

Aus dem Universitätsklinikum Münster
Institut für Translationale Psychiatrie
– Direktor: Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski –

**Erstellung von Machine Learning-Modellen zur Prädiktion einzelner
Krankheitscharakteristika der Depression in Abhängigkeit
von der kognitiven Leistung und der Gehirnstruktur**

INAUGURAL – DISSERTATION

zur

Erlangung des doctor medicinae

der Medizinischen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von Schmitt, Michael Andreas Dominik
aus Wiesbaden

2021

Gedruckt mit Genehmigung der Medizinischen Fakultät der Westfälischen Wilhelms-
Universität Münster

Dekan: Univ.-Prof. Dr. med. Frank Ulrich Müller

1. Berichterstatter: Univ.-Prof. Dr. rer. nat. Dipl.-Psych. Tim Hahn

2. Berichterstatter: Univ.-Prof. Dr. med. André Karch, MSc

Tag der mündlichen Prüfung: 10.12.2021

Aus dem Universitätsklinikum Münster
Institut für Translationale Psychiatrie
- Direktor: Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski -
1. Berichterstatter: Univ.-Prof. Dr. rer. nat. Dipl.-Psych. Tim Hahn
2. Berichterstatter: Univ.-Prof. Dr. med. André Karch, MSc

ZUSAMMENFASSUNG

Erstellung von Machine Learning-Modellen zur Prädiktion einzelner
Krankheitscharakteristika der Depression in Abhängigkeit
von der kognitiven Leistung und der Gehirnstruktur
Schmitt, Michael Andreas Dominik

Hintergrund: Die Major Depressive Disorder (MDD) ist eine der häufigsten psychischen Erkrankungen weltweit. In Bezug auf die Therapieentscheidung ist das Verständnis des Krankheitsverlaufes besonders wichtig. Maschinelles Lernen bietet die Chance, solche Vorhersagen zukünftig zu individualisieren. Ziel dieser Studie ist es, durch prädiktive Analysen den kontrovers diskutierten Zusammenhang zwischen kognitiven Defiziten bzw. strukturellen Gehirnveränderungen und einzelnen Krankheitscharakteristika zu erforschen.

Methoden: Die Proband(inn)en wurden mittels neuropsychologischer Testung und T1-gewichteter MRT-Untersuchung phänotypisiert. Für beide Modalitäten wurde jeweils die Prädizierbarkeit der Diagnose (MDD/Kontrolle), des Schweregrads (Hamilton Depressions-Skala, Beck-Depressions-Inventar) und des Verlaufs (Hospitalisierung, Rezidiv, Funktionsniveau) analysiert. Das Funktionsniveau wurde anhand von Scores (Global Assessment of Functioning, Short Form 36) quantifiziert. Die Stichprobe umfasste $n = 746$ bzw. 432 MDD-Patient(inn)en und $n = 914$ bzw. 615 gesunde Kontrollen (MRT-Daten bzw. neuropsych. Daten). Die Ethikkommissionen der Universitäten Marburg und Münster haben die Studie am 07.02.2014 und 01.08.2014 bewilligt (Aktenzeichen: Studie 07/14 bzw. 2014-422-b-S).

Ergebnisse: Die Prädiktion der Diagnose gelang mit einer gewichteten Genauigkeit von $59.62 (\pm 4.06)$ bzw. $55.34 (\pm 3.57) \%$ (MRT-Daten bzw. neuropsych. Daten). Darüber hinaus konnten lediglich die Short Form 36-Werte (Subskala: körperliche Funktionsfähigkeit) prädiziert werden. Der mittlere absolute Fehler konnte dabei um ca. 9 bzw. 13 % (MRT-Daten bzw. neuropsych. Daten) gegenüber der Vergleichsmetrik (Mittelwert der Zieldaten) reduziert werden, wobei die Pearson-Korrelation $r = 0.24 (\pm 0.14)$ bzw. $0.35 (\pm 0.10)$ betrug.

Diskussion: Dass neben der Diagnose lediglich die Werte der Short Form 36 prädizierbar waren, könnte auf eine motivationale Komponente, die gleichermaßen durch die Prädiktoren und die Short Form 36 erfasst wird, in der Genese der MDD hinweisen. Die Klärung dieser These erfordert weitere Forschungsprojekte. Die klinische Relevanz der ermittelten Modellgüten und die Generalisierbarkeit auf andere Stichproben ist kritisch zu hinterfragen.

Tag der mündlichen Prüfung: 10.12.2021

ERKLÄRUNG

Ich gebe hiermit die Erklärung ab, dass ich die Dissertation mit dem Titel:

Erstellung von Machine Learning-Modellen zur Prädiktion einzelner Krankheitscharakteristika der Depression in Abhängigkeit von der kognitiven Leistung und der Gehirnstruktur

im:

Institut für Translationale Psychiatrie an der Klinik für Psychische Gesundheit der Universitätsklinik Münster

unter der Anleitung von:

Univ.-Prof. Dr. rer. nat. Dipl.-Psych. Tim Hahn

1. selbständig angefertigt,
2. nur unter Benutzung der im Literaturverzeichnis angegebenen Arbeiten angefertigt und sonst kein anderes gedrucktes oder ungedrucktes Material verwendet,
3. keine unerlaubte fremde Hilfe in Anspruch genommen,
4. sie weder in der gegenwärtigen noch in einer anderen Fassung einer in- oder ausländischen Fakultät als Dissertation, Semesterarbeit, Prüfungsarbeit, oder zur Erlangung eines akademischen Grades, vorgelegt habe.

Münster, den 14.12.2021

Michael Schmitt

Inhaltsverzeichnis

1	Theoretischer Hintergrund	1
1.1	Die Major Depressive Disorder (MDD)	1
1.1.1	Definition und Diagnosekriterien der MDD	1
1.1.2	Epidemiologie der MDD	2
1.1.3	Ätiologie der MDD	3
1.1.4	Therapie und Verlauf der MDD	4
1.2	(Neuro-)biologische Korrelate der MDD	4
1.2.1	Kognitive Beeinträchtigungen bei der MDD	5
1.2.2	Strukturelle Gehirnveränderungen bei der MDD	6
1.3	Maschinelles Lernen als Methode	7
1.3.1	Begriffserklärungen	7
1.3.2	Maschinelles Lernen in der psychiatrischen Forschung	11
1.3.3	Kritische Analyse von maschinellem Lernen in der psychiatrischen Forschung	13
1.3.4	Empfehlungen (Best Practices) für Projekte des maschinellen Lernens	14
1.4	Zusammenfassung und Hypothesen	16
2	Daten und Methoden	19
2.1	Die FOR 2107-Studie	19
2.1.1	Ein-/ Ausschlusskriterien	20
2.1.2	Rekrutierung und Studienablauf	21
2.1.3	Erhebung der psychometrischen Daten und Diagnosestellung	21
2.1.4	Erhebung der kognitiven Leistung (neuropsychologische Testbatterie)	24
2.1.5	Erhebung der strukturellen MRT-Daten	29
2.2	Finale Stichprobenwahl und Stichprobenstatistik	31
2.2.1	Stichprobenübersicht (Modalität: neuropsychologische Testung)	35
2.2.2	Stichprobenübersicht (Modalität: strukturelle MRT-Daten)	38
2.3	Implementierung der Methode des maschinellen Lernens	40
2.3.1	Erläuterung der PHOTONAI API und der Pipeline-Architektur	40
2.3.2	Erläuterung der eingesetzten Kreuzvalidierung	41
2.3.3	Transformatoren, Lernverfahren, Hyperparameter und Metriken	46
2.4	Beschreibung der durchgeführten Analysen	53
3	Ergebnisse	56

