische Bilddetails zu trüben (45).

- Intensitäts-Normalisierung: Manche Segmentierungsmethoden sind darauf angewiesen, dass die Intensität der segmentierten Bilddaten der Intensität der Trainingsbilder ähnelt, weshalb ein Intensitäts-Normalisierungsschritt notwendig wird. Hierbei wird der Intensitätsbereich des Zielbildes verändert und auf einen vordefinierten Intensitätsbereich abgebildet (43).

### 1.5.2 Läsionssegmentierung

Im Anschluss an die Vorverarbeitung kommt es zur eigentlichen Bildsegmentierung, wo es zur algorithmischen Abgrenzung von gesuchten Objekten im Bild kommt (46). Die Zahl der verschiedenen Segmentierungsmethoden ist groß und je nach Literatur lassen sie sich in unterschiedliche Kategorien unterteilen, welche in dieser Arbeit nur oberflächlich beschrieben werden (45).

García-Lorenzo et al. wiesen darauf hin, dass die Vermischung von semi-automatischen und automatischen Segmentierungsmethoden im Paper von Mortazavi et al. als Limitation zu sehen ist. Nichtsdestotrotz wurden für diese Diplomarbeit die Kategorien von Mortazavi et al. zur Kategorisierung der Segmentierungstechniken übernommen, da diese zur Übersicht am angemessensten erschienen (43).

Mortazavi et al. schlagen folgende vier Hauptkategorien zur Unterteilung der Segmentierungsmethoden vor: datengesteuerte, statistische, intelligente und deformierbare Modelle. Wenn eine Methode mit Schwellenwertverfahren, Regionenwachstumsverfahren und anderen räumlichen Ansätzen arbeitet, wird sie der datengesteuerten Kategorie zugeordnet. Handelt es sich um eine Methode zur Schätzung von Wahrscheinlichkeitsdichtefunktionen, wird sie in die statistische Gruppe eingeordnet. Intelligente Modelle arbeiten laut Mortazavi et al. mit künstlichen neuronalen Netzwerken. Methoden zur Volumenschätzung zählen zu der Kategorie der deformierbaren Modelle (45). Zum besseren Verständnis wird im Folgenden jede Kategorie anhand eines Beispiels erläutert.

Datengesteuerte Verfahren erwiesen sich als die ersten Techniken zur Segmentierung von MRT-Bildern, wobei Schwellenwertverfahren die Hauptkategorie bilden (45). Schwellenwertverfahren zählen zu den intensitätsbasierten Segmentierungsverfahren (41). Wicks et al. entwickelten ein halbautomatisiertes, intensitätsbasiertes Schwellenwertverfahren, welches Interaktion des Bedieners erfordert, da die beiden Schwellenwerte für dieses Verfahren manuell ausgewählt werden müssen und zum Schluss eine manuelle Überprüfung erforderlich ist. Zuerst muss ein Schwellenwert identifiziert werden, um das Gehirn vom umgebenden Gewebe abzugrenzen und letzteres vom Bild zu entfernen. Der zweite Schwellenwert soll auf einem optimalen Level gewählt werden, um Läsionen zu detektieren. Alle Regionen, welche Intensitäten über dem ausgewählten Schwellenwert besitzen, werden als Läsion identifiziert (50). Schließlich wurden die segmentierten Bilder von Expert\*innen validiert, wobei die Begrenzung schlecht abgegrenzter Läsionen geändert wurde, nicht erkannte Läsionen umzeichnet wurden und falsch klassifiziertes Gewebe gelöscht wurde (41). Filippi et al. versuchten in einer 1996 veröffentlichten Studie zu ermitteln, wie sich minimale Änderungen in der Auswahl des Schwellenwertes auf die berechneten Läsionsvolumina auswirken. Eine Erhöhung des Schwellenwertes um 1%, führte zu einer mittleren Abnahme der berechneten Läsionsvolumina um 15%. Umgekehrt kam es bei einer Verringerung des Schwellenwertes um 1% zu einem mittleren Anstieg der berechneten Läsionsvolumina um 15%. Somit konnten Filippi et al. zeigen, dass die subjektive Wahl eines Schwellenwertes große Auswirkungen auf die Messung des Läsionsvolumens hat und schlugen vor, Guidelines für die Schwellenwert-Wahl einzuführen oder zuverlässig arbeitende automatisierte Segmentierungsmethoden zu entwickeln (42).

Als Beispiel statistischer Verfahren ist eine automatische Segmentierungsmethode für Läsionen der weißen Substanz zu nennen, welche von Anbeek et al. präsentiert wurde (45). Bei diesem Modell findet nach Bildvorverarbeitung ein sogenannter k-Nächste-Nachbarn-Klassifikator (k-NN-Klassifikator), zur Bestimmung der Läsionswahrscheinlichkeit pro Voxel, Anwendung (45). Die Klassifikation erfolgt anhand der nächstgelegenen Nachbarpunkte eines Datenpunktes in einem Merkmalsraum. Stichproben werden dabei abhängig von ihren Merkmalen klassifiziert, wobei jedes Bildvoxel eines MRT-Bildes als eigene Stichprobe behandelt wird. Nach Definition eines Merkmalsraumes (Anbeek et al. verwendeten Voxel-Intensitäten und die räumliche Anordnung eines Voxels als Merkmale), in welchem jede Achse eines der Voxelmerkmale darstellt, wird aus

vorklassifizierten Voxeln eine Lernmenge erstellt. Diese Lernvoxel werden nun im Merkmalsraum an den ihren Merkmalswerten entsprechenden Koordinaten eingetragen. Schließlich wird ein Bildvoxel neuer Patient\*innen-Daten klassifiziert, indem es dem Merkmalsraum hinzugefügt wird und die K Lernvoxel, die ihm am nächsten liegen, untersucht werden. Der neue Fall wird entsprechend der Klassen dieser K Nachbarn klassifiziert – es könnte ihm beispielsweise die häufigste Klasse der K Nachbarn zugewiesen werden. Dadurch ließen sich jedem Voxel Läsionswahrscheinlichkeiten zuordnen, woraus eine Wahrscheinlichkeitskarte erstellt wurde. Durch Anwendung verschiedener Schwellenwerte auf die Wahrscheinlichkeitskarte wurden binäre Segmentierungen der Läsionen erzeugt (51).

Intelligente Modelle basieren vielfach auf Methoden des Maschinellen Lernens, einer Subkategorie der Künstlichen Intelligenz (KI) (52). CAD-Systeme zur Läsionssegmentierung arbeiten häufig auf Grundlage künstlicher neuronaler Netzwerke (artificial neural networks – ANNs), welche zum Deep Learning, was ein Teilbereich des Maschinellen Lernens ist, gezählt werden (45,52). Künstliche neuronale Netzwerke wurden von der Neurobiologie inspiriert. Jedes ANN besteht aus Knoten (analog zum Zellkörper eines Neurons), welche mit anderen Knoten über Verbindungen (analog zu Axonen und Dendriten) kommunizieren (52). Jede der Verbindungen hat ein sogenanntes Gewicht, welches den Eingabe- oder Ausgabewert verändert und dessen Wert während des Trainingsprozesses bestimmt wird, da sich die Gewichtungen zwischen den Knoten durch den Einfluss von Daten verändern können, was als Lernprozess oder auch Training des ANN bezeichnet wird. Somit lernt ein ANN Eingabemuster aus Beispielen zu verstehen und zu klassifizieren (53). ANNs bestehen üblicherweise aus 3 Schichten: einer Eingabeschicht (Input Layer), einer verborgenen Schicht (Hidden Layer) und einer Ausgabeschicht (Output Layer) (54). Jede Schicht eines ANNs kann beliebig viele Knoten enthalten, wobei die Anzahl der Knoten der Ausgabeschicht in der Regel der Anzahl der vorhergesagten Klassen, in welche das neuronale Netzwerk die Input-Daten einteilen soll, entspricht. Die meisten ANNs arbeiten als Feed-Forward-Netzwerke, was bedeutet, dass Informationen von einem Knoten der vorherigen Schicht, beginnend bei der Eingabeschicht, zu einem Knoten der darauffolgenden verborgenen Schicht weitergeleitet werden, bis sie schließlich als Ergebnis in der Ausgabeschicht erscheinen. (52,53). Jeder Knoten besitzt eine Funktion, auch Aktivierungsfunktion genannt, mit welcher er seinen Input in einen Output umwandelt (53). Die Aktivierungsfunktion kann dann mithilfe eines

Schwellenwertes ermitteln, ob das künstliche Neuron reagiert und Informationen weitergibt (54). Zijdenbos et al. entwickelten ein halbautomatisches Segmentierungssystem, genannt MIDAS (Medical Image Display and Analysis System), welches auf Basis eines künstlichen neuronalen Netzwerkes Läsionen der weißen Substanz in MRT-Bildern quantifiziert. MIDAS arbeitet mit drei Eingabeknoten (jeweils einer der Bildmodalitäten, bestehend aus T1-gewichteten, Protonendichte-gewichteten und T2-gewichteten Bildern, entsprechend), einer verborgenen Schicht mit zehn Knoten und einer Ausgabeschicht mit fünf Knoten, entsprechend einer der fünf möglichen Klassen (Hintergrund, weiße Substanz, graue Substanz, Liquor, Läsionen der weißen Substanz) (55).

Auf deformierbare Modelle wird im Zuge des Methodik-Teiles dieser Diplomarbeit genauer eingegangen.

Schließlich kommt es zur Bildnachverarbeitung. Dieser Schritt ist notwendig, um falsch positiv und falsch negativ klassifizierte Pixel zu eliminieren. Dieser Prozess kann manuell oder automatisch erfolgen (45).

## 1.6 Prognose

Die Prognose der Multiplen Sklerose variiert erheblich von Patient\*in zu Patient\*in (1). Der klinische Verlauf entwickelt sich in der Regel über mehrere Jahrzehnte, wobei die mittlere Zeit bis zum Tod etwa 30 Jahre ab Krankheitsbeginn beträgt, was einer Verkürzung der Lebenserwartung um ca. 5-10 Jahre entspricht (14).

Bei den meisten MS-Patient\*innen kommt es zu einer irreversiblen Verschlechterung der Mobilität, welche anhand der EDSS-Skala nach Kurtzke erfasst wird. Im Durchschnitt kommt es nach 8 Jahren Krankheitsdauer zu einer eingeschränkten Gehfähigkeit, nach ca. 20 Jahren ist eine einseitige Gehhilfe erforderlich und nach ca. 30 Jahren benötigen Patient\*innen einen Rollstuhl (56). Trotz allem ist zu erwähnen, dass sich die Aussichten für MS-Patient\*innen dank einer Weiterentwicklung verlaufsmodifizierender Therapeutika in den letzten Jahrzehnten verbessert haben (30). He et al. ermittelten in einer Beobachtungs-Kohortenstudie die Auswirkungen einer frühzeitigen Gabe verlaufsmodifizierender Therapeutika bei MS. In der Studie wurde mit Daten aus dem schwedischen MS-Register gearbeitet. Dieses beinhaltet Patient\*innen-Informationen aus neurologischen Kliniken in ganz Schweden und umfasst dadurch ca. 80% aller prävalenten MS-Fälle in der schwedischen Population. Es konnte gezeigt werden, dass ein frühzeitiger

Start verlaufsmodifizierender Therapeutika mit einer Verbesserung der von den Patient\*innen angegebenen körperlichen Symptome in Verbindung steht (57). In einer weiteren Studie von He et al. wurde der Fragestellung nachgegangen, ob ein früher Beginn mit hochwirksamen Therapeutika, welche traditionell erst nach erfolgloser Behandlung mit Erstlinien-Therapeutika eingesetzt wurden, mit einer geringeren langfristigen Invalidität einhergeht. In diesem Zusammenhang wurde die langfristige Invalidität von Patient\*innen, welche innerhalb von zwei Jahren nach Erkrankungsbeginn mit hochwirksamen Therapeutika begonnen haben, mit der langfristigen Invalidität von Patient\*innen, welche erst nach vier bis sechs Jahren nach Erkrankungsbeginn mit hochwirksamen Therapeutika begonnen haben, verglichen. Auch diese Studie konnte veranschaulichen, dass eine frühere Gabe hochwirksamer Therapeutika (innerhalb der ersten zwei Jahre ab Erkrankungsbeginn) mit einer geringeren langfristigen Invalidität und einem niedrigeren Risiko für ein Fortschreiten der funktionellen Beeinträchtigung assoziiert ist (58).

Bis dato gibt es keinen einzelnen Biomarker, der den individuellen Krankheitsverlauf der Multiplen Sklerose exakt vorhersagen kann. Aus diesem Grund wird derzeit aus der Zusammenschau klinischer Beurteilung, bildgebender Marker und Labormarker eine Behandlungsentscheidung getroffen (1). Es gibt allerdings eine Reihe von Merkmalen, welche die Prognose der Multiplen Sklerose beeinflussen und in folgender Tabelle zusammengefasst werden:

**Tabelle 4: Merkmale mit Auswirkung auf die Prognose der MS**

<b>Bessere Prognose</b>	<b>Schlechtere Prognose</b>
Großes Intervall zwischen erstem und zweitem Schub	Häufige Schübe in den ersten zwei bis fünf Jahren
Keine Funktionseinschränkungen nach fünf Jahren	Erhebliche Funktionseinschränkungen nach fünf Jahren
Unauffällige initiale MRT	Auffällige initiale MRT mit hoher Läsionslast
Komplette Erholung von der ersten neurologischen Episode	Infratentorielle oder spinale Läsionen in der MRT
Optikus-Neuritis oder isolierte sensorische Symptome als Erstmanifestation	Manifestation mit motorischen, zerebellären, sphinkteren oder polyregionalen Symptomen
	Männliches Geschlecht
	Höheres Alter bei Krankheitsbeginn (> 40 Jahre)
	ethnischer Einfluss (Nicht-Kaukasier*innen)
	Vorkommen Liquor-spezifischer oligoklonaler Banden

Quellen: eigene Tabelle in Anlehnung an Hunter, 2016, S. 142 (59), Oh et al., 2018, S. 758 (1), Travers et al., 2022, S. 205 (30)

## 1.7 Therapie

Die Behandlung von Patient\*innen mit Multipler Sklerose erfordert eine multiprofessionelle Herangehensweise und eine umfassende Strategie mit medikamentösen und nicht-medikamentösen Ansätzen für das Schub-Management, die Prävention von Schüben und die Verzögerung des Fortschreitens der Erkrankung (20). Mit der Zulassung verschiedener effizienterer verlaufsmodifizierender Therapien und durch ein verbessertes Verständnis der Krankheitsmechanismen und –entwicklung in den letzten Jahren, haben sich Prognose und Krankheitsverlauf der MS erheblich verändert (2).

Die Therapie der Multiplen Sklerose lässt sich in 3 Kategorien unterteilen: 1. Schubtherapie, 2. verlaufsmodifizierende Therapie (DMTs = disease-modifying treatments) und 3. symptomatische Therapie (4).

Das wichtigste Standbein und die 1. Wahl der Schubtherapie ist die intravenöse Glucocorticoid-Hochdosistherapie mit Methylprednisolon für drei bis fünf Tage. Dieses ist zwar nicht verlaufsmodifizierend, verkürzt aber die Schubdauer. Als Zweitlinien-Schubtherapie kann eine Plasmapherese vor allem bei Versagen der Erstlinientherapie in Betracht gezogen werden. Vor Therapiebeginn ist es natürlich wichtig abzuklären, ob es sich um einen tatsächlichen Schub oder einen Pseudoschub handelt (4).

Die Behandlung der Multiplen Sklerose wurde in den letzten 20 Jahren durch den Einsatz verlaufsmodifizierender Therapien (DMTs) revolutioniert. Das ultimative Ziel der Behandlung mit DMTs ist „NEDA“, was für „no evidence of disease activity“ steht, also „keine Evidenz der Krankheitsaktivität“ (4). NEDA ist somit ein Parameter, der anzeigt, ob eine Therapie effektiv ist. Dies wird anhand klinischer (keine Schübe, keine Krankheitsprogression anhand des EDSS gemessen) und bildgebender (keine neuen T2-Läsionen/keine kontrastmittelanreichernden Läsionen) Werte ermittelt (60). Grundsätzlich lassen sich zwei verschiedene Ansätze zur Anwendung der DMTs unterscheiden: Eskalation und Induktion. Bei der Eskalation wird ein niedriger effektives Therapeutikum mit einem günstigen Sicherheitsprofil für die Initialtherapie ausgewählt und erst bei Therapieversagen, d.h. Krankheitsprogression, auf ein potenteres Therapeutikum umgestiegen. Beim Induktions-Ansatz wird bereits die Initialtherapie mit einem hoch-effektiven Medikament begonnen (2,4,30). Letzteres findet seinen Ursprung in Daten aus Beobachtungsstudien, welche darauf hindeuten, dass eine Erstbehandlung mit

hochwirksamen verlaufsmodifizierenden Medikamenten mit einem geringeren Risiko der Umwandlung einer RRMS in eine SPMS verbunden sein kann (2).