3.1	Beschreibung der Stichproben	56
3.2	Analysen zur Modalität: neuropsychologische Testung	58
3.2.1	Klassifikationsanalyse: MDD vs. HC	58
3.2.2	Klassifikationsanalyse: rezidivierender Verlauf vs. einzelne Episode	60
3.2.3	Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert	61
3.2.4	Regressionsanalyse: HAMD-Wert	61
3.2.5	Regressionsanalyse: BDI-Wert	62
3.2.6	Regressionsanalyse: GAF-Wert	62
3.2.7	Regressionsanalyse: SF-36-Wert (Subskala: körperliche Funktionsfähigkeit)	63
3.3	Analysen zur Modalität: strukturelle MRT-Daten	65
3.3.1	Klassifikationsanalyse: MDD vs. HC	65
3.3.2	Klassifikationsanalyse: rezidivierender Verlauf vs. einzelne Episode	67
3.3.3	Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert	68
3.3.4	Regressionsanalyse: HAMD-Wert	68
3.3.5	Regressionsanalyse: BDI-Wert	69
3.3.6	Regressionsanalyse: GAF-Wert	69
3.3.7	Regressionsanalyse: SF-36-Wert (Subskala: körperliche Funktionsfähigkeit)	70
4	Diskussion	73
4.1	Zusammenfassung der Ergebnisse	73
4.2	Diskussion der Klassifikationsanalysen: MDD vs. HC	74
4.3	Diskussion der Prädizierbarkeit einzelner Krankheitscharakteristika	79
4.4	Diskussion der Methodik des maschinellen Lernens	85
4.5	Limitierungen und Stärken dieser Studie	86
4.6	Fazit und Ausblick	89
5	Danksagung	92
6	Lebenslauf	93
7	Literaturverzeichnis	95
8	Anhang	I
8.1	Ausschlusskriterien (somatische Erkrankungen)	I
8.2	Variablenübersicht (Modalität: neuropsychologische Testung)	II
8.3	Variablenübersicht (Modalität: strukturelle MRT-Daten)	IV

8.4	Probandenaufruf am Studienstandort Münster	VII
8.5	Probandenaufruf am Studienstandort Marburg	IX
8.6	Screening-Fragebogen	XI
8.7	Studienaufklärung und Einverständniserklärung (Allgemein)	XIV
8.8	Studienaufklärung und Einverständniserklärung (Bioproben)	XXVII
8.9	Studienaufklärung und Einverständniserklärung (MRT)	XXXIX
8.10	Skripte der Machine Learning-Pipelines	XLIII
8.11	Ergebnistabellen zur Modalität: neuropsychologische Testung	XLV
8.12	Ergebnistabellen zur Modalität: strukturelle MRT-Daten	LI
8.13	Votum der Ethik-Kommission am Studienstandort Marburg	LVII
8.14	Votum der Ethik-Kommission am Studienstandort Münster	LIX

Abkürzungsverzeichnis

Abkürzung	Bedeutung
BAC	korrigierte Treffergenauigkeit (engl.: balanced accuracy)
BDI	Beck-Depressions-Inventar
BZT	Buchstaben-Zahlen-Test
d2	Test d2
DSM	Diagnostic and Statistical Manual
ENIGMA	Enhancing Neuroimaging Genetics Through Meta-Analysis
FOR 2107	Forschergruppe 2107
GAF	Global Assessment of Functioning
HAMD	Hamilton Depressions-Skala
HC	gesunde Kontrollen (engl.: healthy controls)
IQ	Intelligenzquotient
LOOCV	Leave-One-Out Kreuzvalidierung
M	Mittelwert (engl.: mean)
MAE	mittlerer absoluter Fehler (engl.: mean absolute error)
MDD	schwere depressive Störung (engl.: major depressive disorder)
MPRAGE	Magnetization Prepared Rapid Acquisition Gradient Echo
MRT	Magnetresonanztomographie
MWT-B	Mehrfachwahl-Wortschatz-Intelligenztest
MWW	Mann-Whitney-Wilcoxon Test
RBF	radiale Basisfunktion (engl.: radial basis function)
RWT	Regensburger Wortflüssigkeits-Test
SD	Standardabweichung (engl.: standard deviation)
SF-36	Short Form 36-Fragebogen
SKID-I	Strukturiertes Klinisches Interview Achse-I
SVM	Support Vector Machine
TMT-A/-B	Trail-Making-Test A/B
VLMT	Verbaler Lern- und Merkfähigkeitstest
ZST	Zahlen-Symbol-Test

Abbildungsverzeichnis

Abbildung 1-1: Entwicklungsprozess eines prädiktiven Modells (in Anlehnung an (17,25))	16
Abbildung 2-1: Prozess der Stichprobengenerierung	33
Abbildung 2-2: simple Pipeline-Architektur in PHOTONAI (angelehnt an (64))	41
Abbildung 2-3: Prinzip der k-fachen äußeren Kreuzvalidierung (in Anlehnung an (17,63))	43
Abbildung 2-4: Prinzip der k-fachen inneren Kreuzvalidierung (in Anlehnung an (17,63))	44
Abbildung 2-5: Pipelinearchitektur unter Einsatz eines „Switch“ (angelehnt an (64))	48
Abbildung 2-6: Übersicht über die durchgeführten Analysen	54
Abbildung 3-1: Permutationstest, Klassifikationsanalyse: MDD vs. HC, Modalität: neuropsych. Testung	60
Abbildung 3-2: Permutationstest, Regressionsanalyse: SF-36-Wert (körperliche Funktionsfähigkeit), Modalität: neuropsychologische Testung	65
Abbildung 3-3: Permutationstest, Klassifikationsanalyse: MDD vs. HC, Modalität: strukturelle MRT-Daten	67
Abbildung 3-4: Permutationstest, Regressionsanalyse: SF-36-Wert (körperliche Funktionsfähigkeit), Modalität: strukturelle MRT-Daten	72
Abbildung 4-1: Verteilung der HAMD- und BDI-Werte für beide Modalitäten	77
Abbildung 4-2: Verteilung der GAF- und SF-36-Werte für beide Modalitäten	83
Abbildung 8-1: Probandenaufruf am Studienstandort Münster (Teil 1)	VII
Abbildung 8-2: Probandenaufruf am Studienstandort Münster (Teil 2)	VIII
Abbildung 8-3: Probandenaufruf am Studienstandort Marburg (Teil 1)	IX
Abbildung 8-4: Probandenaufruf am Studienstandort Marburg (Teil 2)	X

Formelverzeichnis

Formel 2-1: Skalierung der Prädiktoren (angelehnt an (125,133))	47
Formel 2-2: Berechnung des gamma-Koeffizienten (= scale) (angelehnt an (125,137,138))	50

Tabellenverzeichnis

Tabelle 2-1: Übersicht über die eingesetzten Fragebögen	22
Tabelle 2-2: Übersicht über die neuropsychologische Testbatterie	25
Tabelle 2-3: Stichprobentabelle 1, Modalität: neuropsychologische Testung, MDD- und HC-Gruppe	36
Tabelle 2-4: Stichprobentabelle 2, Modalität: neuropsychologische Testung, MDD-Gruppe	37
Tabelle 2-5: Stichprobentabelle 3, Modalität: strukturelle MRT-Daten, MDD- und HC-Gruppe	38
Tabelle 2-6: Stichprobentabelle 4, Modalität: strukturelle MRT-Daten, MDD-Gruppe	39
Tabelle 3-1: Klassifikationsanalyse: MDD vs. HC, Modalität: neuropsychologische Testung	59
Tabelle 3-2: Regressionsanalyse: SF-36-Wert, Modalität: neuropsychologische Testung	64
Tabelle 3-3: Klassifikationsanalyse: MDD vs. HC, Modalität: strukturelle MRT-Daten	66
Tabelle 3-4: Regressionsanalyse: SF-36-Wert, Modalität: strukturelle MRT-Daten	71
Tabelle 4-1: Klassifikationsanalyse: MDD vs. HC, beide Modalitäten, verschiedene MDD Subgruppen	76
Tabelle 8-1: Übersicht über die Ausschlusskriterien (somatische Erkrankungen)	I
Tabelle 8-2: Variablenübersicht für die Modalität: neuropsychologische Testung (Teil 1)	II
Tabelle 8-3: Variablenübersicht für die Modalität: neuropsychologische Testung (Teil 2)	III
Tabelle 8-4: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 1)	IV
Tabelle 8-5: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 2)	V
Tabelle 8-6: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 3)	VI
Tabelle 8-7: Klassifikationsanalyse: rezidivierend vs. einzelne Episode, Modalität: neuropsych. Testung	XLVI
Tabelle 8-8: Klassifikationsanalyse: hospitalisiert vs. nicht hosp., Modalität: neuropsych. Testung	XLVII
Tabelle 8-9: Regressionsanalyse: HAMD-Wert, Modalität: neuropsychologische Testung	XLVIII
Tabelle 8-10: Regressionsanalyse: BDI-Wert, Modalität: neuropsychologische Testung	XLIX
Tabelle 8-11: Regressionsanalyse: GAF-Wert, Modalität: neuropsychologische Testung	L
Tabelle 8-12: Klassifikationsanalyse: rezidivierend vs. einzelne Episode, Modalität: strukt. MRT-Daten	LII

Tabelle 8-13: Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert, Modalität: strukt. MRT-Daten	LIII
Tabelle 8-14: Regressionsanalyse: HAMD-Wert, Modalität: strukturelle MRT-Daten	LIV
Tabelle 8-15: Regressionsanalyse: BDI-Wert, Modalität: strukturelle MRT-Daten	LV
Tabelle 8-16: Regressionsanalyse: GAF-Wert, Modalität: strukturelle MRT-Daten	LVI

1 Theoretischer Hintergrund

Bei der schweren depressiven Störung (engl.: major depressive disorder, MDD) handelt es sich um eine der häufigsten psychischen Erkrankungen weltweit (1–6). Die MDD führt zu erheblichen Einschränkungen der Lebensqualität (1,7,8), bedingt eine erhöhte Mortalitätsrate (9,10) und ist durch einen episodenhaften, teils chronischen Verlauf geprägt (8,11–13). Für Betroffene sowie die Gesellschaft gehen mit der Krankheit immense ökonomische sowie emotionale Belastungen einher (3,14–16).

Maschinelles Lernen bietet vor allem im Hinblick auf die rasant wachsende Datenverfügbarkeit die Chance, die Medizin zukünftig zu personalisieren und somit auch die Diagnostik, Therapie und Prognostik im Bereich der MDD zu optimieren und die mit der Erkrankung einhergehenden Belastungen zu reduzieren (17–19).

In dieser Arbeit soll die Technologie des maschinellen Lernens genutzt werden, um basierend auf der kognitiven Leistung und der Gehirnstruktur einzelne Krankheitscharakteristika der MDD zu präzisieren. Durch eine angepasste Behandlungsstrategie könnten somit mildere Krankheitsverläufe ermöglicht werden (8).

Zur Einführung in die Thematik wird im Folgenden zunächst die depressive Störung (s. Kapitel 1.1) inklusive der hier relevanten (neuro-)biologischen Korrelate (s. Kapitel 1.2) vorgestellt. Darauf folgend wird die Methode des maschinellen Lernens zunächst allgemein und dann im Bezug auf die psychiatrische Forschung erläutert (s. Kapitel 1.3). Im Verlauf der beschriebenen Abschnitte wird die explizite Fragestellung dieser Studie erarbeitet und abschließend ausformuliert (s. Kapitel 1.4).

1.1 Die Major Depressive Disorder (MDD)

1.1.1 Definition und Diagnosekriterien der MDD

Die genaue Definition der MDD unterscheidet sich je nach klinischem Klassifikationssystem (20,21). Grundsätzlich handelt es sich um eine affektive Erkrankung, d. h. die Stimmung und die emotionale Verarbeitung von Betroffenen ist beeinträchtigt (22). Im Folgenden werden die Diagnosekriterien anhand des Diagnostic and Statistical Manual (DSM), das als Grundlage für

die Diagnosestellung der Probanden¹ in dieser Studie verwendet wurde, erläutert. In der aktuellen Fassung (DSM-V) sind die Hauptkriterien für die Diagnose einer MDD wie folgt definiert (22):

1. gedrückte Stimmung über die meiste Zeit des Tages,
2. Verlust von Interesse bzw. Freude an annähernd allen Aktivitäten über die meiste Zeit des Tages.

Weitere Nebenkriterien, die im Rahmen einer MDD auftreten können, sind:

3. signifikanter Gewichtsverlust oder -zuwachs bzw. veränderter Appetit,
4. Schlafstörungen (Hypersomnie bzw. Insomnie),
5. motorische Verlangsamung bzw. Anspannung,
6. Erschöpfung bzw. Energielosigkeit,
7. vermindertes Selbstwertgefühl bzw. Schuldgefühle,
8. verminderte Konzentrationsfähigkeit bzw. Entscheidungsschwierigkeiten,
9. Gedanken an den Tod bzw. Suizidpläne oder -handlungen.

Zur Diagnose einer MDD müssen insgesamt fünf Symptome (davon mind. ein Hauptsymptom) für mindestens zwei Wochen am Stück an den meisten Tagen bestehen und klinisch signifikantes Leiden bedingen bzw. funktionale Einschränkungen (z. B. im sozialen oder beruflichen Bereich) verursachen. Die Episode darf dabei nicht auf einen Substanzmissbrauch oder andere somatische Erkrankungen zurückführbar sein und zudem nicht besser durch eine andere psychische Störung erklärbar sein. In der Vorgeschichte darf zudem keine manische oder hypomanische Episode aufgetreten sein (22).