DMTs verringern die Schubfrequenz und reduzieren kurzfristige Funktionseinschränkungen (18). Sie verändern den Verlauf der MS durch Modulation oder Suppression der Immunfunktion (61). Zu den Wirkmechanismen zählen u.a. eine Beeinträchtigung der DNA-Synthese in Lymphozyten, Immunzell-Depletion und eine Veränderung der Zytokinsekretion. Genannte Mechanismen sind aber nur ein Auszug aus dem komplexen Wirkprofil der DMTs und für einige DMTs ist der Wirkmechanismus noch nicht genau geklärt (18).

Die ersten DMTs wurden in den 1990er Jahren zugelassen, wobei es sich um injizierbare Präparate auf Interferon- $\beta$ -Basis und auf Glatiramer-Acetat-Basis handelte, welche inzwischen teilweise weiterentwickelt wurden. Sie führen u.a. zu einer Modulation der antigenpräsentierenden Zellen und verhindern Immunzell-Migration über die Blut-Hirn-Schranke. Später kamen orale Substanzen wie Fingolimod, Dimethylfumarat und Teriflunomid dazu. In den letzten Jahren wurden außerdem noch selektivere S1P-Rezeptor-Modulatoren wie beispielsweise Siponimod entwickelt, welche ebenfalls oral verabreicht werden. Diese führen zu einer Retention der Lymphozyten in lymphatischen Organen (2).

Anfang der 2000er Jahre wurde an einer weiteren vielversprechenden Therapie geforscht – monoklonalen Antikörpern (2). Die Verabreichung erfolgt parenteral und der erste zugelassene monoklonale Antikörper zur Therapie der MS war Natalizumab, welcher das  $\alpha 4\beta 1$ -Integrin an der Lymphozytenoberfläche inhibiert und so deren Migration in das ZNS verhindern soll. Die Therapie mit Natalizumab geht allerdings mit einem erhöhten Risiko für das Auftreten einer PML (progressiven multifokalen Leukenzephalopathie) einher, welche eine opportunistische Infektion mit dem John Cunningham-Virus (JCV) darstellt und in ca. 0,4% der mit Natalizumab behandelten Patient\*innen auftritt. Daher wird Natalizumab nur für JC-Virus-Antikörper-negative Patient\*innen empfohlen (61). Regelmäßige Sicherheitstests sind erforderlich und umfassen die Bestimmung des Antikörpertiters gegen HPyV-2 (humanes Polyomavirus = JC-Virus), sowie MRT-Untersuchungen, um PML-verdächtige Läsionen frühzeitig zu erkennen (2).

In den letzten Jahren eröffnete sich anhand von anti-CD20-Antikörpern eine weitere Therapiemöglichkeit (2). Sie binden an CD20-exprimierende B-Zellen und führen zu deren

Zerstörung (Depletion) (18). Der bedeutendste Wirkstoff aus dieser Gruppe ist Ocrelizumab, welcher besonders erwähnenswert ist, da er der erste und einzige auch für die PPMS zugelassene Wirkstoff ist. Ein weiterer i.v.-verabreichter Antikörper ist Alemtuzumab (2). Dieser bindet selektiv das CD52-Antigen auf Lymphozyten und führt so zu einer Depletion dieser (18). Trotz seiner Potenz als Wirkstoff in der Behandlung der Multiplen Sklerose ist Alemtuzumab inzwischen eher als „Reserve-Medikament“ einzustufen und wird nur bei ausgewählten Patient\*innen-Gruppen verabreicht, da es potenziell schwerwiegende Nebenwirkungen aufweist (62).

Zuletzt ist das oral verabreichte Purin-Analogon Cladribin zu nennen, welches ursprünglich in der Onkologie eingesetzt wurde (2,18). Es wird in sich rasch vermehrende Zellen, wie Lymphozyten, aufgenommen und als fehlerhafter Baustein in die DNA eingebaut, wodurch es zum Zelltod kommt (18).

Die Therapie mit verlaufsmodifizierenden Medikamenten ist eine lebenslange, es sei denn es kommt zu einem Krankheitsdurchbruch oder es treten unerwünschte Wirkungen auf, welche einen Medikamentenwechsel erfordern. Außerdem zeigen mehrere Beobachtungsstudien, dass ältere Patient\*innen mit langer Therapiedauer, welche im Sinne eines Krankheitsstillstands über einen längeren Zeitraum klinisch und radiologisch stabil geblieben sind, ein geringes Risiko für ein Wiederauftreten von Krankheitsaktivität besitzen und dadurch ein kontrolliertes Absetzen der DMTs veranlasst werden kann. Aktuell existieren dahingehend allerdings keine gültigen Richtlinien (18).

Das Therapieansprechen von MS-Patient\*innen auf DMTs erweist sich in der Praxis als sehr heterogen. Kliniker\*innen müssen das individuelle Ansprechen von Patient\*innen auf Therapien wiederholt prüfen, um optimale Therapieentscheidungen und damit einhergehend einen frühzeitigen Wechsel auf ein alternatives, potenziell wirksameres Therapeutikum zu ermöglichen. Rio et al. schlugen ein Scoring-System vor, welches anhand aktiver MRT-Läsionen, der Schubanzahl und des EDSS jene Patient\*innen identifizieren kann, welche ein hohes Risiko für ein schlechtes Therapieansprechen haben. Sormani et al. versuchten diesen sogenannten „Rio-Score“ zu vereinfachen und veröffentlichten den „modifizierten Rio-Score“ (63). Der modifizierte Rio-Score umfasst zwei Kriterien und wird nach dem ersten Jahr der Therapie erhoben: 1. MRT-Kriterium (1 Punkt für Patient\*innen mit >4 neuen T2-Läsionen), 2. Schub-Kriterium (1 Punkt für Patient\*innen mit einem Schub, 2 Punkte für Patient\*innen mit  $\geq 2$  Schüben) (64). Anhand

des Scores werden Patient\*innen in geringes Risiko (Score= 0), mittleres Risiko (Score=1) und hohes Risiko (Score= 2-3) für ein schlechtes Therapieansprechen nach einem Jahr Therapie eingestuft (63).

Abschließend kurz zu erwähnen ist die symptomatische Therapie der Multiplen Sklerose, welche ebenfalls große Bedeutung hat. Diese erweist sich als sehr unterschiedlich, je nach Symptomatik. Als Beispiele sind hier die Verabreichung von Botulinumtoxin bei Spastizität, Laxantien bei Obstipation oder intermittierender Selbstkatheterismus bei erhöhtem Residualvolumen in der Harnblase (4).

## 1.8 Fragestellung und Zielsetzung

Es lässt sich aus den in vorangegangenen Abschnitten dargestellten Überlegungen und Fakten erkennen, dass die Läsionssegmentierung von MRT-Bildern bei Multipler Sklerose große Relevanz für die Diagnostik, Verlaufsuntersuchungen, Monitoring, aber vor allem das Management der Erkrankung hat (43–45). In den letzten Jahren wurden immer effizientere Therapien verfügbar, welche mehr Sorgfalt in deren Auswahl durch Differenzierung tatsächlich „krankheitsaktiver“ Patient\*innen von Patient\*innen mit „Therapieversagen“ bei zuvor etablierter Medikation benötigen (2).

Zugleich ist die bisherige, standardmäßig angewandte manuelle Segmentierung mit klaren Einschränkungen verbunden, da diese durch die große zu analysierende MRT-Datenmenge zeitaufwändig ist, Expertise erfordert, wobei eine flächendeckende Verfügbarkeit erfahrener Neuroradiolog\*innen zum aktuellen Zeitpunkt nicht gegeben ist und wegen der subjektiven Beurteilung durch Menschenhand starken Sensitivitätsschwankungen hinsichtlich der Läsionsbeurteilung unterliegt (43–45).

In der Routine konnten bisher noch keine ausreichend objektivierbaren Standards in der radiologischen Interpretation von MRT-Aufnahmen bei MS-Patient\*innen etabliert werden. Aktuell werden MRT-Aufnahmen in der klinischen Praxis zumeist von Radiolog\*innen ohne Spezialisierung im Bereich der Neuroradiologie mittels Augenmaß verglichen, wohingegen der Trend zu objektivierbareren und reproduzierbareren Methoden gehen muss, da von diesen Befunden wichtige Entscheidungen wie Therapieansprechen, -versagen oder -wechsel abhängen. Außerdem sind neue Bildgebungs-Standards wie 3D-FLAIR relativ schnell zu akquirieren und zu standardisieren (3). Damit wird die automatisierte Segmentierung von MS-Läsionen immer bedeutsamer.

In dieser Diplomarbeit wurde in Zusammenarbeit mit dem Hospital Universitari Vall d'Hebron in Barcelona der Prototyp eines von dem Spin-Off-Unternehmen Tensor Medical entwickelten entscheidungsunterstützenden Systems, genannt copilot.ms, zur automatischen Läsionssegmentierung auf Basis von MRT-Bildern bei Multipler Sklerose, an einer Kohorte von Patient\*innen des Grazer MS-Zentrums getestet.

Folgende Fragestellung soll mit dieser Diplomarbeit beantwortet werden: Ist es möglich mittels automatischer Segmentierung von T2-gewichteten MRT-Bildern mithilfe eines entscheidungsunterstützenden Systems und anschließendem Validieren durch Expert\*innen gleiche Genauigkeit in der Erkennung von neuen T2-Läsionen bei MS-Patient\*innen zu erreichen wie durch alleinige manuelle Segmentierung?

Hypothese dieser Arbeit war, dass durch automatische Segmentierung mittels copilot.ms und folgende Validierung durch Expert\*innen eine Verbesserung der Läsions-Analyse in MRT-Bildern erreicht werden kann, indem eine validierte, automatische Auswertung eine gleiche Genauigkeit wie die manuelle Segmentierung mit dem Vorteil der zeitlichen Effizienz aufweist.

## 2 Material und Methoden

### 2.1 Studienaufbau

Diese Studie ist Teil einer internationalen, multizentrischen Validierungsstudie, um Qualitätskriterien des Prototyps eines Programmes zur automatischen Läsionssegmentierung, genannt Copilot.ms, an im Routinealltag mittels eines standardisierten MRT-Untersuchungsprotokolls erhobenen Daten aus dem MS-Zentrum Graz zu testen. Die Durchführung erfolgte in Zusammenarbeit mit dem Hospital Universitari Vall d'Hebron in Barcelona und der Spin-Off-Firma Tensor Medical, welche Copilot.ms entwickelte.

Wie bereits dargestellt, ist die Identifikation und Segmentierung MS-typischer, hyperintenser T2-Läsionen für das Monitoring der Erkrankung von entscheidender Bedeutung (3). Daher war es Ziel der Studie, die automatische Erkennung neuer T2-Läsionen in MRT-Folgeuntersuchungen mit Copilot.ms zu analysieren, zu validieren und die Leistung des Analyseprogramms mit der bisher gebräuchlichen, manuellen Segmentierung zu vergleichen.

### 2.2 Patient\*innen-Kohorte

Um die Studienpopulation hinsichtlich klinischer und demographischer Charakteristika näher beschreiben zu können, wurden zwischen Mai und Juli 2022 Patient\*innen-Daten aus dem Krankenhaus-Informationssystem openMEDOCS, sowie der RDA-Datenbank für MS, welche eine zentrale Datenbank wissenschaftsrelevanter Daten für die medizinische Forschung darstellt, am Universitätsklinikum für Neurologie Graz extrahiert und gesammelt.

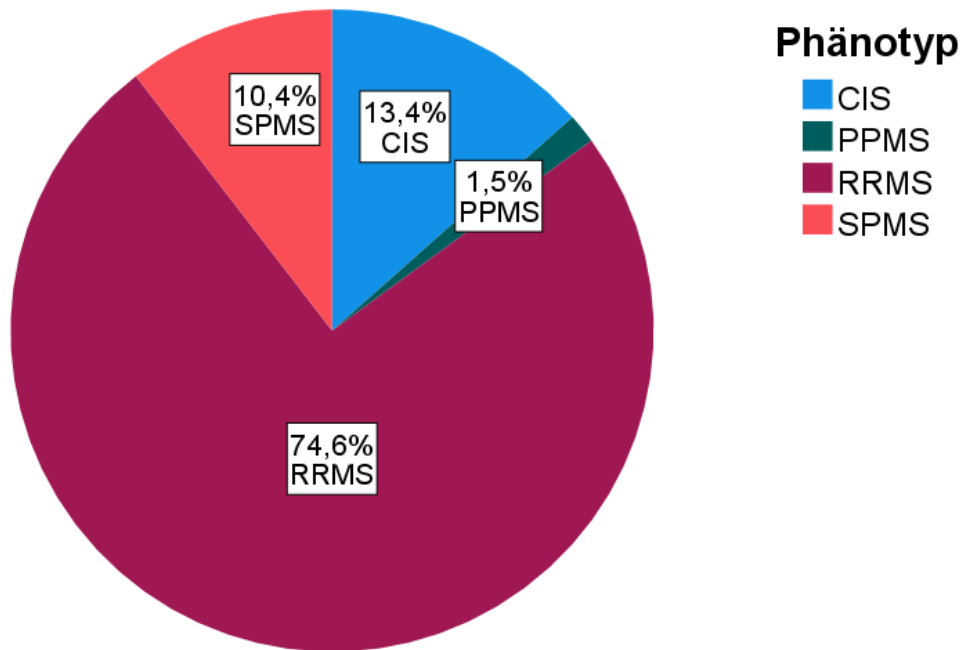
Die in dieser retrospektiven Studie analysierten MRT-Bilder stammen von einer Patient\*innen-Kohorte von 67 Patient\*innen des Grazer MS-Zentrums der Universitätsklinik für Neurologie Graz. Die Kohorte setzt sich aus 39 (58,2%) Frauen und 28 (41,8%) Männern zusammen, mit einem mittleren Alter von  $40,1 \pm 10,5$  Jahren (Altersspanne 20,4 -57,9 Jahre) zum Zeitpunkt des Follow-Up (mittleres Follow-Up Intervall  $319 \pm 82$  Tage, Follow-Up-Intervallspanne 84-434 Tage). Somit zeigt sich in der Kohorte eine Geschlechterverteilung von Frauen zu Männern von 3:2, was insofern der für Multiple Sklerose bekannten Geschlechterverteilung entspricht, als Frauen häufiger betroffen sind, wenngleich Männer in dieser Kohorte überrepräsentiert sein dürften.

Die Krankheitsdauer in der Studienpopulation betrug im Mittel  $10,3 \pm 7,3$  Jahre. EDSS-Werte wurden zum Zeitpunkt der Baseline-MRT, sowie der Follow-Up-MRT erhoben. Da EDSS-Werte keine Normalverteilung zeigen, wurden Median und IQR (interquartile range = Interquartilsabstand) zur Beschreibung der Werteverteilung der EDSS anstatt des Mittelwertes und der Standardabweichung herangezogen. Der IQR ist der Abstand zwischen dem dritten und ersten Quartil ( $IQR = 3.\text{Quartil} - 1.\text{Quartil}$ ) und umfasst somit die mittleren 50% der Daten. Der EDSS-Median zum Baseline- und Follow-Up-Zeitpunkt betrug jeweils 1. Der errechnete IQR für Baseline- und Follow-Up-EDSS betrug jeweils 3 (1. Quartil= 0, 3. Quartil= 3).

**Tabelle 5: Deskriptive Statistik zur Studienpopulation**

	<b>Mittelwert</b>	<b>Standard- Abweichung</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Spannweite</b>
<b>Alter (bei Follow Up), Jahre</b>	40,1	10,5	20,4	57,9	37,9
<b>Krankheitsdauer, Jahre</b>	10,3	7,3	1,0	27,0	26,0
<b>Follow-Up-Intervall, Tage</b>	319	82	84	434	350
	<b>Median</b>	<b>IQR</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Spannweite</b>
<b>EDSS Baseline</b>	1	3	0	6	6
<b>EDSS Follow-Up</b>	1	3	0	6	6

Weiters wurde die Studienpopulation hinsichtlich des MS-Phänotyps (CIS, RRMS, SPMS, PPMS) untersucht. Die RRMS stellte mit 74,6% (50 Patient\*innen) den häufigsten Phänotyp dar, gefolgt von dem CIS mit 13,4% (9 Patient\*innen), der SPMS mit 10,4% (7 Patient\*innen) und schließlich der PPMS mit 1,5% (ein Fall in der Kohorte).



**Abbildung 8: Kreisdiagramm zur Phänotypen-Verteilung in der Studienpopulation**

82,1% der Studienpopulation erhielten während des Studienzeitraumes eine verlaufsmodifizierende Therapie. 10,4% der Patient\*innen erlitten im Intervall jeweils einmalig einen Schub.

## 2.3 MRT-Untersuchungen und Einschlusskriterien

In dieser Studie wurden bereits standardisiert akquirierte MRT-Daten, welche an einem 3T Forschungs-Magnetresonanztomographen aufgenommen wurden, zur Analyse herangezogen.

Voraussetzung für die Teilnahme an der Studie war das Vorhandensein von zwei T2-gewichteten 3D-FLAIR-Sequenz MRT-Bildern pro Patient\*in, die im Abstand von 6-24 Monaten an demselben Magnetresonanztomographen (MAGNETOM PRISMA, SIEMENS Healthineers) aufgenommen wurden. Die MRT-Bilder wurden zwischen dem 2. Dezember 2019 und dem 10. Februar 2021 akquiriert.

Die Anzahl der eingeschlossenen Patient\*innen wurde durch die maximale Anzahl der verfügbaren Follow-Up-Scans auf demselben Magnetresonanztomographen nach demselben Protokoll bestimmt.

Weitere Patient\*innen-Charakteristika zum Studieneinschluss wurden im Einklang mit dem primären Studienziel nicht spezifiziert.

Patient\*innen-Daten wurden im Zuge dieser Studie pseudonymisiert verarbeitet.

Zur Durchführung der Studie wurde ein Votum von der Ethikkommission der Medizinischen Universität Graz mit EK-Nummer 31-432 ex 18/19 eingeholt. Eine schriftliche, informierte Einwilligung von allen teilnehmenden Patient\*innen liegt vor.

## 2.4 MRT-Analysen und Vergleiche

Um die Möglichkeit der Einbindung von Copilot.ms in den klinischen Alltag zu testen, ist es notwendig, die Ergebnisse der Plattform mit Referenzmethoden hinsichtlich der Fähigkeit, neue T2-Läsionen korrekt zu erkennen, zu vergleichen.

Es wurden drei verschiedene Methoden herangezogen, um die Anzahl neuer T2-Läsionen in den vorhandenen Baseline- und Follow-Up-MRT-Bildern unserer Studienpopulation zu bestimmen und zu vergleichen:

- 1) Rein manuelle Segmentierung (visuelle Analyse) = VA
- 2) Rein automatisierte Analyse mit der Plattform Copilot.ms = AA
- 3) Vorläufige Analyse durch die Plattform Copilot.ms, gefolgt von einer visuellen Validierung = AAValid

### 2.4.1 Manuelle Segmentierung

Die manuelle Segmentierung stellt den bisherigen Goldstandard dar. Um die Anzahl der durch visuelle Analyse erkannten T2-Läsionen zu ermitteln, wurden zwischen Mai und Juli 2022 die den MRT-Bildern zugehörigen neuroradiologischen standardisierten Berichte aus openMEDOCS analysiert. Daraus extrahierte Informationen zur Änderung der MS-Läsionszahl der durch einen neuroradiologisch auf dem Gebiet der MS einschlägig erfahrenen Befundersteller der Universitätsklinik Graz identifizierten, neuen T2-Läsionen wurden zum statistischen Vergleich tabellarisch in einem Excel-Dokument zusammengefasst.

### 2.4.2 Automatisierte Analyse

Die automatisierte Analyse der MRT-Bilder erfolgte mittels Copilot.ms, einem von der Firma Tensor Medical entwickelten, entscheidungsunterstützenden System zur automatischen, supervidierten Läsionssegmentierung. Copilot.ms stellt ein CAD-System dar, welches auf Basis eines zur Erkennung von neuen T2-Läsionen trainierten neuronalen Netzwerkes arbeitet. Neue Läsionen werden anhand eines Algorithmus erkannt, welcher

auf einem deformierbaren Modell, sogenannten Deformationsfeldern (deformation fields), fußt (65).

#### 2.4.2.1 Deformierbare Modelle

Techniken zur automatischen Detektion von T2-Läsionen lassen sich in zwei Kategorien einteilen: intensitätsbasierte und deformationsbasierte Ansätze. Intensitätsbasierte Ansätze, denen zufolge aufeinanderfolgende Scans anhand eines Punkt-zu-Punkt- (Voxel-zu-Voxel) Vergleiches analysiert werden, wurden bereits im Zuge der Einleitung dieser Diplomarbeit behandelt. Deformationsbasierte Ansätze analysieren zwei aufeinanderfolgende Scans (Baseline und Follow-Up) anhand von Deformationsfeldern (66).

Die beiden oben genannten Ansätze basieren auf zwei Modellen des Läsionswachstums, welche in der Realität meist kombiniert auftreten: Gewebetransformation und Gewebedeformation. Intensitätsbasierte Ansätze beruhen auf Gewebetransformation, wobei Läsionen anhand von Intensitätsveränderungen erkannt werden, ohne dass sich dabei Gewebe verschiebt. Die molekulare Struktur des Gewebes verändert sich an Ort und Stelle und das Läsionsvolumen ist jener Bereich des MRT-Bildes, welcher eine andere Gewebszusammensetzung und folglich andere Signalintensitäten (im engeren Sinne Grauwerte) aufweist (67).

Gewebedeformation beschreibt hingegen jenes Phänomen, dass einige Läsionen lediglich durch starke Verformung des Gewebes charakterisiert werden können, ohne erkennbare Änderungen der Bildintensität. Es findet sich eine Änderung des Gewebes im Gehirn, welche sich aber nur durch Ausdehnung, Kontraktion oder Verformung des Gewebes erkennen lässt, da dieses Gewebe dieselbe Protonendichte wie Hirngewebe besitzt und dadurch keine Grauwertänderungen auftreten (67).

Deformationsfelder können durch nichtlineare Registrierung zweier aufeinanderfolgender MRT-Scans erhalten werden (66,68). Unter nichtlinearer Bildregistrierung versteht man Methoden, bei denen zu registrierende Bilder geometrische Unterschiede aufweisen, welche durch Ähnlichkeitstransformationen nicht ausgeglichen werden können (47). Nichtlineare Registrierungsverfahren erzeugen ein Vektorfeld, mit welchem Deformationen zwischen zwei in Folge entstandenen Bildern (beispielsweise einer Baseline- und Follow-Up-MRT) definiert werden können. Dieses Vektorfeld kann zur Erkennung sich entwickelnder Prozesse, wie beispielsweise neuer T2-Läsionen, herangezogen werden (66). Durch Registrierung von Baseline- und Follow-Up-Bild

kommt es im Bereich neuer Läsionen zu örtlichen Verzerrungen der Transformationsmatrix, da die neuen Läsionen im Baseline-Bild nicht vorhanden sind. Diese Verzerrungen lassen sich mittels Deformationsfeldern beschreiben. Schließlich kann ein Schwellenwert für die Verzerrung festgelegt werden, wodurch neue Läsionen automatisch identifiziert werden können (66–69).

#### 2.4.2.2 Ablauf der automatisierten Analyse

Bei Copilot.ms handelte es sich zum Zeitpunkt der Analysen, welche von Mai bis Juli 2022 stattfanden, um eine Online-Plattform, welche über den Link <https://copilot.tensormedical.ai/> erreicht werden konnte. Zugangsdaten zur Verwendung des Prototypen, welcher ausschließlich Forschungszwecken diente, wurden von der Firma Tensor Medical zur Verfügung gestellt.

Um die MRT-Bilder mit Copilot.ms auswerten zu können, mussten sie zuerst aus dem PACS-Archiv, welches der Bilddaten-Speicherung im Radiologischen-Informationssystem am Universitätsklinikum Graz dient, exportiert werden. In weiterer Folge wurden die Bilder in die Plattform Copilot.ms mittels „Drag and Drop“ importiert.

Nach Upload der Bilddaten (pro Patient\*in eine Baseline- und eine Follow-Up-MRT-Untersuchungsserie, vgl. Einschlusskriterien), erhält man auf der Startseite der Online-Plattform eine Übersicht über alle hochgeladenen MRT-Daten, geordnet nach pseudonymisierter Patienten\*innen-ID.

Nach dem Upload werden die MRT-Bilder sofort automatisch hinsichtlich neuer T2-Läsionen segmentiert. Die Analyse durch copilot.ms findet in einer Cloud, in die die MRT-Daten über die Website im Upload-Prozess hochgeladen werden, im Servernetzwerk der Firma Tensor Medical statt. Die Dauer des Upload- und Analyseprozesses ist abhängig von der Auslastung des Servernetzes, dauert im Durchschnitt jedoch nur wenige Minuten. Das bedeutet Benutzer\*innen müssen lediglich den Upload durchführen und erhalten bereits im nächsten Schritt, ohne ihr Zutun, eine Aussage über die Anzahl der neuen T2-Läsionen im Follow-Up-MRT-Bild. Copilot.ms gibt die absolute Anzahl neuer T2-Läsionen an, oder beschreibt den Status als „stable“, wenn keine neuen Läsionen gefunden wurden.

Um einen Fall abzuschließen, erfordert Copilot.ms eine Validierung der automatisierten Segmentierung. Dies erfolgte allerdings erst im nächsten Schritt der Analysen und war für die rein automatisierte Läsions-Analyse noch nicht von Bedeutung.

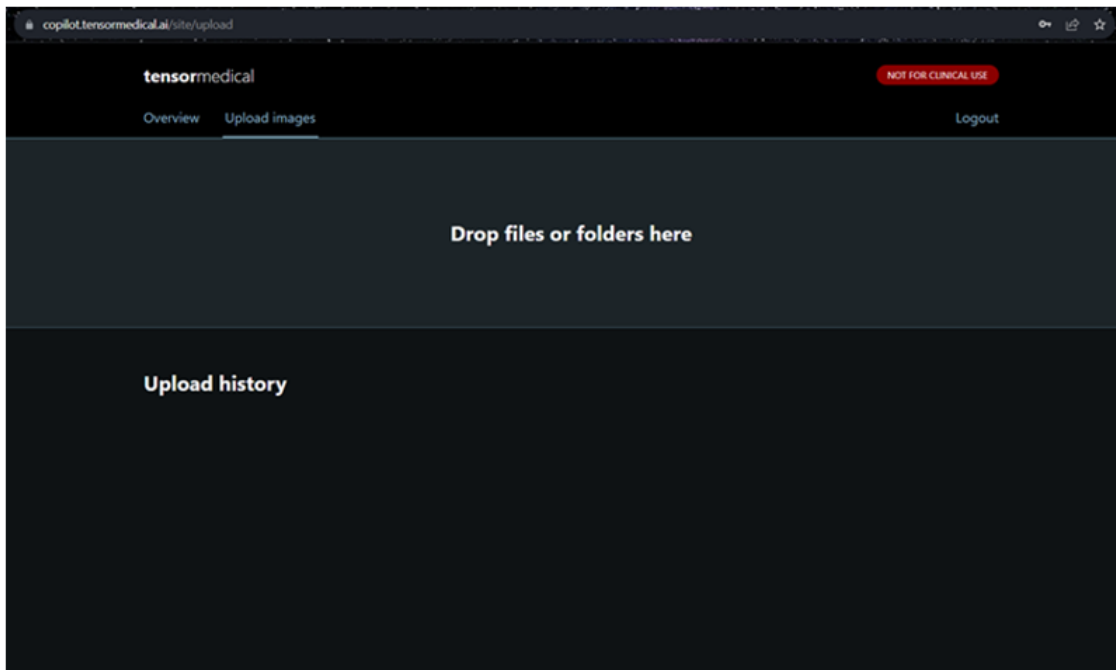
Die unter „Status“ in der Startseitenübersicht (Abb.10) herauszulesenden absoluten Zahlen neuer T2-Läsionen wurden ebenfalls für spätere, statistische Vergleiche tabellarisch in Excel übertragen.

### 2.4.3 Automatisierte Analyse mit Validierung

Die dritte Methode zur Ermittlung der Anzahl neuer T2-Läsionen in MRT-Daten der Studienpopulation war eine vorausgegangene automatisierte Segmentierung mittels Copilot.ms und anschließende manuelle Validierung der gefundenen Läsionen. Da Copilot.ms als entscheidungsunterstützendes System entwickelt wurde, ist letztlich der radiologische Validierungsprozess durch Expert\*innen entscheidend.

Im Zuge des Validierungsprozesses, welcher über den Button „validate“ auf der Startseite der Plattform (Abb. 10) gestartet werden kann, zeigt Copilot.ms die von dem Programm segmentierten Läsionen durch bunte Umzeichnung der Läsion an. Bediener\*innen des Programmes können die von Copilot.ms als neue Läsionen bezeichneten Bereiche als diese bestätigen, oder fälschlicherweise als Läsionen erkannte Bereiche korrigieren und diese „falsch positiven“ Läsionen löschen. Außerdem verfügt die Plattform über die Möglichkeit, dass Bediener\*innen manuell Läsionen einzeichnen können. Dies erlaubt eine Korrektur der durch das Programm nicht erkannten Läsionen und dient somit der Berichtigung falsch negativer Ergebnisse.

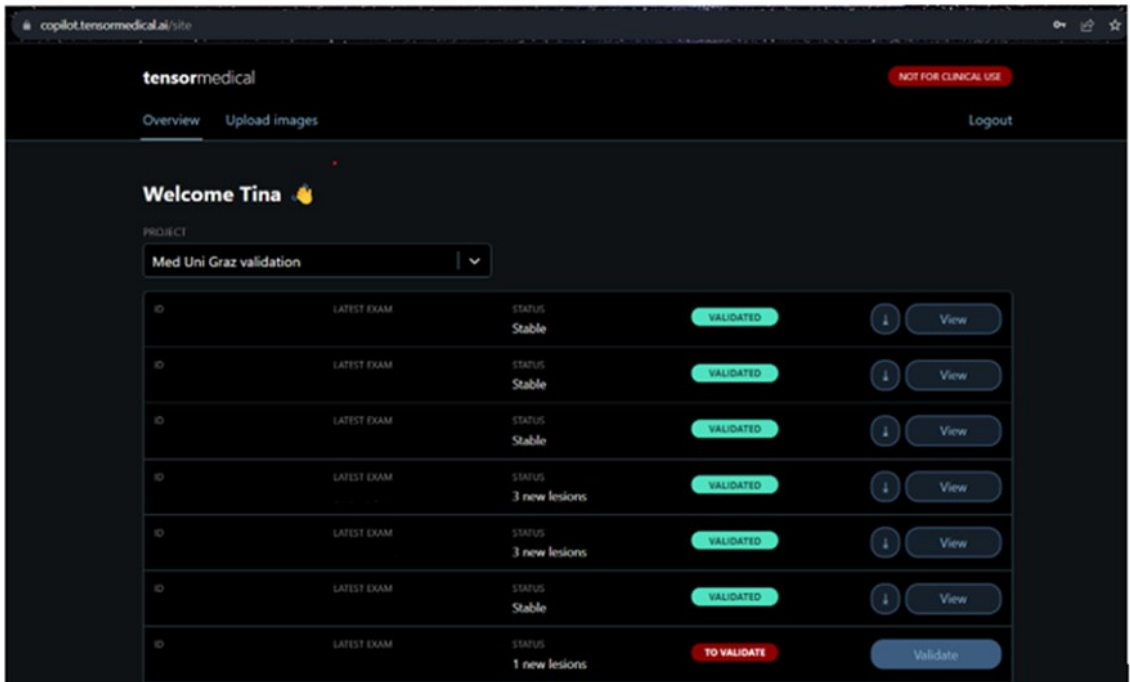
Die Handhabung von copilot.ms lässt sich somit in drei Schritte zusammenfassen, welche in den Abbildungen 9-11 anhand von Screenshots der Online-Plattform zur besseren Veranschaulichung dargestellt werden. Die pseudonymisierten Patient\*innen-IDs wurden in diesen Beispiel-Abbildungen unkenntlich gemacht.



**Schritt 1**

Hochladen der MRT-Daten mittels „drag and drop“

Abbildung 9: Schritt 1 - Upload der MRT-Daten



**Schritt 2**

Übersicht über die hochgeladenen MRT-Daten nach Patient\*innen-ID geordnet inklusive des durch copilot.ms ermittelten Läsions- und Validierungsstatus auf der Startseite

Abbildung 10: Schritt 2 - Startseitenansicht von copilot.ms



**Schritt 3**  
Benutzeroberfläche für den Validierungsprozess

**Abbildung 11: Schritt 3 - Validierungsprozess**

Die Benutzeroberfläche zur Validierung lässt sich in Abbildung 11 erkennen. Auf der linken Seite ist das Baseline-MRT-Bild zu erkennen, auf der rechten Seite der Follow-Up-Scan. Der von Copilot.ms als neue T2-Läsion erkannte Bereich ist in der rechten MRT-Aufnahme rot markiert.