1.1.2 Epidemiologie der MDD

Nach Angabe der Global Burden of Disease Studie litten 2017 weltweit (Prävalenz) über 163 Millionen Menschen an einer MDD, über 106 Millionen an einer dysthymen Störung und über 45 Millionen an einer bipolaren Störung (engl.: bipolar disorder) (1). Die Angaben zur Lebenszeitprävalenz in der Literatur sind nicht eindeutig. Laut einer Studie, die in den Vereinigten Staaten von Amerika durchgeführt wurde, beträgt die Lebenszeitprävalenz

¹ Aufgrund der besseren Lesbarkeit wird in dieser Arbeit das generische Maskulinum verwendet. Entsprechende Begriffe sind geschlechtsunabhängig zu verstehen und beinhalten sowohl Frauen als auch Männer.

16.2 % (n = 9090) (3). In einer anderen Studie, die ebenfalls die US-Bevölkerung untersuchte, wurde eine entsprechende Lebenszeitprävalenz von 2.6 % (n = 36 309) ermittelt (4). Die Auswertung einer niederländischen Studie ergab wiederum eine Lebenszeitprävalenz von 18.7 % (n = 6646) (5). In einer breiter angelegten Vergleichsstudie, in der 8 Länder mit niedrigem und mittlerem Einkommen sowie 10 Länder mit hohem Einkommen untersucht wurden, wurde eine Lebenszeitprävalenz von 11.1 % respektive 14.6 % errechnet (6).

An den Lebensjahren gemessen, die eine betroffene Person mit einer Erkrankung und deren Beeinträchtigungen lebt (engl. Bezeichnung: Years Lived with Disability), ist die MDD mit einem Wert von > 30 weltweit unter den drei relevantesten Erkrankungen (1).

Die Rezidivrate der MDD beträgt entsprechend einer Metaanalyse für die durchschnittliche Bevölkerung bis zu 35 % nach 15 Jahren. Dieser Wert, der in etwa repräsentativ für die allgemeine Bevölkerung ist, ist vergleichsweise niedrig. Untersucht man hingegen Patienten spezialisierter psychiatrischer Kliniken, weisen diese entsprechend der Studie eine deutlich höhere Rückfallquote von 60 % nach 5 Jahren, 67 % nach 10 Jahren und 85 % nach 15 Jahren auf (11).

1.1.3 Ätiologie der MDD

Die Ätiologie der MDD ist nach dem heutigen Stand der Forschung nicht abschließend geklärt. Depressive Störungen werden als ein heterogenes Krankheitsbild mit multifaktorieller Genese verstanden. Dabei spielen insbesondere genetische Prädispositionen und Umweltfaktoren eine entscheidende Rolle (11,17,23–25). Pathophysiologisch geht man u. a. davon aus, dass es zu einer Dysregulation von neurobiologischen Transmittern kommt, wobei insbesondere das monoaminerge System betroffen zu sein scheint (v. a. die Wirkstoffe Serotonin, Noradrenalin und Dopamin) (26–28). Diese These stützt sich vor allem darauf, dass wirksame Antidepressiva mitunter genau diese Wirkstoffsysteme beeinflussen (26–28).

Andere Mechanismen wie bspw. eine Beteiligung des Immunsystems oder die Rolle von psycho-sozialen Faktoren sind in der Diskussion, zumal nicht alle Betroffenen gleichermaßen von den verfügbaren Antidepressiva profitieren (28–30).

Die Mechanismen, die dem episodenhaften Verlauf der Krankheit zu Grunde liegen, sind bis heute wenig verstanden (11,31–33). Bisherige Studien haben klinische Faktoren (z. B. die Dauer und Schwere der bisherigen Episoden, die Schwere der Restsymptomatik am Ende

einer Episode, die Häufung von depressiven Episoden oder die Beeinträchtigung der kognitiven Leistungsfähigkeit) als Risikofaktoren für die Entwicklung eines Rezidivs identifiziert (11,31–33).

1.1.4 Therapie und Verlauf der MDD

Zur Behandlung der MDD existieren eine Vielzahl an medikamentösen (34) und nicht-medikamentösen Behandlungsmethoden (35). Zu letzteren zählen bspw. verschiedene Formen der Psycho- und Verhaltenstherapien, die Elektrokonvulsionstherapie, Sporttherapien und neuere digital gestützte Ansätze (35,36). Die Forschung belegt, dass derzeitige Therapieressourcen nicht immer effektiv eingesetzt werden – zumal die individuelle Wirkung der unterschiedlichen Therapien bei einzelnen Patienten unzureichend erforscht ist (35) – und dass eine Optimierung der Therapie die Gesundheit von Betroffenen über einen langen Zeitraum deutlich verbessern könnte (8).

In Bezug auf den episodenhaften Verlauf der MDD implizieren die Ergebnisse einer Meta-Analyse, dass präventive psychologische Therapien (z. B. kognitive Verhaltenstherapien) das Risiko eines Rezidivs – im Gegensatz zu der üblicherweise zeitlich begrenzten Behandlung in Akutphasen oder zur singulären Langzeittherapie mit Antidepressiva – reduzieren könnten (13). Die Kombination aus einem präzisen System zur Vorhersage des individuellen Krankheitsverlaufes und dem Einsatz entsprechender Therapien könnte helfen, die Krankheitslast von Betroffenen zu mindern (11). Im Rahmen einer Studie der Weltgesundheitsorganisation (16) wurde allerdings festgestellt, dass gerade MDD-Patienten mit schweren Verläufen vielerorts über einen langen Zeitraum unbehandelt bleiben. Eine Neuverteilung der verfügbaren Behandlungsressourcen unter Beachtung besonders schwerer Fälle und solcher, die zur Zeit mild sind, im Verlauf jedoch ein hohes Intensivierungsrisiko aufweisen, könnte das bestehende Problem der Unterbehandlung lösen (16).

1.2 (Neuro-)biologische Korrelate der MDD

Wie in Kapitel 1.1.3 beschrieben, handelt es sich bei der MDD um ein heterogenes Krankheitsbild mit multifaktorieller, nicht abschließend geklärter Genese. Um das Verständnis über die Krankheitsentstehung, verschiedene Verlaufsformen und potentielle Behandlungsoptionen zu verbessern, beschäftigt sich die gegenwärtige psychiatrische Forschung mit unterschiedlichsten Modalitäten wie z. B. der Neuroanatomie (bildgebende Verfahren), Umweltfaktoren (Persönlichkeits-, Risiko- und Schutzfaktoren, Kindheitstraumata usw.),

polygenen Risikofaktoren, neurochemischen Faktoren (Transmittern), der Elektroenzephalografie und weiteren (31,37,38). Die in dieser Arbeit erstellten prädiktiven Modelle treffen ihre Vorhersagen basierend auf der kognitiven Leistung bzw. der Gehirnstruktur, weshalb diese beiden Modalitäten im Folgenden genauer erläutert werden. Die Begründung der Auswahl erfolgt im weiteren Verlauf der vorliegenden Arbeit (s. Kapitel 1.4).

1.2.1 Kognitive Beeinträchtigungen bei der MDD

Viele Studien im Bereich der neuropsychologischen psychiatrischen Forschung haben gezeigt, dass die MDD mit Einschränkungen in unterschiedlichen kognitiven Funktionsbereichen einhergeht (39–45). In einigen Studien wurde gezeigt, dass kognitive Einschränkungen schon in der ersten depressiven Episode auftreten (44) und zudem über die Dauer einer symptomgeprägten MDD-Episode hinaus bestehen bleiben können (39,41).

Andere Studien konnten hingegen eine deutliche Verminderung der kognitiven Einschränkungen nach erfolgreicher Therapie einer depressiven Episode zeigen (46,47). Weitere Studien implizieren, dass z. B. ein hohes Alter (48), psychiatrische Komorbiditäten (49) oder der Depressions-Subtyp (50) beeinflussen können, welche kognitiven Bereiche bei dem Vorliegen einer MDD besonders von eingeschränkter Funktionsfähigkeit betroffen sind. In der Literatur wurden depressionsspezifische Einschränkungen in bestimmten Funktionsbereichen ermittelt, sodass diskutiert wird, dass es sich nicht um einen von anderen Faktoren abhängigen Nebeneffekt handelt, der bspw. auf mangelnder Aufmerksamkeit von depressiven Patienten basiert, sondern vielmehr um ein Merkmal der MDD (41).

In einigen Studien trat keine Korrelation zwischen der Depressionsschwere und der Performance in der neuropsychologischen Testung auf, während andere Arbeiten einen solchen Zusammenhang vor allem für schwierige Aufgaben feststellen konnten (41). In einer weiteren Meta-Analyse wurde eine Korrelation zwischen der Depressionsschwere und bestimmten Funktionsbereichen (episodisches Gedächtnis, exekutive Funktionen und Verarbeitungsgeschwindigkeit) festgestellt, wohingegen andere Bereiche (semantisches oder visuell-räumliches Gedächtnis) nicht betroffen waren (40). In einer dritten Meta-Analyse wurde eine Assoziation zwischen der psychomotorischen Geschwindigkeit sowie der Merkfähigkeit und der klinischen Ausprägung der MDD ermittelt, wohingegen dies für die Aufmerksamkeitsspanne und die exekutiven Funktionen nicht der Fall war (44).

Die Mehrheit der zitierten Studien belegt, dass die MDD mit kognitiven Einschränkungen einhergeht. Die Frage, ob Korrelationen zwischen den kognitiven Einschränkungen und einzelnen Krankheitscharakteristika der Depression bestehen, scheint nicht abschließend geklärt zu sein, da es diesbezüglich in der Literatur widersprüchliche Angaben gibt.

1.2.2 Strukturelle Gehirnveränderungen bei der MDD

Aus zwei umfangreichen Meta-Analysen des ENIGMA Konsortiums (Enhancing Neuroimaging Genetics Through Meta-Analysis) ergibt sich, dass bei der Erforschung von strukturellen Gehirnveränderungen bei depressiven Probanden mittels Magnetresonanztomographie (MRT) unterschiedliche Effekte festgestellt wurden (51,52). Das genaue Muster der strukturellen Veränderungen ist bis heute nicht abschließend geklärt und inkonsistent beschrieben (51,52). Problematisch erscheint, dass viele MRT-Studien nur kleine Stichproben umfassen, unterschiedliche Bildverarbeitungsprozesse anwenden und depressive Erkrankungen zudem heterogen sind (51–53). Grundsätzlich können bei MDD-Patienten subkortikale und kortikale Veränderungen beobachtet werden (51–53).

Subkortikale Veränderungen:

Eine Meta-Analyse (51) ergab, dass Probanden mit einer MDD verglichen mit gesunden Kontrollen in den 15 untersuchten Studien ein signifikant geringeres Hippocampus-Volumen aufwiesen. Dieser Effekt zeigte sich allerdings maßgeblich bei Probanden mit rezidivierender Depression, wohingegen Probanden mit nur einer Episode keinen Unterschied zu gesunden Kontrollen aufwiesen (51). Es wurde keine Korrelation zwischen der Depressionsschwere, dem mittleren Alter oder dem Anteil an Probanden, die Antidepressiva einnahmen, und den Gehirnveränderungen festgestellt (51). In zwei weiteren Meta-Analysen wurden bei MDD-Patienten vor allem in Bereichen der emotionalen Verarbeitung (u. a. Hippocampus, Striatum) geringere subkortikale Gehirnvolumina als bei gesunden Kontrollen festgestellt (54,55).

Kortikale Veränderungen:

In einer Meta-Analyse (52), auf der der folgende Abschnitt basiert, wurde gezeigt, dass die kortikale graue Substanz bei erwachsenen MDD-Patienten (> 21 Jahre) im orbitofrontalen Kortex, im anterioren und posterioren Gyrus Cinguli, in der Inselrinde sowie in den Temporallappen in den betrachteten Studien dünner ausgeprägt war als bei gesunden Kontrollen. Am stärksten war der Effekt bei Patienten in ihrer ersten Episode und bei Patienten, die erstmals im Erwachsenenalter an Depressionen erkrankten. Probanden in ihrer Adoleszenz (\leq 21 Jahre) wiesen eine geringere Gesamtoberflächengröße, regionale

Reduktionen in frontalen Regionen sowie in den visuellen, somatosensorischen und motorischen Regionen des Gehirns auf. Besonders ausgeprägt waren die Effekte hier bei Probanden mit rezidivierender Erkrankung. Korrelationen zwischen der Depressionsschwere und kortikalen Veränderungen konnten nicht festgestellt werden. MDD-bedingte kortikale Veränderungen könnten entsprechend dieser Studie sehr dynamisch und altersabhängig sein (52). Auch im Bereich des Kortex wiesen depressive Patienten vor allem in Bereichen, die in die emotionale Verarbeitung involviert sind (u. a. frontaler Kortex, Gyrus cinguli, orbitofrontaler Kortex), verringerte Volumina auf, wie zwei weitere Meta-Analysen belegen (54,55).

Wie auch in den obigen Abschnitten beschrieben, gibt es dabei die Vermutung, dass vor allem persistente Formen der MDD (z. B. rezidivierende Verläufe, länger andauernde Episoden) mit stärkeren Hirnveränderungen assoziiert sind (56).

1.3 Maschinelles Lernen als Methode

Im Jahre 1997 definierte Tom Mitchell maschinelles Lernen (engl.: machine learning) formal wie folgt:

„Definition: A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .“ (57)

Mit anderen Worten (angelehnt an (57,58)): Maschinelles Lernen bedeutet, dass ein Algorithmus durch das Lernen aus Daten Wissen generiert. Durch das Identifizieren von Mustern in den vorliegenden Daten wird ein Modell, das den Zusammenhang zwischen den Prädiktoren und Zielvariablen abbildet, erstellt. Dabei müssen die genauen Parameter und Berechnungsvorschriften des Modells nicht vollständig durch den Anwender festgelegt werden. Vielmehr lernt der Algorithmus diese in einem Trainingsprozess anhand von Beispielen selbst – deshalb der Name.