Es können nun alle Schichten des vorliegenden Baseline- und Follow-Up-Bildes synchron durchgescrollt werden, um die Läsionsanzahl manuell zu validieren und abschließend zu bestätigen.

Nach erfolgter Validierung stellt Copilot.ms automatisch einen endgültigen Bericht über alle neuen Läsionen zusammen, welcher wie folgt aussehen kann:

### Patient Information

Patient ID

Date added 28/04/2022 12:30 UTC

### Imaging Sessions

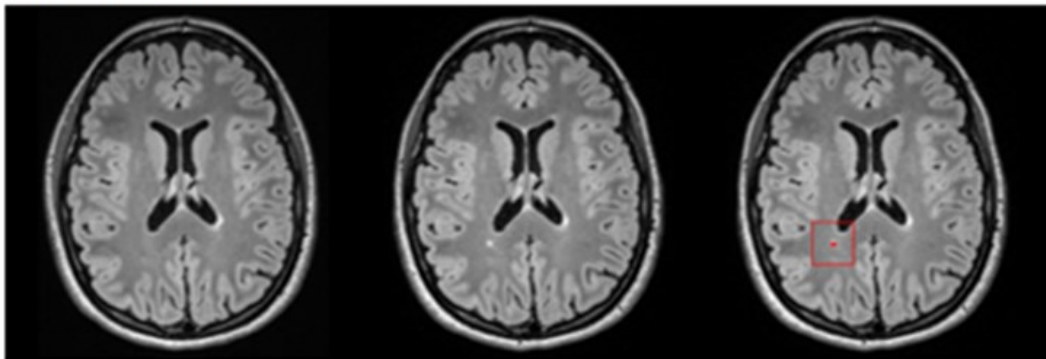
ID	Acquisition Date	Modalities	Scanner
Baseline		FLAIR	
Followup		FLAIR	

### Analysis Results

Months between 11 months

New lesions 2

#### Lesion 1 (Axial Slice 165)

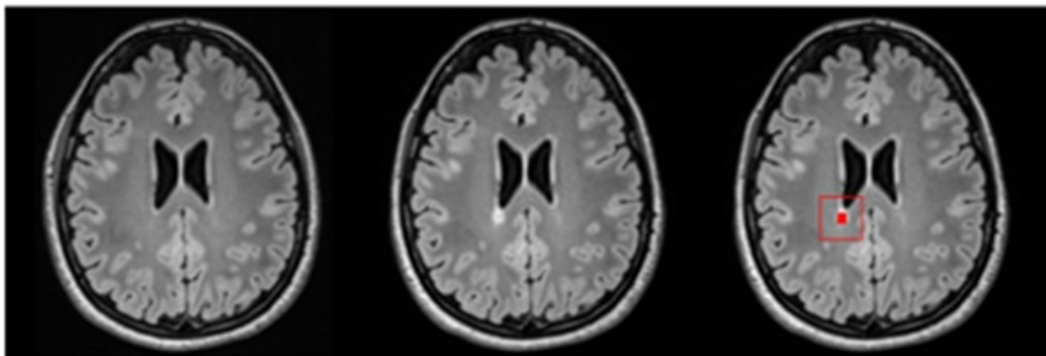


(a) Baseline

(b) Followup

(c) New Lesion

#### Lesion 2 (Axial Slice 174)



(a) Baseline

(b) Followup

(c) New Lesion

**WARNING: This is a research tool, do not use it for clinical purposes**

© Tensor Medical 2021  
www.tensormedical.ai

Abbildung 12: Beispiel Abschlussbericht Copilot.ms

Die validierten, absoluten Zahlen der neuen T2-Läsionen wurden auch bei dieser Methode zur statistischen Auswertung tabellarisch in Excel zusammengefasst.

## 2.5 Statistische Auswertung

Die Datenanalyse dieser Diplomarbeit wurde mit dem Statistikprogramm SPSS (Statistical Package for the Social Sciences) der Firma IBM, Version 28, durchgeführt.

Deskriptive, statistische Analysen (absolute und prozentuelle Häufigkeitsrechnungen, Mittelwert und Standardabweichung, Maximum, Minimum) wurden angewandt, um die Patient\*innen-Kohorte hinsichtlich demographischer und klinischer Eigenschaften zu charakterisieren.

Für die drei zuvor beschriebenen Methoden (VA, AA, AAValid) wurde die absolute Zahl der Summe neu detektierter T2-Läsionen ermittelt und mittels passender Statistik tabellarisch und grafisch dargestellt.

Zum Vergleich der Güte der beiden Methoden AA und AAValid mit dem bisherigen Goldstandard, der Methode VA, wurden je Methode Sensitivität, Spezifität und negativer Vorhersagewert auf Basis von Kreuztabellen mit entsprechenden Konfidenzintervallen berechnet.

Außerdem wurden die drei Methoden anhand eines Vorzeichentests auf Gleichheit untersucht. Dieser wurde als angemessenes Testverfahren gewählt, weil in dieser Diplomarbeit mit verbundenen Stichproben gearbeitet wurde, da alle statistisch verwerteten Daten von denselben 67 Patient\*innen stammen und auf demselben Scanner durchgeführt wurden, die Datenvariablen, sowie die Differenzen der Variablen, nicht normalverteilt sind und der Vorzeichentest keine Voraussetzungen bezüglich der Verteilungsform macht. Für die Analysen wurde ein Signifikanzniveau von  $p < 0,05$  als signifikant angesehen.

## 3 Ergebnisse

### 3.1 Absolute Zahlen neuer T2-Läsionen

Initial war die absolute Zahl neu detektierter T2-Läsionen je nach Methode und Patient\*in von Interesse. Von insgesamt 67 Studienteilnehmer\*innen, von denen 67 MRT-Scan-Paare vorlagen, traten bei 17 Teilnehmer\*innen (25,3%) mittels visueller Analyse (Methode VA) insgesamt 40 neue T2-Läsionen, bei 23 Teilnehmer\*innen (34,3%) mittels automatischer Analyse (Methode AA) insgesamt 56 neue T2-Läsionen und bei 22 Teilnehmer\*innen (32,8%) mittels validierter automatisierter Analyse (Methode AAValid) insgesamt 61 neue T2-Läsionen im Follow-Up-Scan auf. Die restliche Studienpopulation zeigte kein Auftreten neuer T2-Läsionen.

Jene 17 Proband\*innen, welche mittels **visueller Analyse (Methode VA)** neue T2-Läsionen aufwiesen, waren im Mittel zum Follow-Up-Zeitpunkt  $36,0 \pm 9,8$  Jahre alt, bei einem mittleren Follow-Up-Intervall von  $286 \pm 95$  Tagen (Follow-Up-Intervallspanne 84-378 Tage). Sie setzten sich aus 11 Frauen (64,7%) und 6 Männern (35,3%) zusammen und hatten durchschnittlich eine Krankheitsdauer von  $7,8 \pm 5,7$  Jahren.

Drei (17,6%) der 17 Proband\*innen litten an einem CIS, eine\*r (5,9%) an einer SPMS und der Großteil mit 13 Proband\*innen (76,5%) an einer RRMS. Nur zwei (11,8%) der 17 Patient\*innen erhielten während des Studienzeitraumes eine verlaufsmodifizierende Therapie. Bei sechs (35,3%) der 17 Patient\*innen kam es im Intervall zwischen beiden MRT-Aufnahmen zu einem einmaligen Schubgeschehen.

Die EDSS-Werte jener 17 Patient\*innen betragen im Median sowohl zum Baseline-, als auch zum Follow-Up-Zeitpunkt jeweils 1, mit einem IQR zum Baseline-Zeitpunkt von 3 (1. Quartil = 0, 3. Quartil = 3) und einem IQR zum Follow-Up-Zeitpunkt von 2 (1. Quartil = 2, 3. Quartil = 0).

In der **automatischen Analyse (AA)** wurden bei 23 Teilnehmer\*innen neue T2-Läsionen detektiert. Jene 23 Teilnehmer\*innen setzten sich aus 15 Frauen (65,2%) und 8 Männern (34,8%) zusammen und hatten im Mittel zum Follow-Up-Zeitpunkt ein Alter von  $36,5 \pm 10,6$  Jahren, bei einem mittleren Follow-Up-Intervall von  $308 \pm 86$  Tagen (Follow-Up-Intervallspanne 84-391 Tage).

Fünf (21,7%) der 23 Proband\*innen wiesen ein CIS auf, 17 (73,9%) eine RRMS und eine\*r (4,3%) eine SPMS, bei einer durchschnittlichen Krankheitsdauer von  $8,1 \pm 6,5$  Jahren.

Die EDSS-Werte dieser 23 Teilnehmer\*innen betragen im Median sowohl zum Baseline-, als auch zum Follow-Up-Zeitpunkt jeweils 1, mit einem IQR = 3 (1. Quartil = 0, 3. Quartil = 3) zum Baseline-Zeitpunkt und einem IQR = 2 (1. Quartil = 0, 3. Quartil = 2) zum Follow-Up-Zeitpunkt.

Nur zwei Patient\*innen (8,7%) erhielten eine verlaufsmodifizierende Therapie während des Studienzeitraumes und bei fünf (21,7%) der 23 Teilnehmer\*innen kam es im MRT-Intervall zu einem einmaligen Schubgeschehen.

Zuletzt wird die Gruppe der 22 Teilnehmer\*innen, bei denen mittels **validierter automatisierter Analyse (AAValid)** neue T2-Läsionen identifiziert wurden, hinsichtlich ihrer klinischen und demographischen Merkmale beschrieben. Jene 22 Proband\*innen waren im Durchschnitt  $35,9 \pm 10,3$  Jahre alt mit einem mittleren Follow-Up-Intervall von  $306 \pm 87$  Tagen (Follow-Up-Intervallspanne 84-391 Tage).

Unter den 22 Proband\*innen fanden sich 14 Frauen (63,6%) und acht Männer (36,4%) mit einer mittleren Krankheitsdauer von  $7,7 \pm 6,2$  Jahren. Die RRMS ist mit 16 Fällen (72,7%) auch in dieser Gruppe am häufigsten vertreten. Fünf Patient\*innen (22,7%) litten an einem CIS und eine\*r (4,5%) an einer SPMS.

Von den 22 Patient\*innen erhielten nur zwei (9,1%) eine verlaufsmodifizierende Therapie im Studienzeitraum und bei fünf (22,7%) kam es zu einem einmaligen Schubgeschehen im MRT-Intervall.

Die EDSS-Werte verhielten sich ähnlich wie in den beiden zuvor beschriebenen Teilnehmer\*innen-Gruppen, mit einem Median von 1 zum Baseline- und Follow-Up-Zeitpunkt und einem Baseline-IQR = 3 (1. Quartil = 0, 3. Quartil = 3), sowie einem Follow-Up-IQR = 2 (1. Quartil = 0, 3. Quartil = 2).

Im Median traten null neue Läsionen bei allen drei Methoden auf. Methode VA und AAValid fanden maximal sieben, Methode AA maximal zehn neue T2-Läsionen in der vorliegenden Studien-Kohorte. Als Streuungsmaß wurde für jede Methode der IQR berechnet. Für alle drei Methoden (VA, AA, AAValid) ergab sich ein IQR=1 (1. Quartil = 0, 3. Quartil = 1), was bedeutet, dass es sich bei 50% der neu detektierten T2-Läsionen um keine oder eine neue Läsion handelt.

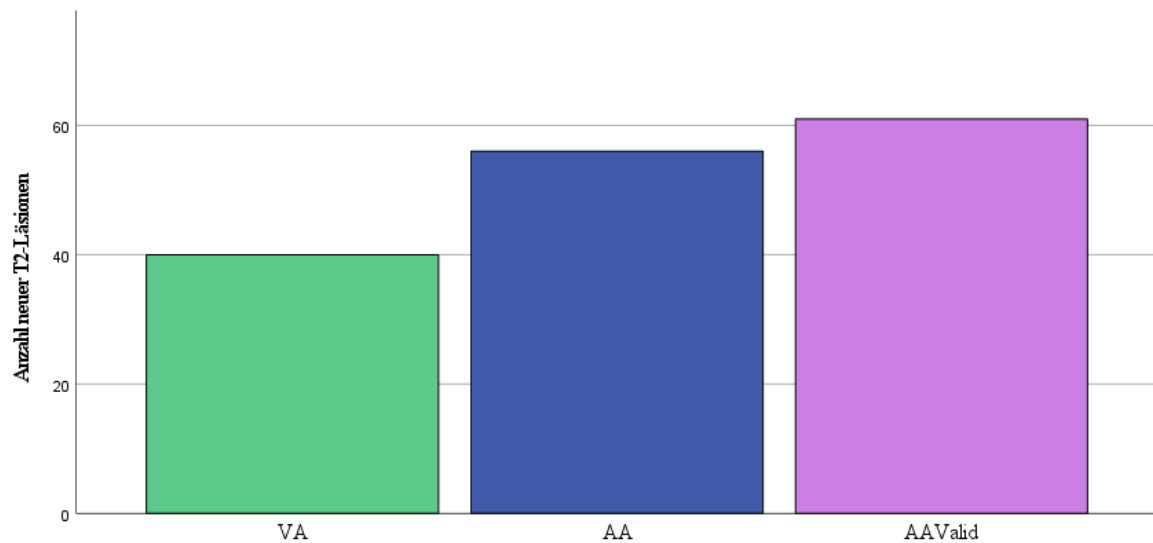
**Tabelle 6: Deskriptive Statistik neu detektierter T2-Läsionen nach Methode über die Studienkohorte (67 Teilnehmer\*innen und 67 MRT-Scan-Paare)**

<b>Absolute Zahl neu detektierter T2-Läsionen</b>					
<b>Methode</b>	<b>Median</b>	<b>IQR</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Spannweite</b>
<b>VA<sup>a</sup></b>	0	1	0	7	7
<b>AA<sup>b</sup></b>	0	1	0	10	10
<b>AAValid<sup>c</sup></b>	0	1	0	7	7
<p><b>a. VA= visuelle Analyse; entspricht Anzahl neuer Läsionen laut neuroradiologischem Bericht</b>  <b>b. AA= automatische Analyse; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms</b>  <b>c. AAValid= automatische Analyse + Validierung; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms und anschließender manueller Validierung</b></p>					

Tabelle 7 und Abbildung 13 dienen als Übersicht der Summe neuer T2-Läsionen, welche je nach Methode identifiziert wurden:

**Tabelle 7: Vergleich der Summe neuer T2-Läsionen nach Methode über die gesamte Studien-Kohorte (67 Teilnehmer\*innen und 67 MRT-Scan-Paare)**

<b>Absolute Zahl neuer T2-Läsionen nach Methode</b>	
<b>Summe neuer Läsionen</b>	
<b>VA<sup>a</sup></b>	40
<b>AA<sup>b</sup></b>	56
<b>AAValid<sup>c</sup></b>	61
<p><b>a. VA= visuelle Analyse; entspricht Anzahl neuer Läsionen laut neuroradiologischem Bericht</b>  <b>b. AA= automatische Analyse; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms</b>  <b>c. AAValid= automatische Analyse + Validierung; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms und anschließender manueller Validierung</b></p>	



VA= visuelle Analyse; entspricht Anzahl neuer Läsionen laut radiologischem Bericht  
 AA= automatische Analyse; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms  
 AAValid= automatische Analyse + Validierung; entspricht Anzahl neuer Läsionen ermittelt durch Copilot.ms und anschließender manueller Validierung

**Abbildung 13: Balkendiagramm zur Anzahl neuer T2-Läsionen nach Methode über die gesamte Studien-Kohorte (67 Teilnehmer\*innen und 67 MRT-Scan-Paare)**

Es lässt sich erkennen, dass mittels visueller Analyse (VA) insgesamt 40, mittels automatisierter Analyse (AA) 56 und mittels automatisierter Analyse und anschließender Validierung (AAValid) 61 neue T2-Läsionen in den Follow-Up-Scans gefunden wurden. Diese Diskrepanz, dass mittels automatisierter Segmentierung plus Validierung (AAValid) mehr Läsionen detektiert wurden als durch die rein automatisierte Analyse (AA), lässt sich dahingehend erklären, dass sich bei Methode AAValid T2-Läsionen zu den automatisiert detektierten addieren, welche durch visuelle Validierung erkannt wurden und durch die automatisierte Analyse unentdeckt blieben. Auf diesen Punkt wird im Kapitel Diskussion näher eingegangen.