1.3.1 Begriffserklärungen

Im Folgenden werden wichtige Begriffe des maschinellen Lernens definiert und erläutert. Im Hinblick auf die weitgehend englischsprachige Nomenklatur im Bereich des maschinellen Lernens werden die englischen Bezeichnungen ebenfalls genannt.

Definition eines prädiktiven Modells

Die Analyse von Korrelationen innerhalb einer Stichprobe (engl. Bezeichnung: in-sample correlations) stellt den Hauptteil der heutigen psychiatrischen Forschung dar (25,59). Valide prädiktive Modelle zeichnen sich durch ihre Generalisierbarkeit auf neue Individuen aus, die nicht aus der Stichprobe stammen (engl. Bezeichnung: out-of-sample data), die zur Entwicklung des Modells genutzt wurde (59–61). Erstgenannte Korrelations-Studien führen im allgemeinen zu einer Überschätzung der Effektstärken, sodass prädiktive Modelle notwendig sind, um die translationale bzw. klinische Relevanz der Erkenntnisse einordnen zu können (59,60,62). In der Regel bauen prädiktive Modelle auf Mechanismen auf, die zuvor in Korrelations-Studien erforscht wurden (59,60). Der Begriff Prädiktion wird in dieser Arbeit also in Bezug auf generalisierbare Modelle verwendet, die anhand von Testdaten evaluiert werden.

Lernstile

Neben bestärkendem Lernen (engl.: reinforcement learning) werden üblicherweise zwei verschiedene Lernstile unterschieden (58,63): überwachtes (engl.: supervised learning) und unüberwachtes Lernen (engl.: unsupervised learning). In dieser Arbeit wird ausschließlich überwachtes Lernen verwendet, weshalb andere Lernstile nicht näher erläutert werden. Überwachtes Lernen zeichnet sich dadurch aus, dass die Zielvariable, die prädiziert werden soll, in der Trainingsphase bekannt ist.

Klassifikations- und Regressionsprobleme

Im Bereich des überwachten Lernens werden Klassifikations- und Regressionsprobleme unterschieden. Bei Klassifikationsproblemen steht die Zuordnung zu einer diskreten Klasse im Fokus (bspw. krank/gesund), wohingegen bei Regressionsproblemen eine kontinuierliche Variable (bspw. Alter) prädiziert wird (58,63).

Lernverfahren

Im Zentrum eines Modells im Bereich des maschinellen Lernens steht das eingesetzte Lernverfahren (engl.: estimator), das je nach Anwendungsfall variiert werden kann (63). Es gibt viele unterschiedliche Lernverfahren (z. B. Neuronale Netze, Entscheidungsbäume usw.) (63), weshalb im Methodenteil (Kapitel 2.3) nur die für diese Arbeit relevanten vorgestellt werden. Neben dem eigentlichen Lernverfahren enthält das Modell meist weitere Elemente bspw. zur Vorverarbeitung der Daten (63,64), welche im Folgenden Transformatoren (engl.: transformer) genannt werden.

Trainingsprozess, Bias und Varianz

Im Trainingsprozess werden die Parameter des eingesetzten Lernverfahrens (bspw. durch die Implementierung einer Kostenfunktion) so angepasst, dass der von dem Modell prädizierte Wert möglichst wenig von dem entsprechenden wahren Wert der Trainingsdaten abweicht (58,63). Wird das Modell durch die Wahl des Lernverfahrens und der Parameter allerdings zu spezifisch auf die Trainingsdaten angepasst (engl.: overfitting), ist die Generalisierbarkeit bei der Anwendung auf unbekannte Daten nicht mehr gegeben (58,63). Wird ein zu starres Modell generiert, kann der Zusammenhang zwischen den Prädiktoren und Zielvariablen nicht mehr präzise abgebildet werden (63).

Das optimale prädiktive Modell ist dabei eine Balance zwischen zu flexiblen Modellen (niedriger Bias, hohe Varianz) und zu starren Modellen (hoher Bias, niedrige Varianz) (engl. Bezeichnung: bias-variance-tradeoff) (63). Als Bias (Verzerrung) werden Abweichungen zwischen den Prädiktionen und den wahren Werten durch die Wahl des Lernverfahrens, als Varianz die Sensitivität des Modells gegenüber dem Datenset bezeichnet (63). Zum Training des Modells eingesetzte Samples werden im Folgenden als Trainingsset (Trainingsdaten) bezeichnet.

Hyperparameter

Neben den Parametern, die interne Variablen des Lernverfahrens sind und wie beschrieben im Lernprozess angepasst werden, existieren spezifische globale Einstellungsparameter für die einzelnen Lernverfahren (sog. Hyperparameter) (58,63,64). Die Anzahl und Art der Hyperparameter unterscheidet sich je nach Lernverfahren (63,64). Die Werte der einzelnen Hyperparameter werden vor dem Trainingsprozess festgelegt. Als Hyperparameteroptimierung wird die optimale Auswahl dieser Hyperparameter bezeichnet (63,64). Der Prozess der Hyperparameteroptimierung basiert in dieser Arbeit auf dem Testen verschiedener Hyperparameterkombinationen (alternativ auf Expertenwissen oder Erfahrung) (63,64). Dazu wird der Trainingsprozess mehrfach mit verschiedenen Hyperparameterkombinationen durchgeführt (17,64). Anhand im Training unbekannter Daten kann im Validierungsprozess das Modell mit der entsprechenden Hyperparameterkombination gewählt werden, das die beste Performance (gemessen an der entsprechenden Metrik) bietet (17,58,64). Ein Sample, das zur Bewertung und zur Auswahl einer Hyperparameterkombination genutzt wird, wird im Folgenden Validierungsset genannt.

Evaluation und Kreuzvalidierung

Nach erfolgreichem Trainings- und Hyperparameteroptimierungsprozess wird das Modell anhand von dem Modell bisher unbekanntem Daten evaluiert (17,58). Ein solches zur Evaluation genutztes Sample wird in dieser Arbeit als Testset (Testdaten) bezeichnet. Idealerweise wird zur finalen Evaluation der Generalisierbarkeit des Modells ein vollkommen separates externes Testset z. B. aus einer anderen Stichprobe/Studie genutzt (17,58).

Im Rahmen der Modellentwicklung stammt das Testset allerdings meist aus derselben Stichprobe, die auch zum Training und zur Hyperparameteroptimierung genutzt wird (17,58). Die vorliegenden Daten werden dazu üblicherweise vor dem Beginn des Trainings-/Evaluationsprozesses in sich nicht überschneidende Gruppen zum Training und zum Testen sowie je nach Modell auch zur Hyperparameteroptimierung aufgeteilt (17,58). Um einem Bias durch einmalige Ziehung solcher Subgruppen aus der Stichprobe vorzubeugen, ist im Bereich des maschinellen Lernens eine Kreuzvalidierung üblich (17,58,63). Dabei wird die Aufteilung in Trainings-/Validations- und Testset mehrfach zufällig durchgeführt, sodass das Modell anhand unterschiedlicher Samples trainiert und getestet wird und am Ende die Performance aller Durchgänge gemittelt werden kann (17,58,63). Es gibt verschiedene Arten der Kreuzvalidierung, wie z. B. die verschachtelte Kreuzvalidierung (engl.: nested cross-validation) und die Leave-One-Out-Kreuzvalidierung (engl.: leave-one-out cross-validation) (17,24,58).