### 3.2 Sensitivität, Spezifität, Vorhersagewerte

Um Aussagen über die Güte der jeweiligen Methode zur Läsionssegmentierung machen zu können, wurden die beiden Gütekriterien Sensitivität und Spezifität berechnet. Sie sind als grundlegende Maßstäbe für die diagnostische Genauigkeit zu sehen (70).

Der vorliegende Datensatz musste zur Durchführung von Sensitivitäts- und Spezifitätsberechnungen mittels SPSS umkodiert werden, da Kreuztabellen als Basis der Berechnungen dienen.

Anstatt der absoluten Anzahl der gefundenen Läsionen pro Patient\*in und pro Methode, wurde eine binäre Kodierung gewählt. „0“ steht in der neuen Kodierung für „keine neue Läsion durch die Methode gefunden“, „1“ steht für „mindestens eine neue Läsion durch die Methode gefunden“. Die binäre Kodierung basiert nicht mehr auf der exakten Läsionszahl, sondern lediglich auf der Fähigkeit der Methode neue Läsionen zu erkennen. Vereinfacht gesagt gibt sie auch an, in wie vielen Fällen der Patient\*innen-Kohorte die untersuchte Methode (AA oder AAValid) Läsionen erkannt hat.

Im nächsten Schritt wurden mittels SPSS aus den binär kodierten Daten zwei Kreuztabellen erstellt. Die visuelle Analyse (VA) wurde in beiden Fällen als Goldstandard, welcher als Vergleichswert zur Berechnung der Sensitivität und Spezifität nötig ist, herangezogen.

In der ersten Kreuztabelle werden die automatische Analyse (AA) und die visuelle Analyse (VA) gegenübergestellt, in der zweiten Kreuztabelle die validierte automatische Analyse (AAValid) und die visuelle Analyse (VA).

**Tabelle 8: Kreuztabelle automatische Analyse x visuelle Analyse**

		VA <sup>b</sup>		Gesamt
		0	1	
AA <sup>a</sup>	0	43	1	44
	1	7	16	23
Gesamt		50	17	67

0= keine neue Läsion gefunden  
1= mindestens eine neue Läsion gefunden  
a. AA= automatische Analyse  
b. VA= visuelle Analyse = GOLDSTANDARD

**Tabelle 9: Kreuztabelle validierte automatische Analyse x visuelle Analyse**

		VA <sup>b</sup>		Gesamt
		0	1	
AAValid <sup>a</sup>	0	44	1	45
	1	6	16	22
Gesamt		50	17	67

0= keine neue Läsion gefunden  
1= mindestens eine neue Läsion gefunden  
a. AAValid= validierte, automatische Analyse  
b. VA= visuelle Analyse = GOLDSTANDARD

Anhand beider Kreuztabellen lassen sich nun für die Methoden AA und AAValid Sensitivität und Spezifität berechnen, da sie folgende wichtige Kennwerte enthalten:

- falsch positives Ergebnis (FP): eine neue Läsion wurde von der Methode erkannt, ohne tatsächliches Vorliegen einer neuen Läsion
- falsch negatives Ergebnis (FN): eine neue Läsion liegt vor, diese wurde aber von der Methode nicht erkannt
- richtig positives Ergebnis (RP): eine neue Läsion wurde korrekt als solche identifiziert
- richtig negatives Ergebnis (RN): es liegen keine neuen Läsionen vor und es wurden von der Methode keine Läsionen erkannt

Sensitivität und Spezifität werden anhand folgender Formeln berechnet (44):

$$\text{Sensitivität} = \frac{RP}{RP+FN}$$

$$\text{Spezifität} = \frac{RN}{RN+FP}$$

Die Berechnungen von Sensitivität und Spezifität wurden mithilfe von MedCalc (MedCalc Software Ltd., Ostend, Belgium, [https://www.medcalc.org/calc/diagnostic\\_test.php](https://www.medcalc.org/calc/diagnostic_test.php)) durchgeführt. Sensitivität und Spezifität werden mit einem 95%-Konfidenzintervall angegeben, wobei MedCalc „exakte“ Clopper-Pearson-Konfidenzintervalle berechnet.

MedCalc wurde außerdem zur Berechnung des positiven und negativen Vorhersagewertes herangezogen. Hierbei handelt es sich um statistische Größen, die allgemein die Wahrscheinlichkeit ermitteln, mit welcher ein diagnostischer Test die korrekte Aussage, z.B. Diagnose, liefert (71). Dies lässt sich auf diese Diplomarbeit folgendermaßen umlegen, als dass der positive Vorhersagewert (PPV = positive predictive value) die Wahrscheinlichkeit angibt, dass eine neue Läsion vorliegt, wenn diese durch eine der Methoden erkannt wird. Umgekehrt gibt der negative Vorhersagewert (NPV = negative predictive value) die Wahrscheinlichkeit an, dass keine neue Läsion vorliegt, wenn die jeweilige Methode keine identifizieren konnte.

Beide Vorhersagewerte werden prävalenzabhängig nach folgenden Formeln ermittelt (71):

$$\text{PPV} = \frac{\text{Sensitivität} + \text{Prävalenz}}{(\text{Sensitivität} * \text{Prävalenz}) + (1 - \text{Spezifität}) * (1 - \text{Prävalenz})}$$

$$NPV = \frac{\text{Spezifität} \cdot (1 - \text{Prävalenz})}{(1 - \text{Sensitivität}) \cdot \text{Prävalenz} + \text{Spezifität} \cdot (1 - \text{Prävalenz})}$$

Da es sich bei der vorliegenden Studienpopulation um Patient\*innen des Grazer MS-Zentrums an der Universitätsklinik für Neurologie handelt, wurde die von Salhofer-Polanyi et al. ermittelte Prävalenz der Multiplen Sklerose für Österreich von 158,9 pro 100 000 Einwohner\*innen herangezogen (8). Bei dem von MedCalc erstellten 95%-Konfidenzintervall des Vorhersagewertes handelt es sich um ein Logit-Konfidenzintervall nach Mercaldo et al. 2007.

**Tabelle 10: Leistung der automatischen und validierten automatischen Methode in der Läsionserkennung**

	AA	AAValid
<i>RP</i>	16	16
<i>RN</i>	43	44
<i>FP</i>	7	6
<i>FN</i>	1	1
<i>Sensitivität</i>	94,12%	94,12%
<i>(95%-Konfidenzintervall)</i>	(71,31-99,85%)	(71,31-99,85%)
<i>Spezifität</i>	86,00%	88,00%
<i>(95%-Konfidenzintervall)</i>	(73,26-94,18%)	(75,69-95,47%)
<i>Positiver Vorhersagewert</i>	1,06%	1,23%
<i>(95%-KI)</i>	(0,53-2,10%)	(0,58-2,60%)
<i>Negativer Vorhersagewert</i>	99,99%	99,99%
<i>(95%-Konfidenzintervall)</i>	(99,93-100,00%)	(99,93-100,00%)

Tabelle 10 gibt einen Überblick über die vier wichtigen Kenngrößen RP, RN, FP, FN, welche aus den Kreuztabellen in Tabelle 8 und Tabelle 9 ermittelt wurden.

Für Methode AA konnten eine Sensitivität von 94,12% (71,31-99,85%) und eine Spezifität von 86,00% (73,26- 94,18%) berechnet werden. Methode AAValid besitzt eine Sensitivität von 94,12% (71,31-99,85%) und eine Spezifität von 88,00% (75,69-95,47%). Somit besitzen die beiden untersuchten Methoden eine übereinstimmende Sensitivität und unterscheiden sich lediglich in der Spezifität.

Der negative Vorhersagewert beider Methoden liegt bei 99,99% (99,93-100,00%) und zeigt somit ebenfalls übereinstimmende Werte. Der positive Vorhersagewert der Methode AA liegt bei 1,06% (0,53-2,10%). Jener der Methode AAValid liegt bei 1,23% (0,58-2,60%).

### 3.3 Test zum Vergleich der Methoden

Im Fall dieser Studie wurde ein Vorzeichentest als passendes Testverfahren ausgewählt, wenngleich dieser zur Beantwortung der Fragestellung gewisse Limitationen aufweist, welche im Kapitel Diskussion erläutert werden.

Der Vorzeichentest wird als Testverfahren für zwei abhängige Stichproben mit nicht normalverteilten Variablen herangezogen. Da die erhobenen und statistisch verarbeiteten Daten dieser Studie von denselben 67 Studienteilnehmer\*innen stammen und auf demselben MRT-Scanner durchgeführt wurden, wird mit verbundenen bzw. abhängigen Stichproben gearbeitet.

Mittels Shapiro-Wilk-Test und Kolmogorov-Smirnov-Test wurde überprüft, ob die Variablen der abhängigen Stichproben, sowie die Differenzen der Variablen normalverteilt sind. Beide Tests auf Normalverteilung arbeiten mit der Nullhypothese, dass die überprüften Daten normalverteilt sind. Ein p-Wert mit einem Signifikanzniveau  $< 0,05$  bedeutet, dass die Nullhypothese abgelehnt wird und die überprüfte Stichprobe keine Normalverteilung der Variablen zeigt. Sowohl für die Variablen dieser Studie, als auch deren Differenzen ergab sich ein  $p < 0,001$  im Shapiro-Wilk- und Kolmogorov-Smirnov-Test, wodurch die Nullhypothese der Normalverteilung abgelehnt wurde, die Stichproben nicht symmetrisch verteilt sind und der Vorzeichentest als nicht-parametrischer Test gewählt wurde. Er stellt die nicht-parametrische Alternative für den t-Test für abhängige Stichproben dar.

Mittels Vorzeichentest wurde überprüft, ob sich die zentralen Tendenzen der abhängigen Stichproben der Methoden VA und AA sowie VA und AAValid unterscheiden. Ein Signifikanzniveau von  $p < 0,05$  wurde als statistisch signifikant angesehen. Die Nullhypothese eines Vorzeichentests besagt, dass sich beide Methoden in ihren zentralen Tendenzen nicht unterscheiden, wodurch sich für die vorliegende Studie folgende Null- und Alternativhypothesen ergeben:

**Tabelle 11: Null- und Alternativhypothesenübersicht des durchgeführten Vorzeichentests**

<b>Vorzeichentest<sup>1</sup> für Methode VA und AA</b>	
<b>Nullhypothese<sup>1</sup></b>	Methode VA = Methode AA
<b>Alternativhypothese<sup>1</sup></b>	Methode VA ≠ Methode AA
<b>Vorzeichentest<sup>2</sup> für Methode VA und AAValid</b>	
<b>Nullhypothese<sup>2</sup></b>	Methode VA = Methode AAValid
<b>Alternativhypothese<sup>2</sup></b>	Methode VA ≠ Methode AAValid

Somit entsprechen die beiden Nullhypothesen der Forschungshypothese dieser Studie, dass die manuelle Segmentierung (VA) und die automatische Segmentierung mittels copilot.ms (AA) bzw. die Expert\*innen-validierte automatische Segmentierung mittels copilot.ms (AAValid) gleiche Ergebnisse hinsichtlich der absoluten Anzahl neu detektierter T2-Läsionen erzielen. Die beiden Alternativhypothesen besagen, dass sich die beiden Methoden in ihren zentralen Tendenzen unterscheiden und sie somit keine gleichen Ergebnisse hinsichtlich der absoluten Anzahl neu detektierter T2-Läsionen erzielen.

Folgende Tabellen sollen zur Übersicht über die Ergebnisse der beiden Vorzeichentests dienen:

**Tabelle 12: Häufigkeit der errechneten Paardifferenzen mittels Vorzeichentest**

<b>Häufigkeiten</b>		
		N= Anzahl
<b>Vorzeichentest<sup>1</sup></b> Methode VA - Methode AA	Negative Differenzen <sup>a,d</sup>	14
	Positive Differenzen <sup>b,e</sup>	6
	Bindungen <sup>c,f</sup>	47
	Gesamt	67
<b>Vorzeichentest<sup>2</sup></b> Methode VA - Methode AAValid	Negative Differenzen <sup>a,d</sup>	14
	Positive Differenzen <sup>b,e</sup>	3
	Bindungen <sup>c,f</sup>	50
	Gesamt	67
a. Detektierte T2-Läsionen VA < Detektierte T2-Läsionen AA b. Detektierte T2-Läsionen VA > Detektierte T2-Läsionen AA c. Detektierte T2-Läsionen VA = Detektierte T2-Läsionen AA d. Detektierte T2-Läsionen VA < Detektierte T2-Läsionen AAValid e. Detektierte T2-Läsionen VA > Detektierte T2-Läsionen AAValid f. Detektierte T2-Läsionen VA = Detektierte T2-Läsionen AAValid		

Der Vorzeichentest berechnet für jeden Fall die Differenzen zwischen den beiden Variablen und klassifiziert jene als positiv, negativ oder verbunden.

In Tabelle 11 lässt sich die Anzahl der positiven und negativen Paardifferenzen, sowie der Bindungen ablesen. Vereinfacht gesagt bedeutet dies für den **Vorzeichentest<sup>1</sup> (Vergleich der Methoden VA und AA)**, dass in 14 Fällen mit der Methode VA weniger T2-Läsionen als mit der Methode AA, in 6 Fällen mit der Methode VA mehr T2-Läsionen als mit der Methode AA und in 47 Fällen gleich viele T2-Läsionen anhand beider Methoden detektiert wurden. Analog dazu gilt für den **Vorzeichentest<sup>2</sup> (Vergleich der Methoden VA und AAValid)**, dass in 14 Fällen mit der Methode VA weniger T2-Läsionen als mit der Methode AAValid, in 3 Fällen mit der Methode VA mehr T2-Läsionen als mit der Methode AAValid und in 50 Fällen exakt gleich viele T2-Läsionen anhand beider Methoden detektiert wurden.

**Tabelle 13: Teststatistik der Vorzeichentests**

<b>Teststatistiken der Vorzeichentests</b>		
	<b>Vorzeichentest<sup>1</sup></b> Methode VA - Methode AA	<b>Vorzeichentest<sup>2</sup></b> Methode VA - Methode AAValid
Exakte Signifikanz (2-seitig)	,115	,013

Tabelle 12 fasst die Teststatistiken der beiden durchgeführten Vorzeichentests zusammen. Ein Signifikanzniveau von  $p < 0,05$  gilt als statistisch signifikant, Somit wird bei einem Signifikanzniveau von  $p < 0,05$  die Nullhypothese verworfen.

Der **Vorzeichentest<sup>1</sup> (Vergleich der Methoden VA und AA)** ergab einen p-Wert von 0,115 und somit keinen statistisch signifikanten Unterschied der zentralen Tendenzen beider Methoden. Die Nullhypothese wird beibehalten. Für den **Vorzeichentest<sup>2</sup> (Vergleich der Methoden VA und AAValid)** errechnete sich ein p-Wert von 0,013, welcher kleiner als das vorgegebene Signifikanzniveau und somit einen statistisch signifikanten Unterschied zwischen den beiden verglichenen Methoden anzeigt. Die Nullhypothese, dass Gleichheit zwischen den beiden Methoden besteht, kann somit verworfen werden.

## 4 Diskussion

Die Multiple Sklerose ist eine chronische Autoimmunerkrankung, welche das Zentralnervensystem betrifft und mit weltweit ca. 2,3 Mio. Erkrankten als eine der häufigsten Ursachen neurologischer Beeinträchtigung junger Erwachsener gilt (1). Die Magnetresonanztomographie stellt das wichtigste paraklinische Diagnose- und Monitoringwerkzeug der Erkrankung dar (35). Krankheitsaktivität lässt sich mit der MRT sehr gut anhand neuer oder sich vergrößernder T2-Läsionen durch den Vergleich mit früheren, bestenfalls innerhalb von  $\leq 1$  Jahr durchgeführten MRT-Aufnahmen erkennen (3). Da die Krankheitsaktivität ein relevanter Entscheidungsfaktor in der Therapieauswahl bzw. -adaptation der MS ist, hat die Läsionssegmentierung in MRT-Aufnahmen nicht nur große Bedeutung für Diagnostik, Verlaufsuntersuchungen, und Monitoring der Erkrankung, sondern vor allem auch für das Krankheitsmanagement (2).