Weitere Begriffe

Aus Gründen der besseren Nachvollziehbarkeit werden an dieser Stelle noch einige weitere relevante Begriffe definiert (in Anlehnung an (58,63,64)): Die Eingabedaten, auf deren Basis das Modell die Vorhersagen trifft, werden im Folgenden Prädiktoren (engl.: features/predictors) genannt. Die Variablen oder Klassen, die das Modell prädiziert, werden Zielvariablen/Label (engl.: target variable bzw. label) genannt. Der englischsprachige Begriff Machine Learning bleibt im Hinblick auf einen besseren Lesefluss bewusst neben der deutschsprachigen Bezeichnung erhalten. In Kapitel 2.3 findet sich eine detaillierte Beschreibung der in dieser Arbeit angewendeten Methoden des maschinellen Lernens.

1.3.2 Maschinelles Lernen in der psychiatrischen Forschung

Methodologisch gesehen werden in der psychiatrischen Forschung überwiegend gruppenstatistische Analysen durchgeführt (17,18,25,65). Maschinelles Lernen als Methode findet sich vor allem in der jüngeren Literatur (31,66,67).

Gründe für den Einsatz von maschinellem Lernen

Wie in Kapitel 1.1.3 am Beispiel der MDD dargestellt, haben psychiatrische Erkrankungen eine sehr heterogene und individuelle Natur. Neuere Studien implizieren zudem, dass unterschiedliche Subtypen der bekannten psychiatrischen Erkrankungen existieren, die sich bspw. bezüglich der Symptomatik (68) und des Therapieansprechens (69) unterscheiden. Vorherige Arbeiten in der psychiatrischen Forschung haben gezeigt, dass einzelne Variablen (univariate Ansätze) meist keine ausreichend stabile Grundlage zur Vorhersage bestimmter Krankheitsaspekte bieten, sondern verschiedene Prädiktoren zum Einsatz kommen sollten (25,31,66,70). Das ist in anderen medizinischen Disziplinen mitunter anders, wo bspw. für die Therapieentscheidung bei Krebspatienten einzelne annähernd perfekte Biomarker identifiziert werden konnten (18). Die beschriebene Heterogenität bildet sich ebenfalls in der Art und Menge klinisch erhobener Daten ab und kann mittels klassischer Statistik kaum erfasst werden (17–19). Machine Learning-Algorithmen sind dagegen in der Lage, hochkomplexe heterogene Daten simultan auszuwerten (multivariate Ansätze), um individuelle Muster zu identifizieren und basierend auf den eigenen Rückschlüssen personalisierte Vorhersagen zu treffen (17–19,65). Im Hinblick auf die steigende Anzahl an verfügbaren Daten birgt die Methode das Potential, die Medizin grundlegend zu verändern (17–19).

Anwendungsbereiche

Im Bereich der Psychiatrie lassen sich dabei u. a. drei große Anwendungsbereiche für prädiktive Modelle identifizieren (19,25), die im Folgenden beispielhaft vorgestellt werden:

1. Individuelle Prädiktion der Wirksamkeit einer therapeutischen Intervention,
2. Verbesserung der Differenzialdiagnostik von psychiatrischen Erkrankungen,
3. Vorhersage von individuellen Risiken im Zuge der Therapieplanung.

1. Die Abwägung, ob eine therapeutische Intervention (z. B. pharmakologisch oder psychotherapeutisch) bei einem individuellen Patienten wirkt, stellt bei psychiatrischen Erkrankungen im Hinblick auf die zahlreichen Nebenwirkungen pharmakologischer Therapien und die Compliance der Patienten eine große Herausforderung dar (25). In der Literatur konnte

durch maschinelles Lernen die Wirksamkeit von Antidepressiva anhand verschiedener Modalitäten (u. a. MRT) (66) und die Effektivität der Elektrokrampftherapie bei MDD-Patienten anhand von strukturellen MRT-Daten (71) prädiziert werden. Zudem wurden behandlungsresistente Depressionsverläufe anhand von klinischen Daten identifiziert (72). Dieser Anwendungsbereich bietet die Chance, den derzeit vielerorts gegenwärtigen Trial-and-Error-Prozess im Zuge der Ermittlung der individuell optimalen Therapie zu ersetzen (25).

2. Die Differenzialdiagnose psychiatrischer Störungen im klinischen Umfeld ist nicht zuletzt aufgrund von fehlenden oder unspezifischen Biomarkern (24,73), sich überschneidender genetischer Genese bei unterschiedlicher klinischer Ausprägung (74), hoher Komorbidität von verschiedenen psychiatrischen Krankheitsbildern (3,4) sowie einer mangelnden Präzision in der Erhebung von Symptomen sehr komplex (24,75) und stellt eine große Herausforderung dar (17,25). Das betrifft z. B. die Unterscheidung zwischen einer Angststörung und einer MDD (76) oder zwischen bi- und unipolaren Patienten vor Ausbruch einer manischen Episode (25,77). Durch die Anwendung von maschinellem Lernen gelang in der Literatur bspw. die Unterscheidung von MDD-Probanden und gesunden Kontrollen anhand von funktionellen MRT-Daten (67,78), EEG Daten (79) oder der Dynamik der Herzfrequenz (80). Auch hinsichtlich der Vorhersage der Krankheitsschwere wurden mittels Machine Learning erste Fortschritte erzielt (81,82).

3. Im Hinblick auf den rezidivierenden episodenhaften Verlauf von affektiven Erkrankungen (11) ist die Prädiktion individueller Risiken und somit die gezielte Anwendung von präventiven Therapien ein wichtiger Anwendungsbereich (25). In der Literatur konnte mittels Machine Learning bspw. die zukünftige Beeinträchtigung des sozialen Funktionsniveaus von Patienten mit neu eingetretener depressiver Erkrankung anhand von klinischen Faktoren und MRT-Daten (83) oder das Risiko einer erneuten stationären Behandlung von MDD-Patienten innerhalb von zwei Jahren anhand von verschiedenen Prädiktoren (u. a. MRT-Daten, Biomarker, klinische Faktoren) (31) mit besserer Präzision als durch den Zufall prädiziert werden.

Die dargestellte Literatur unterstreicht die Chancen, die sich durch maschinelles Lernen in der psychiatrischen Forschung ergeben.

1.3.3 Kritische Analyse von maschinellem Lernen in der psychiatrischen Forschung

Eine kritische Betrachtung von Machine Learning-Studien im Bereich der psychiatrischen Forschung deckt neben den in Kapitel 1.3.2 dargestellten Chancen auch einige Risiken auf (25). Die ermittelten Modellgüten variieren in der Literatur deutlich und zudem lassen sich methodologische Inkonsistenzen feststellen, die zu einer systematischen Überschätzung der Ergebnisse führen können (17,58).