Die bisherige, standardmäßig angewandte manuelle Läsionssegmentierung, welche in der Praxis meist mittels Augenmaß durch Radiolog\*innen ohne neuroradiologische Spezialisierung aufgrund fehlender flächendeckender Verfügbarkeit erfahrener Neuroradiolog\*innen durchgeführt wird, ist mit klaren Einschränkungen verbunden. Durch große zu analysierende MRT-Datenmengen ist sie zeitaufwändig, erfordert Expertise und ist aufgrund der subjektiven Beurteilung durch Menschenhand starken Sensitivitätsschwankungen unterlegen (43–45). Da wichtige Entscheidungen vor allem hinsichtlich der Therapie von diesen Befunden abhängen, muss der Trend zu objektivierbaren und reproduzierbaren Methoden gehen. Es lässt sich erkennen, dass eine automatisierte Läsionssegmentierung immer bedeutsamer wird.

Die Entwicklung eines solchen Tools zur automatisierten Läsionsanalyse ist Gegenstand aktueller Forschung. In dieser Diplomarbeit wurde in Zusammenarbeit mit dem Hospital Universitari Vall d'Hebron in Barcelona der Prototyp eines von dem Spin-Off-Unternehmen Tensor Medical entwickelten entscheidungsunterstützenden Systems, genannt copilot.ms, zur automatisierten Läsionssegmentierung auf Basis von MRT-Aufnahmen bei MS an einer Kohorte von Patient\*innen des Grazer MS-Zentrums getestet.

Ziel war es, die Fragestellung zu beantworten, ob es mittels automatischer Segmentierung von T2-gewichteten MRT-Aufnahmen durch copilot.ms und anschließendem Vidieren durch Expert\*innen möglich ist, gleiche Genauigkeit in der Erkennung von neuen T2-Läsionen bei MS-Patient\*innen zu erhalten wie durch alleinige manuelle Segmentierung.

Ließe sich gleiche Genauigkeit durch automatisierte, validierte Analyse erreichen, würde eine automatische Vorauswertung den klaren Vorteil der zeitlichen Effizienz aufweisen.

Die Ergebnisse dieser Diplomarbeit zeigten Unterschiede zwischen der konventionellen visuellen Analyse und der automatischen, sowie validiert automatisierten Analyse. Die automatische und validiert automatische Analyse fanden insgesamt mehr neue T2-Läsionen und hatten ein geringeres Risiko radiologisch aktive Patient\*innen zu übersehen als die manuelle Segmentierung. Zugleich zeigten Methode AA und AAValid zum aktuellen Zeitpunkt noch Einschränkungen in der zuverlässigen Detektion neuer T2-Läsionen und erzielten lediglich bei stabilen Patient\*innen zuverlässige Ergebnisse.

Insgesamt wurden 67 Studienteilnehmer\*innen mit jeweils 67 MRT-Scan-Paaren untersucht. Davon traten bei 17 Teilnehmer\*innen (25,3%) mittels visueller Analyse (Methode VA) insgesamt 40 neue T2-Läsionen, bei 23 Teilnehmer\*innen (34,3%) mittels automatischer Analyse (Methode AA) insgesamt 56 neue T2-Läsionen und bei 22 Teilnehmer\*innen (32,8%) mittels validierter automatisierter Analyse (Methode AAValid) insgesamt 61 neue T2-Läsionen im Follow-Up-Scan auf. Im Median traten null neue Läsionen bei allen drei Methoden auf, mit einem IQR=1. Somit handelte es sich bei 50% der neu detektierten Läsionen um keine oder eine neue Läsion.

Es zeigt sich somit, dass durch die konventionelle visuelle Analyse (manuelle Segmentierung) weniger Läsionen detektiert wurden, als durch die automatische bzw. die validierte automatische Analyse. Eine mögliche Erklärung hierfür ist der Unterschied im Setting der Analysen. Die konventionelle visuelle Analyse findet unter Zeitdruck im klinischen Alltag statt, wodurch die Sensitivität in der Läsionserkennung aufgrund mangelnder zeitlicher Ressourcen, aber auch fehlender personeller Ressourcen (mangelnde flächendeckende Verfügbarkeit erfahrener Neuroradiolog\*innen) sinken kann. Die validierte automatische Analyse durch Expert\*innen im Rahmen der Studie fand ohne jenen Zeitdruck mit dem einzigen Ziel der Identifizierung neuer T2-Läsionen statt, was eingehendere Analysen ermöglichte. Es lässt sich außerdem eine Diskrepanz in der detektierten Läsionszahl zwischen automatischer und validiert automatischer Analyse erkennen. Dies ergibt sich durch den Validierungsschritt. Durch erneute visuelle Expert\*innen-Analyse in dem bereits erwähnten Setting ohne Zeitdruck nach zuvor erfolgter automatischer Vorauswertung mittels copilot.ms, wird eine neuerliche Analyse der vorgefilterten MRT-Daten ermöglicht, was die Chancen, neue Läsionen zu übersehen,

minimiert und wodurch im Validierungsschritt durch die automatische Analyse als falsch negativ erkannte Läsionen korrigiert werden können, wodurch die insgesamt detektierte Läsionszahl steigt.

Die klinischen und demografischen Merkmale jener Teilnehmer\*innen, die je nach Methode neue T2-Läsionen aufwiesen, sind vergleichbar. Bei allen drei Methoden waren ca. zwei Drittel der Patient\*innen, bei denen neue T2-Läsionen detektiert wurden, weiblich. In allen drei Gruppen war die RRMS der am häufigsten auftretende Phänotyp. In jeder der drei Gruppen erhielten jeweils nur zwei der Teilnehmer\*innen mit neuen T2-Läsionen eine verlaufsmodifizierende Therapie. Das mittlere Follow-Up-Intervall (Methode VA  $286 \pm 95$  Tage, Methode AA  $308 \pm 86$  Tage, Methode AAValid  $306 \pm 87$ ) ist bei den Methoden AA und AAValid vergleichbar, bei der Methode VA allerdings etwas kürzer. Im Schnitt wiesen die Patient\*innen mit neuen T2-Läsionen ein ähnliches Alter (Methode VA  $36,0 \pm 9,8$  Jahre, Methode AA  $36,5 \pm 10,6$  Jahre, Methode AAValid  $35,9 \pm 10,3$  Jahre) und eine ähnliche Krankheitsdauer (Methode VA  $7,8 \pm 5,7$  Jahre, Methode AA  $8,1 \pm 6,5$  Jahre, Methode AAValid  $7,7 \pm 6,2$  Jahre) auf. Im Median ließ sich bei Teilnehmer\*innen mit neu aufgetretenen Läsionen bei allen drei Methoden ein geringer EDSS-Wert von 1 sowohl zum Baseline- als auch Follow-Up-Zeitpunkt beobachten (IQR-Baseline=3, IQR-Follow-Up= 2), was im Median auf einen geringen Grad der funktionellen Beeinträchtigung in der untersuchten Studienkohorte mit neuen T2-Läsionen schließen lässt.

Zur Berechnung der diagnostischen Genauigkeit der automatischen bzw. validiert automatischen Analyse wurden Sensitivität, Spezifität und Vorhersagewerte berechnet. Als Referenzwert zur Erstellung der Vierfeldertafel wurde der aktuelle Goldstandard, die manuelle Segmentierung (Methode VA), herangezogen. Eine Sensitivitäts-/Spezifitätsberechnung für die manuelle Segmentierung (Methode VA) wurde in dieser Studie nicht durchgeführt, da hierfür ein „Referenz-Goldstandard“ erstellt werden hätte müssen, um einen Vergleich zwischen Methode VA und „Referenz-Goldstandard“ zu ermöglichen. Da sich der Fokus dieser Studie allerdings auf die Performance der automatischen bzw. validiert automatischen Analyse richtete und zum aktuellen Zeitpunkt kein anderer Goldstandard als die manuelle Segmentierung existiert, wurde dies nicht durchgeführt. Es wurden die Gütekriterien Sensitivität und Spezifität somit nur für Methode AA und AAValid berechnet.

Methode AA fand insgesamt mehr falsch positive Ergebnisse als Methode AAValid. Durch den Validierungsschritt gelang es also, als falsch positiv identifizierte Ergebnisse und damit fälschlicherweise als radiologisch aktiv identifizierte Teilnehmer\*innen auszusortieren und entsprechend zu berichtigen. Dies zeigt die Relevanz des Validierungsschrittes. Insgesamt war der Anteil als falsch positiv detektierter Ergebnisse bei beiden Methoden relativ hoch (7 FP bei Methode AA; 6 FP bei Methode AAValid). Somit besteht wie bereits angeführt in beiden Situationen die Problematik, dass in 6 bzw. 7 Fällen radiologisch inaktive Patient\*innen als fälschlicherweise radiologisch aktiv erkannt wurden.

Hinsichtlich der Anzahl an falsch negativen Ergebnissen lieferten beide Methoden überschneidende Ergebnisse mit jeweils nur einer falsch negativ detektierten Läsion.

Die angeführten Überlegungen spiegeln sich in weiterer Folge auch in Sensitivität und Spezifität der Methoden wider. Hinsichtlich der Sensitivität schnitten die automatische und die validiert automatische Analyse mit 94,12% (95%-KI 71,31-99,85%) identisch ab. Aufgrund der geringen Anzahl falsch negativer Ergebnisse lässt sich somit eine hohe Sensitivität erzielen. Der negative Vorhersagewert ist mit 99,99% (95%-KI 99,93-100,00%) für beide Methoden sehr hoch. Es lässt sich daraus schließen, dass sich beide Methoden sehr gut dazu eignen stabile Patient\*innen, also jene, wo tatsächlich keine neue Läsion vorliegt, wenn die Methode keine identifizieren konnte, zu erkennen.

Die Spezifität fiel mit 86,00% (95%-KI 73,26-94,18%) bei Methode AA und 88,00% (95%-KI 75,69-95,47%) bei Methode AAValid geringer als die Sensitivität aus. Die etwas höhere Spezifität bei Methode AAValid ergibt sich durch die geringere Anzahl falsch positiver Ergebnisse.

Die positiven Vorhersagewerte und somit die Wahrscheinlichkeit, dass tatsächlich eine neue Läsion vorliegt, wenn sie durch eine Methode als solche erkannt wurde, fallen bei beiden Methoden sehr gering aus (Methode AA 1,06% (95%-KI 0,53-2,10%), Methode AAValid 1,23% (95%-KI 0,58-2,60%)). Dies ergibt sich unter anderem durch den Einfluss der Prävalenz, welche in dieser Studie mit 158,9 pro 100 000 Einwohner\*innen anhand der von Salhofer-Polanyi et al. für Österreich ermittelten Prävalenz für MS in Österreich festgelegt wurde (8). Daraus ergibt sich eine prozentuelle Prävalenz von 0,1589%, welche den positiven Vorhersagewert dahingehend beeinflusst, dass bei einer geringen Prävalenz die Gruppe der Nicht-aktiven-Patient\*innen unter allen Untersuchten größer ist, wodurch

relativ gesehen mehr falsch positive Ergebnisse auftreten können, was zu einem geringeren positiven Vorhersagewert führt. Nichtsdestotrotz zeigen die geringen positiven Vorhersagewerte beider Methoden in dieser Studie die Einschränkungen der automatischen und validiert automatischen Analyse auf. Aufgrund der geringen Wahrscheinlichkeit eine neue Läsion korrekterweise als solche zu erkennen, eignen sich die beiden Methoden aktuell noch nicht ausreichend, um Patient\*innen im klinischen Setting zuverlässig hinsichtlich ihrer radiologischen Aktivität anhand neu aufgetretener T2-Läsionen zu detektieren. Das Risiko Patient\*innen als fälschlicherweise radiologisch aktiv zu deklarieren ist zu groß. Da dies eine mögliche Therapieveränderung als Folge hätte, welche jedoch nicht indiziert wäre, ist die Anwendungssicherheit nicht ausreichend gegeben. Zu beachten ist, dass die angewandte Prävalenz nur gültig ist, wenn die MRT-Aufnahmen einer beliebigen Person in Österreich mit Methode AA oder Methode AAValid untersucht werden. Da das primäre Ziel von copilot.ms der Einsatz im Krankheitsmonitoring der MS, bei bereits diagnostizierten Patient\*innen ist, wodurch eine Gruppe von Personen mit gruppenspezifisch höherer Prävalenz der Erkrankung untersucht wird, lässt sich der berechnete positive Vorhersagewert nur eingeschränkt umlegen.

Zusammenfassend lässt sich hinsichtlich der Gütekriterien sagen, dass beide Methoden eine hohe Sensitivität aufweisen und nur wenige falsch negative Ergebnisse erzielen, was bedeutet, dass nur wenige neue Läsionen mittels automatischer oder validiert automatischer Analyse übersehen werden. Zugleich besitzen beide Methoden eine hohe Zahl falsch positiver Ergebnisse und einen geringen positiven Vorhersagewert, wodurch sich beide Methoden zum aktuellen Zeitpunkt nicht uneingeschränkt zur sicheren Detektion neuer Läsionen eignen, da die Wahrscheinlichkeit gering ist, dass es sich bei durch die Methoden gefundenen Läsionen tatsächlich um neue T2-Läsionen handelt. Durch einen Validierungsschritt und keine vollautomatische Analyse lässt sich die Zahl falsch positiver Ergebnisse reduzieren, weshalb ein Validierungsschritt zur Verbesserung der Spezifität sinnvoll erscheint. Dies entspräche auch der Grundidee des Einsatzes von copilot.ms, da dieses als entscheidungsunterstützendes System fungieren soll und somit einen Validierungsschritt benötigt. Aufgrund des prozentuell hohen negativen Vorhersagewertes eignen sich die beiden untersuchten Methoden jedoch sehr gut, um stabile Patient\*innen, ohne neu aufgetretene Läsionen, zuverlässig als solche zu identifizieren.

Zusätzlich zu den ermittelten Gütekriterien wurden die automatische und die validiert automatische Methode mit dem Goldstandard, der visuellen Analyse (Methode VA), mittels geeignetem Testverfahren verglichen. Es wurde aufgrund der gegebenen Stichprobe und Datenvariablen der Vorzeichentest als einziger, die Voraussetzungen erfüllender Test gewählt. Vorab ist zu erwähnen, dass dieser gewisse Limitationen aufweist. Anhand des Vorzeichentests können lediglich die zentralen Tendenzen einer abhängigen Stichprobe verglichen werden. Das Ausmaß des Unterschiedes, sollte ein Unterschied vorliegen, kann damit nicht getestet werden. Es wurden in beiden Fällen die Nullhypothesen geprüft, dass sich die Methoden AA oder AAValid nicht von der Methode VA unterscheiden und somit gleichwertig eingesetzt werden können, was zur Beantwortung der Forschungsfrage nützen soll, da somit die Nullhypothese der Forschungshypothese dieser Studie entspricht. Der Vorzeichentest für Methode VA und AA zeigte mit  $p = 0,015$  (Signifikanzniveau  $p < 0,05$ ) keinen signifikanten Unterschied zwischen den beiden Methoden. Im Kontext der Studie lässt sich daraus schließen, dass die beiden Methoden gleiche Ergebnisse hinsichtlich der absolute Anzahl neu detektierter T2-Läsionen erzielen.

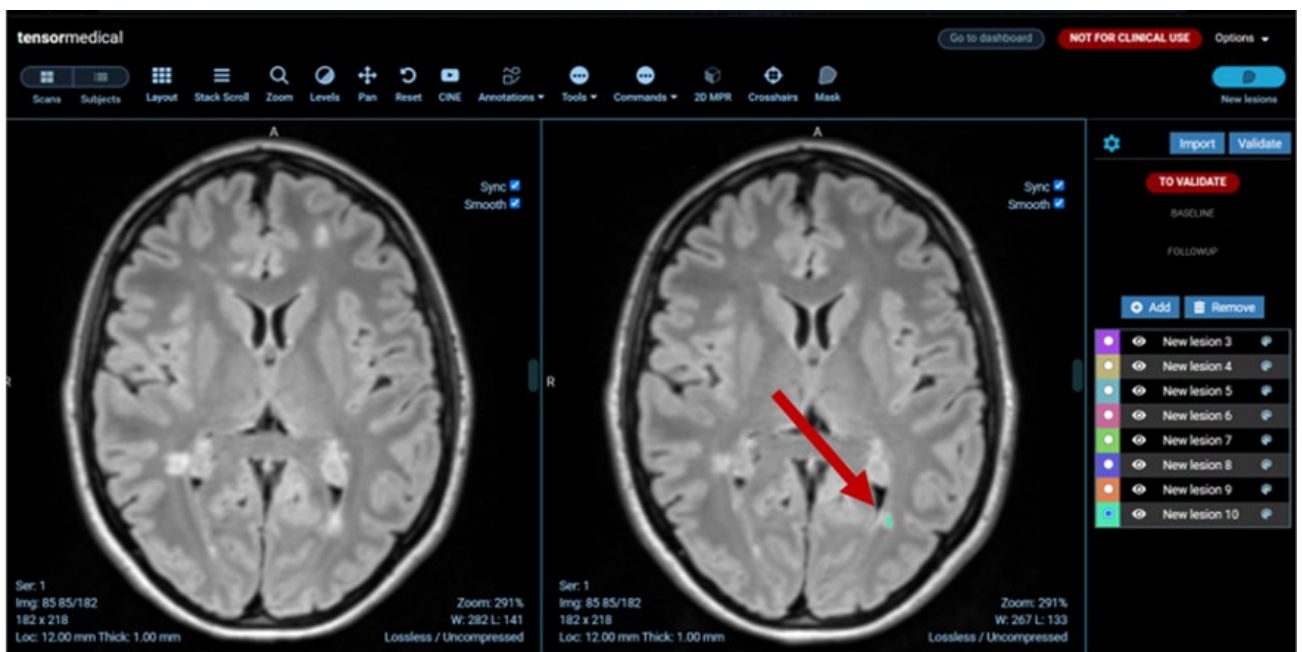
Für den Vergleich der Methoden VA und AAValid ergab der Vorzeichentest ein  $p = 0,013$  (Signifikanzniveau  $p < 0,05$ ) und somit einen statistisch signifikanten Unterschied zwischen beiden Methoden. Sie liefern somit keine vergleichbaren Ergebnisse angesichts der neu gefundenen Läsionen. In welchem Ausmaß dieser Unterschied vorhanden ist lässt sich jedoch, wie eingangs erwähnt, nicht ermitteln.

Somit lässt sich letztlich sagen, dass anhand des Vorzeichentests nur eine Gleichheit der Methode AA im Vergleich zu VA gezeigt werden konnte, was in Anbetracht der Forschungsfrage den Schluss nahelegt, dass nur Methode AA und VA gleichwertig eingesetzt werden können. Dieser Schluss ist nicht vollständig richtig, da die zuvor angeführten Gütekriterien zur Bewertung der Performance der jeweiligen Methode von großer Bedeutung sind und nicht außer Acht gelassen werden dürfen. Bezieht man deren Ergebnisse ein, zeigt sich ein Vorteil der Methode AAValid gegenüber der Methode AA. Die Beantwortung der Forschungsfrage kann somit nur in Zusammenschau des Tests und der errechneten Gütekriterien erfolgen.

Aus den angeführten Überlegungen zeigt sich eine klare Stärke des entscheidungsunterstützenden Systems in der hohen Sensitivität. Das Risiko Läsionen und damit radiologisch aktive Patient\*innen zu übersehen wird, wenn auch auf Kosten der

Spezifität, verringert, was in der klinischen Praxis von Relevanz ist, da so Patient\*innen, welche möglicherweise Therapieanpassungen benötigen würden, identifiziert werden können. Weiters spricht die zeitliche Effizienz für den Einsatz eines entscheidungsunterstützenden Systems. Trotz Validierungsschritt, welcher bei zuverlässiger Performance des Programmes nur wenige Minuten benötigt, würde die klinische Arbeit mit einem entscheidungsunterstützenden System Zeit- und Ressourceneffizienter erfolgen können.

Im Rahmen dieser Studie fielen Limitationen in der Anwendbarkeit der geprüften Version von copilot.ms auf. Eine dieser Limitationen ist die eingeschränkte Erkennung sich vergrößernder Läsionen durch copilot.ms. In einigen Fällen wurden in der Follow-Up-Aufnahme sich vergrößert dargestellte Läsionen, welche bereits in der Baseline-Untersuchung vorhanden waren, als neue T2-Läsionen erkannt. Zudem wurden in mehreren Fällen vorbestehende Läsionen als neu aufgetretene Läsionen detektiert, was mit der zuvor angeführten Anzahl falsch positiver Ergebnisse korreliert. Dies soll anhand von folgendem Beispiel dargestellt werden:



**Abbildung 14: Vorbestehende Läsion fälschlicherweise als neue Läsion erkannt**

Eine weitere Limitation von copilot.ms ist die fehlende Erkennung von „shrinking lesions“ im Verlauf. Im klinischen Alltag werden nicht nur neue oder sich vergrößernde Läsionen bei der Multiplen Sklerose, sondern auch sich verkleinernde oder verschwindende Läsionen beobachtet. Diese Läsionen werden als „shrinking lesions“ bezeichnet und in der

klinischen Routine bei der manuellen Segmentierung berücksichtigt. Der prognostische Wert und die genaue Bedeutung dieser Läsionen für den Krankheitsverlauf sind bis dato nicht ausreichend geklärt und Gegenstand aktueller Forschung (72). Nichtsdestotrotz wäre es für den klinischen Einsatz von copilot.ms von Interesse, die Funktion der Detektion von „shrinking lesions“ zu implementieren. Folgende Abbildungen sollen diese Limitation beispielhaft darstellen:

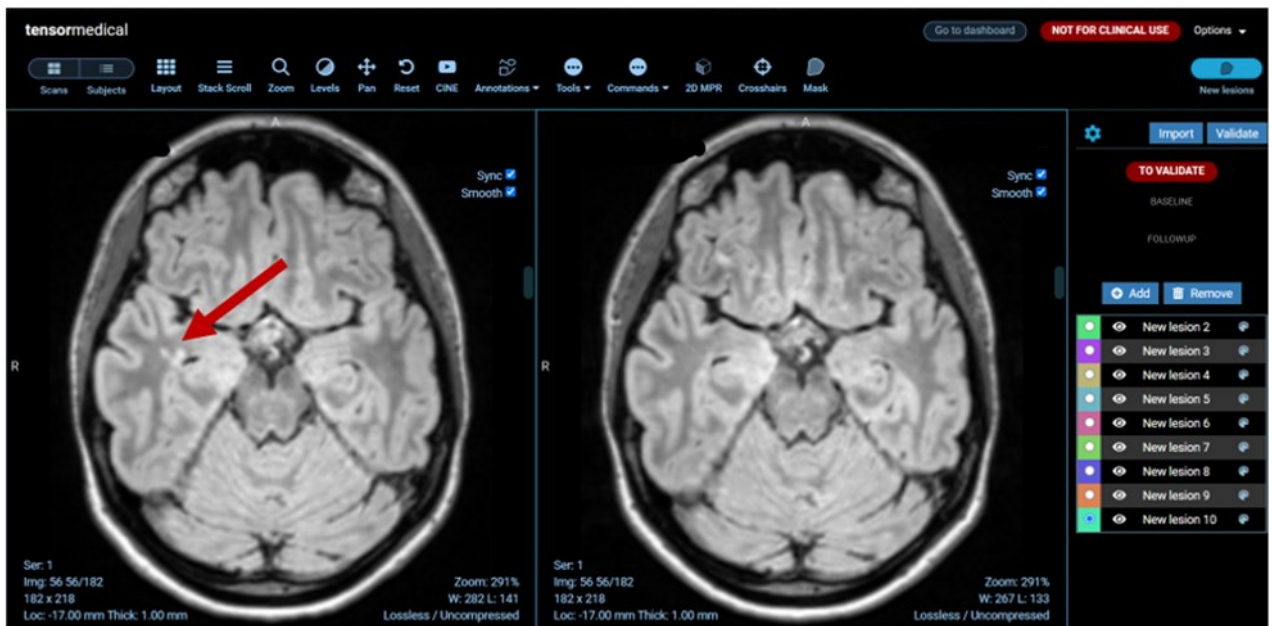


Abbildung 15: Beispiel für nicht erkannte shrinking lesion

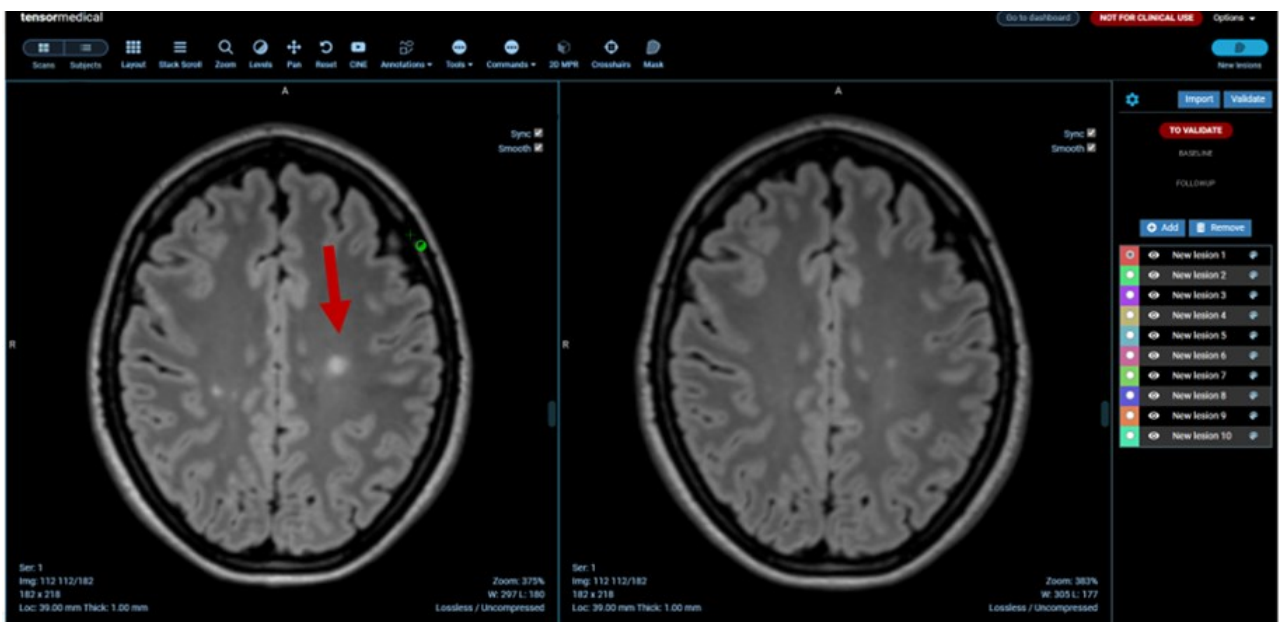


Abbildung 16: Beispiel für nicht erkannte shrinking lesion

Der niedrige positive Vorhersagewert ist ebenfalls als Limitation des untersuchten entscheidungsunterstützenden Systems zu sehen, da es somit seine erwartete Aufgabe, neue T2-Läsionen gleich zuverlässig wie neuroradiologische Expert\*innen zu detektieren, nicht ausreichend erfüllt. Weiters limitierend für die uneingeschränkte Anwendung von Copilot.ms ist, dass es keine für das Sicherheitsmonitoring der Erkrankung relevanten Aspekte, wie beispielsweise das Auftreten einer PML erkennt. Zudem werden keine Komorbiditäten, wie Neoplasien, Infarkte oder Aneurysmen erfasst, was als Lücke im Sicherheitsmonitoring gesehen werden kann. Zu bedenken ist außerdem, dass es sich bei Copilot.ms um kein diagnostisches Tool handelt, wodurch das Risiko der Falschanwendung besteht, da das System MS-Läsionen zum aktuellen Zeitpunkt nicht von mikroangiopathischen Läsionen im Gehirn, oder zerebralen Läsionen einer MOGAD oder einer NMOSD unterscheiden kann.

Zuletzt sollte der Aspekt des Datenschutzes erwähnt werden. In der Verarbeitung von Patient\*innen-Daten dieser Studie wurde grundlegend das Datenschutzgesetz angewandt und Patient\*innen-Daten ausschließlich pseudonymisiert verarbeitet. Nichtsdestotrotz darf nicht außer Acht gelassen werden, dass die Analyse der MRT-Daten in einer Cloud der Firma Tensor Medical stattfindet. Somit werden die Daten nicht in einem lokalen Rechner verarbeitet, sondern werden in das Servernetzwerk der Firma Tensor Medical ausgelagert und dort für einen gewissen Zeitraum gespeichert. Die Daten können nach dem Upload über eine Internetverbindung mittels Webbrowser und einen entsprechenden Link aufgerufen werden. Tensor Medical arbeitet auf Basis einer strengen Datenschutzrichtlinie (<https://www.tensormedical.ai/privacy-policy>), welche auf der Datenschutz-Grundverordnung (EU) 2016/679 (DSGVO) fußt. Trotzdem besteht bei Speicherung der Daten in einer Cloud größere Gefahr für Sicherheitslecks, da die Daten, wie bereits erwähnt, nicht lokal verarbeitet werden.

Als Limitation dieser Studie ist die kleine Kohorte von 67 untersuchten Patient\*innen zu nennen. Der Anteil radiologisch aktiver Patient\*innen war dadurch relativ gering (Methode VA als Goldstandard mit 17/67 Patient\*innen = 25,4%), was gemeinsam mit der herangezogenen Prävalenz zur Berechnung der Vorhersagewerte, welche die gruppenspezifisch höhere Prävalenz wie zuvor erläutert nicht berücksichtigt, Sensitivität, Spezifität und Vorhersagewerte beeinflusst. Die Generalisierung der Ergebnisse dieser Studie hinsichtlich der Performance von Copilot.ms ist eingeschränkt, da die untersuchten MRT-Aufnahmen an nur einem einzigen MRT-Scanner mit demselben standardisierten

MRT-Protokoll akquiriert wurden. Folglich kann die Performance bei MRT-Daten, welche an einem anderen MRT-Scanner mit unterschiedlicher Magnetfeldstärke oder Impulssequenz erstellt wurden, variieren. Eine weitere Einschränkung dieser Studie ist die limitierte Aussagekraft des Vorzeichentests, wobei aufgrund der Konfiguration der Datenvariablen nur der Vorzeichentest angewandt werden konnte. Durch zusätzliche Berechnung der Gütekriterien der einzelnen Methoden konnte die Forschungsfrage schließlich beantwortet werden. Als Schwäche dieser Studie ist zudem zu nennen, dass kein Referenz-Goldstandard erstellt wurde, wodurch eine Berechnung der Gütekriterien des aktuellen Goldstandards (visuelle Analyse bzw. manuelle Segmentierung, Methode VA) nicht möglich war. Dies hätte einen Vergleich der Gütekriterien des Goldstandards mit den Gütekriterien der untersuchten Methoden AA und AAValid erlaubt.

Als Stärke dieser Studie ist die Durchführung als Teil einer multizentrischen Studie zu nennen. Dadurch ist insgesamt eine Analyse größerer Datenmengen möglich, wodurch weitgehendere Schlussfolgerungen und größere wissenschaftliche Aussagekraft erreicht werden. Eine weitere Stärke dieser Studie ist die Verarbeitung vergleichbarer Patient\*innen-Daten, da die MRT-Aufnahmen am gleichen MRT-Forschungsscanner, mit dem gleichen Protokoll erstellt wurden. Zuletzt zählen die Relevanz in der klinischen Praxis und Aktualität des in dieser Studie untersuchten Forschungsgebietes zu Stärken dieser Diplomarbeit.

Anhand der angeführten Limitationen und berechneten Gütekriterien lässt sich erkennen, dass eine uneingeschränkte Anwendung von Copilot.ms in der geprüften Version noch nicht möglich ist. Ein Ansatz zur Verbesserung der angeführten Limitationen wäre ein weitreichenderes „Training“ von Copilot.ms. Da es sich bei dem System um ein neuronales Netzwerk handelt, ist seine Erkennungsfähigkeit und Performance stark von den zum Training des Programmes eingespeisten MRT-Daten abhängig. Copilot.ms müsste somit auch mit einer großen Anzahl an MRT-Daten von Patient\*innen mit einer PML, Komorbiditäten wie beispielsweise Aneurysmen oder differenzialdiagnostisch wichtigen Erkrankungen, wie MOGAD oder NMOSD Läsionen, konfrontiert und trainiert werden, um diese identifizieren zu können. Zur ausgedehnteren Beleuchtung dieser Aspekte wären Folgestudien mit einem umfassender trainierten neuronalen Netzwerk nötig. Dadurch wäre es möglich die Anwendbarkeit dieses entscheidungsunterstützenden Systems, mit verbesserter Performance hinsichtlich der Spezifität und der Sicherheitsaspekte, weiter zu explorieren.

## 4.1 Conclusio

Zusammenfassend zeigen die Ergebnisse dieser Studie, dass automatische Segmentierungsmethoden eine höhere Sensitivität als die konventionelle, manuelle Segmentierung hinsichtlich der Identifikation neuer T2-Läsionen und somit krankheitsaktiver MS-Patient\*innen haben. Diese höhere Sensitivität entsteht allerdings auf Kosten der Spezifität, wobei automatische Methoden mehr falsch positive Ergebnisse liefern, wodurch ein Validierungsschritt durch Expert\*innen zur Steigerung der Spezifität und zuverlässigeren Läsionsdetektion nötig wird. Somit ist ein entscheidungsunterstützendes System, welches einen solchen Validierungsschritt beinhaltet, vielversprechender für den klinischen Einsatz als vollautomatische Segmentierungsverfahren, weil fehlerhafte Segmentierungen validiert und korrigiert werden können. Da ein entscheidungsunterstützendes System auf einem automatischen Segmentierungsverfahren basiert, ist es relevant, dass jenes automatische Verfahren zuverlässige Ergebnisse in der Vorauswertung von MRT-Daten liefert. Das in dieser Studie untersuchte neuronale Netzwerk zur Entscheidungsunterstützung liefert zuverlässige Ergebnisse beim Verlaufsmonitoring von stabilen Patient\*innen, muss allerdings für einen Einsatz im klinischen Alltag noch überarbeitet werden, um neue T2-Läsionen korrekt identifizieren zu können und schlussendlich Therapieentscheidungen davon abhängig zu machen. Aufgrund der mangelnden zeitlichen und personellen Ressourcen wäre der Einsatz eines entscheidungsunterstützenden Systems in der klinischen Praxis wertvoll und wäre besonders hinsichtlich der zeitlichen Effizienz im klaren Vorteil gegenüber der konventionellen manuellen Segmentierung.

Mit der entsprechenden Weiterentwicklung vor allem in Hinblick auf Spezifität und Verbesserung des positiven Vorhersagewertes, sowie Implementation der Erkennung von Aspekten des Sicherheitsmonitorings (Vorliegen einer PML oder zerebraler Komorbiditäten) und der Fähigkeit MS-Läsionen von zerebralen Läsionen anderer Ätiologie (mikroangiopathisch, NMOSD, MOGAD) zu differenzieren, könnten entscheidungsunterstützende Systeme in Zukunft einen bedeutenden Beitrag im klinischen Alltag liefern.

## Literaturverzeichnis

1. Oh J, Vidal-Jordana A, Montalban X. Multiple sclerosis: Clinical aspects. Vol. 31, *Current Opinion in Neurology*. Lippincott Williams and Wilkins; 2018. p. 752–9.
2. Wiendl H, Gold R, Berger T, Derfuss T, Linker R, Mäurer M, et al. Multiple Sclerosis Therapy Consensus Group (MSTCG): position statement on disease-modifying therapies for multiple sclerosis (white paper). Vol. 14, *Therapeutic Advances in Neurological Disorders*. SAGE Publications Ltd; 2021.
3. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Vol. 20, *The Lancet Neurology*. Elsevier Ltd; 2021. p. 653–70.
4. Doshi A, Chataway J. Multiple sclerosis, a treatable disease. Vol. 16, *Clinical Medicine*. 2016.
5. Belbasis L, Bellou V, Evangelou E, Ioannidis JPA, Tzoulaki I. Environmental risk factors and multiple sclerosis: An umbrella review of systematic reviews and meta-analyses. *Lancet Neurol*. 2015 Mar 1;14(3):263–73.
6. Al-Sakran LH, Marrie RA, Blackburn DF, Knox KB, Evans CD. Establishing the Incidence and Prevalence of Multiple Sclerosis in Saskatchewan. *Canadian Journal of Neurological Sciences*. 2018 May 1;45(3):295–303.
7. Dobson R, Giovannoni G. Multiple sclerosis – a review. Vol. 26, *European Journal of Neurology*. Blackwell Publishing Ltd; 2019. p. 27–40.
8. Salhofer-Polanyi S, Cetin H, Leutmezer F, Baumgartner A, Blechinger S, Dal-Bianco A, et al. Epidemiology of Multiple Sclerosis in Austria. *Neuroepidemiology*. 2017 Oct 1;49(1–2):40–4.
9. Aarli JA, Abramsky O, Browne P, Chandraratna D, Angood C, Tremlett H, et al. Section Editors Global Perspectives ATLAS OF MULTIPLE SCLEROSIS 2013: A GROWING GLOBAL PROBLEM WITH WIDESPREAD INEQUITY [Internet]. 2014. Available from: <http://www.msif.org/about-ms/publications->
10. Koch-Henriksen N, Magyari M. Apparent changes in the epidemiology and severity of multiple sclerosis. Vol. 17, *Nature Reviews Neurology*. Nature Research; 2021. p. 676–88.
11. Hawkes CH, Giovannoni G, Lechner-Scott J, Levy M, Waubant E. Multiple sclerosis and migration revisited. Vol. 34, *Multiple Sclerosis and Related Disorders*. Elsevier B.V.; 2019. p. A1–2.
12. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. MULTIPLE SCLEROSIS Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis [Internet]. Vol. 375, *Science*. 2022. Available from: <https://www.science.org>

13. Robinson WH, Steinman L. Epstein-Barr virus and multiple sclerosis. Vol. 375, Science. American Association for the Advancement of Science; 2022. p. 263–4.
14. Compston A, Coles A. Multiple sclerosis. Vol. 372, The Lancet. Elsevier B.V.; 2008. p. 1502–17.
15. Kim W, Patsopoulos NA. Genetics and functional genomics of multiple sclerosis. Vol. 44, Seminars in Immunopathology. Springer Science and Business Media Deutschland GmbH; 2022. p. 63–79.
16. Karussis D. The diagnosis of multiple sclerosis and the various related demyelinating syndromes: A critical review. Vols. 48–49, Journal of Autoimmunity. Academic Press; 2014. p. 134–42.
17. Lassmann H. Pathology and disease mechanisms in different stages of multiple sclerosis. Vol. 333, Journal of the Neurological Sciences. Elsevier B.V.; 2013. p. 1–4.
18. McGinley MP, Goldschmidt CH, Rae-Grant AD. Diagnosis and Treatment of Multiple Sclerosis: A Review. Vol. 325, JAMA - Journal of the American Medical Association. American Medical Association; 2021. p. 765–79.
19. Lubetzki C, Zalc B, Williams A, Stadelmann C, Stankoff B. Remyelination in multiple sclerosis: from basic science to clinical translation [Internet]. Vol. 19, Lancet Neurol. 2020. Available from: [www.thelancet.com/neurology](http://www.thelancet.com/neurology)
20. Olek MJ. Multiple Sclerosis. Vol. 174, Annals of Internal Medicine. American College of Physicians; 2021. p. ITC81–96.
21. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Soelberg Sørensen P, Alan Thompson DmsJ, et al. VIEWS & REVIEWS Defining the clinical course of multiple sclerosis The 2013 revisions. 2014.
22. Klineova S, Lublin FD. Clinical course of multiple sclerosis. Cold Spring Harb Perspect Med. 2018 Sep 1;8(9).
23. Okuda DT, Mowry EM, Beheshtian A, Waubant E, Baranzini SE, Goodin DS, et al. Incidental MRI anomalies suggestive of multiple sclerosis The radiologically isolated syndrome [Internet]. 2009. Available from: [www.neurology.org](http://www.neurology.org).
24. Okuda DT, Siva A, Kantarci O, Inglese M, Katz I, Tutuncu M, et al. Radiologically isolated syndrome: 5-year risk for an initial clinical event. PLoS One. 2014 Mar 5;9(3).
25. Yamout BI, Alroughani R. Multiple Sclerosis. Semin Neurol. 2018 Apr 1;38(2):212–25.
26. Fisniku LK, Brex PA, Altmann DR, Miszkief KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: A 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain. 2008 Mar;131(3):808–17.

27. Rovaris M, Confavreux C, Furlan R, Kappos L, Comi G, Filippi M. Secondary progressive multiple sclerosis: current knowledge and future challenges. *Lancet Neurol.* 2006 Apr 1;5(4):343–54.
28. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Vol. 17, *The Lancet Neurology.* Lancet Publishing Group; 2018. p. 162–73.
29. Uitdehaag BMJ, Barkhof F, Coyle PK, Gardner JD, Jeffery DR, Mikol DD. The changing face of multiple sclerosis clinical trial populations. Vol. 27, *Current Medical Research and Opinion.* 2011. p. 1529–37.
30. Travers BS, Tsang BKT, Barton JL. Multiple Sclerosis - Diagnosis, disease-modifying therapy and prognosis. Vol. 51, Reprinted from *AJGP.* 2022.
31. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Vol. 391, *The Lancet.* Lancet Publishing Group; 2018. p. 1622–36.
32. Banwell B, Bennett JL, Marignier R, Kim HJ, Brilot F, Flanagan EP, et al. Diagnosis of myelin oligodendrocyte glycoprotein antibody-associated disease: International MOGAD Panel proposed criteria. Vol. 22, *The Lancet Neurology.* Elsevier Ltd; 2023. p. 268–82.
33. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* 2014 Mar 25;14(1).
34. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology [Internet].* 1983 Nov 1;33(11):1444–1444. Available from: <https://www.neurology.org/lookup/doi/10.1212/WNL.33.11.1444>
35. Hemond CC, Bakshi R. Magnetic resonance imaging in multiple sclerosis. *Cold Spring Harb Perspect Med.* 2018 May 1;8(5).
36. Filippi M, Preziosa P, Rocca MA. Mri in multiple sclerosis: What is changing? Vol. 31, *Current Opinion in Neurology.* Lippincott Williams and Wilkins; 2018. p. 386–95.
37. Tintore M, Cobo-Calvo A, Carbonell P, Arrambide G, Otero-Romero S, Río J, et al. Effect of Changes in MS Diagnostic Criteria Over 25 Years on Time to Treatment and Prognosis in Patients With Clinically Isolated Syndrome. *Neurology.* 2021 Oct 26;97(17):E1641–52.
38. Sahraian MA, Radue EW, Haller S, Kappos L. Black holes in multiple sclerosis: Definition, evolution, and clinical correlations. Vol. 122, *Acta Neurologica Scandinavica.* 2010. p. 1–8.
39. Truyen L, Van Waesberghe ; J H T M, Van Walderveen ; M A A, Van Oosten ; B W, Polman ; C H, Hommes ; O R, et al. Accumulation of hypointense lesions (“black holes”) on T<sub>2</sub> spin-echo MRI correlates with disease progression in multiple sclerosis. Vol. 47, *NEUROLOGY.* 1996.

40. Lukas C, Sailer M. Magnetresonanztomografie bei Patienten mit Multipler Sklerose: Relevanz in der Diagnose und Verlaufsbeurteilung 1. Fortschritte der Neurologie Psychiatrie. 2015 Sep 2;83(9):522–34.
41. Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. Vol. 118, Brain. 1995.
42. Filippi M, Rovaris M, Campi A, Pereira C, Comi G. Semi-automated thresholding technique for measuring lesion volumes in multiple sclerosis: Effects of the change of the threshold on the computed lesion loads. Acta Neurol Scand. 1996;93(1):30–4.
43. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Vol. 17, Medical Image Analysis. 2013. p. 1–18.
44. Danelakis A, Theoharis T, Verganelakis DA. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Vol. 70, Computerized Medical Imaging and Graphics. Elsevier Ltd; 2018. p. 83–100.
45. Mortazavi D, Kouzani AZ, Soltanian-Zadeh H. Segmentation of multiple sclerosis lesions in MR images: A review. Vol. 54, Neuroradiology. 2012. p. 299–320.
46. Handels H. Medizinische Bildverarbeitung [Internet]. 2nd ed. Wiesbaden: Vieweg+Teubner; 2009. Available from: <http://link.springer.com/10.1007/978-3-8348-9571-4>
47. Goshtasby A, Staib L, Studholme C, Terzopoulos D. Nonrigid image registration: Guest editors' introduction. Vol. 89, Computer Vision and Image Understanding. Academic Press Inc.; 2003. p. 109–13.
48. Ganzetti M, Wenderoth N, Mantini D. Intensity inhomogeneity correction of structural MR images: A data-driven approach to define input algorithm parameters. Front Neuroinform. 2016 Mar 15;10(MAR).
49. Gerig G, Kubler O, Kikinis R, Jolesz FA. Nonlinear Anisotropic Filtering of MRI Data. Vol. 1, IEEE TRANSACTIONS ON MEDICAL IMAGING. 1992.
50. Wicks DAG, Tofts PS, Miller DH, Du Boulay GH, Feinstein A, Sacares RP, et al. Volume measurement of multiple sclerosis lesions with magnetic resonance images A preliminary study. Vol. 34, Neuroradiology. 1992.
51. Anbeek P, Vincken KL, Van Osch MJP, Bisschops RHC, Van Der Grond J. Probabilistic segmentation of white matter lesions in MR imaging. Neuroimage. 2004 Mar;21(3):1037–44.
52. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Peter Campbell J. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol. 2020;9(2).

53. Grossi E, Buscema M. Introduction to artificial neural networks. *Eur J Gastroenterol Hepatol* [Internet]. 2007 Dec;19(12):1046–54. Available from: <http://journals.lww.com/00042737-200712000-00005>
54. Kriegeskorte N, Golan T. Neural network models and deep learning. *Current Biology* [Internet]. 2019 Apr;29(7):R231–6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0960982219302040>
55. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. Vol. 13, *IEEE TRANSACTIONS ON MEDICAL IMAGING*. 1994.
56. Swanton J, Fernando K, Miller D. Early prognosis of multiple sclerosis. In: *Handbook of Clinical Neurology*. Elsevier B.V.; 2014. p. 371–91.
57. He A, Spelman T, Manouchehrinia A, Ciccarelli O, Hillert J, McKay K. Association between early treatment of multiple sclerosis and patient-reported outcomes: A nationwide observational cohort study. *J Neurol Neurosurg Psychiatry*. 2023 Apr 1;94(4):284–9.
58. Brown JWL, Phd S, Kalincik T, Spelman T, Hillert J, He A, et al. Timing of high-efficacy therapy for multiple sclerosis: a retrospective observational cohort study [Internet]. Vol. 19, *Articles Lancet Neurol*. 2020. Available from: [www.thelancet.com/neurology](http://www.thelancet.com/neurology)
59. Hunter SF. Overview and diagnosis of multiple sclerosis. *Am J Manag Care* [Internet]. 2016 Jun;22(6 Suppl):s141-50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27356023>
60. Pandit L. No evidence of disease activity (NEDA) in multiple sclerosis - Shifting the goal posts. Vol. 22, *Annals of Indian Academy of Neurology*. Wolters Kluwer Medknow Publications; 2019. p. 261–3.
61. Hauser SL, Cree BAC. Treatment of Multiple Sclerosis: A Review. Vol. 133, *American Journal of Medicine*. Elsevier Inc.; 2020. p. 1380-1390.e2.
62. Gross RH, Krieger S. Alemtuzumab in multiple sclerosis: an update. Vol. 5, *Neurodegenerative disease management*. 2015. p. 225–32.
63. Sormani MP, Rio J, Tintorè M, Signori A, Li D, Cornelisse P, et al. Scoring treatment response in patients with relapsing multiple sclerosis. *Multiple Sclerosis Journal*. 2013;19(5):605–12.
64. Tutuncu M, Altintas A, Dogan B V., Uygunoglu U, Kale Icen N, Elmalı Karakaya A, et al. The use of Modified Rio score for determining treatment failure in patients with multiple sclerosis: retrospective descriptive case series study. *Acta Neurol Belg*. 2021 Dec 1;121(6):1693–8.
65. Rovira A, Corral JF, Auger C, Valverde S, Vidal-Jordana A, Oliver A, et al. Assessment of automatic decision-support systems for detecting active T2 lesions in multiple sclerosis patients. *Multiple Sclerosis Journal*. 2022 Jul 1;28(8):1209–18.

66. Cabezas M, Corral JF, Oliver A, Díez Y, Tintoré M, Auger C, et al. Improved automatic detection of new t2 lesions in multiple sclerosis using deformation fields. *American Journal of Neuroradiology*. 2016 Oct 1;37(10):1816–23.
67. Thirion JP, Calmon G. Deformation Analysis to Detect and Quantify Active Lesions in Three-Dimensional Medical Image Sequences. Vol. 18, *IEEE TRANSACTIONS ON MEDICAL IMAGING*. 1999.
68. Lladó X, Ganiler O, Oliver A, Martí R, Freixenet J, Valls L, et al. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology*. 2012 Aug;54(8):787–807.
69. Rey D, Subsol G, Delingette H, Ayache N. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis [Internet]. Vol. 6, *Medical Image Analysis*. 2002. Available from: [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)
70. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. Second. New York: Wiley; 2014.
71. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. *BMJ* [Internet]. 1994 Jul 9;309(6947):102–102. Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj.309.6947.102>
72. Pongratz V, Schmidt P, Bussas M, Grahl S, Gaser C, Berthele A, et al. Prognostic value of white matter lesion shrinking in early multiple sclerosis: An intuitive or naïve notion? *Brain Behav*. 2019 Dec 1;9(12).