Kleine homogene Stichproben

Die Prädiktionen eines Machine Learning-Modells basieren auf den zum Lernen zur Verfügung stehenden Daten (17). Steht eine kleine homogene Stichprobe als Trainingsset zur Verfügung, so ist kritisch zu hinterfragen, inwiefern die Eigenschaften der Zielpopulation durch die Stichprobe repräsentiert sind und ob das Modell generalisierbar ist (17,18). Die Evaluation dieses Umstandes bleibt eine Herausforderung, da in der psychiatrischen Forschung meist keine unabhängigen Testsets zur Verfügung stehen (17,18). Werden kleine homogene Stichproben genutzt, kann dies zudem zu einer Überschätzung der Genauigkeit führen (84,85). Durch die Analyse verschiedener Stichprobengrößen wurde das in der Literatur auch für die Unterscheidung von MDD-Probanden und gesunden Kontrollen (Modalität: MRT-Daten) gezeigt (85).

Datenlecks

Datenlecks sind ein weiteres methodisches Problem, das zur Überschätzung der Genauigkeit führen kann. Wie in Kapitel 1.3.1 beschrieben, wird die vorhandene Stichprobe im Rahmen der Kreuzvalidierung in separate Trainings- und Validierungs-/Testsets aufgeteilt. Werden vor der Anwendung der Kreuzvalidierung bspw. fehlende Datenpunkte basierend auf Informationen der gesamten Stichprobe ersetzt, sind die Trainings- und Testdaten nicht mehr unabhängig voneinander (17). In der Evaluation anhand der Testdaten würde die Genauigkeit des Modells in diesem Fall überschätzt werden, da das Modell indirekt schon in der Lernphase Zugriff auf Daten aus dem Testset hatte (17). Das lässt sich durch eine unabhängige Implementierung der einzelnen Verarbeitungsschritte vermeiden (17).

Verhältnis zwischen Stichprobengröße und Prädiktorenanzahl

Entsprechend der in den Kapiteln 1.1.3 und 1.3.2 beschriebenen Heterogenität psychiatrischer Erkrankungen werden in der Forschung multivariate Analysen mit einer großen Anzahl an Prädiktoren (z. B. hunderttausende MRT-Voxel pro Proband) bei oftmals kleinem Stichprobenumfang (n) durchgeführt (17,25,86). Dieses ungleiche Verhältnis wird

auch Fluch der Dimensionalität (engl.: curse of dimensionality) genannt (25,63). Um diesem Ungleichgewicht Rechnung zu tragen, gibt es verschiedene technische Möglichkeiten zur Reduktion der Dimensionalität (25).

Klinische Anwendung von prädiktiven Modellen

Trotz der Fortschritte in der Forschung werden prädiktive Machine Learning-Modelle im (psychiatrischen) Gesundheitswesen bisher kaum eingesetzt (17). Um die prognostische Stabilität, die Nützlichkeit und die Performance im klinischen Alltag zu evaluieren, ist es unumgänglich, dass die Modelle in Zukunft vermehrt in der klinischen Praxis getestet und weiterentwickelt werden (17,19,58). Die psychiatrische Versorgung basiert auf dem persönlichen Kontakt zwischen den Therapeuten und den Patienten, sodass es unwahrscheinlich scheint, dass prädiktive Modelle den klinischen Entscheidungsprozess komplett ersetzen (17). Realistischer erscheint es, dass die Modelle zukünftig unterstützend eingesetzt werden, insofern sich zeigt, dass die Kombination aus Mensch und Technik signifikant bessere Ergebnisse erzielt (17).

1.3.4 Empfehlungen (Best Practices) für Projekte des maschinellen Lernens

Um in dieser Arbeit eine sinnvolle technische Architektur der Machine Learning-Modelle zu gewährleisten und um die Nachvollziehbarkeit des Entwicklungsprozesses zu ermöglichen, werden in diesem Kapitel entsprechende Empfehlungen (Best Practices) von Experten vorgestellt (25). Dies geschieht insbesondere im Hinblick auf die in Kapitel 1.3.3 vorgestellten Anwendungsrisiken.

Der Entwurf von validen Machine Learning-Modellen erfordert sowohl technisches Wissen im Bereich der Methode als auch Expertenwissen im entsprechenden Anwendungsgebiet (engl.: domain knowledge) (25). Methodisches Wissen ist auch deshalb entscheidend, da die Theorie zeigt, dass kein einzelner Algorithmus für alle denkbaren Probleme (universell) stets die im Vergleich optimale Performance bieten kann (sog. „no free lunch theorem“) (25,87). Zusätzlich besteht die Gefahr von Implementierungsinkonsistenzen. Erfahrung im Anwendungsgebiet ermöglicht es, die Zusammenhänge, die für das prädiktive Modell entscheidend sind, optimal aufzubereiten (25).

Damit ein prädiktives Modell für die klinische Praxis geeignet ist, sollte es einfach zu bedienen sein, idealerweise auf einer angemessenen (tendenziell kleinen) Anzahl an Prädiktoren

basieren, die leicht im klinischen Alltag zu erheben sind, und die Empfehlungen simpel und eindeutig darstellen (18).

In Anlehnung an Hahn et al. (25) und Cearns et al. (17) wird im folgenden Abschnitt eine strukturierte Vorgehensweise zur Entwicklung von prädiktiven Analyseprojekten im Bereich der psychiatrischen Forschung vorgestellt (s. auch Abbildung 1-1):

1. Initial gilt es, die Projektziele festzulegen und zu entscheiden, ob die Prädiktion der Zielvariable überhaupt notwendig ist, d. h. einen Mehrwert generiert. Zudem sollte der Geltungsbereich für das Modell abgesteckt werden (z. B. eine bestimmte Zielpopulation). Davon hängt die Stichprobe ab, die zur Entwicklung des Modells erhoben und genutzt wird. Nicht zuletzt sollte geprüft werden, ob der Zusammenhang zwischen den Prädiktoren und der Zielvariable auch in Zukunft stabil bleibt.

2. Der zweite Schritt umfasst die Datenerhebung, wobei zunächst identifiziert werden muss, welche Prädiktoren eine Vorhersage der Zielvariable erlauben. Das erfordert Expertenwissen bzw. die Erforschung theoretisch möglicher Zusammenhänge. Zusätzlich sollte sichergestellt werden, dass das Modell effizient implementiert wird (z. B. durch die Identifikation redundanter Daten). Außerdem sollte geprüft werden, ob zur Prüfung der gewünschten Generalisierbarkeit weitere unabhängige Datensätze notwendig sind. Die Wahl eines ausreichend großen Datensatzes wird empfohlen.

3. Im dritten Schritt steht das eigentliche theoretisch-mathematische Modell im Fokus. Hier stellt sich u. a. die Frage, welches Lernverfahren und welche Form der Kreuzvalidierung genutzt werden soll. Dabei wird die verschachtelte k-fache Kreuzvalidierung empfohlen. Zudem ist sicherzustellen, dass keine Datenlecks entstehen, wozu eine Pipeline-Architektur empfohlen wird.

4. Im vierten Schritt soll das validierte Modell dann in die klinische Praxis integriert werden. Von großer Bedeutung ist hier die Frage, in welcher Form es den verschiedenen Stakeholdern (Praxen, Kliniken usw.) zur Verfügung gestellt werden kann. Nicht zuletzt muss sichergestellt werden, dass das Modell auch in Zukunft valide funktioniert, was eine kontinuierliche Überwachung und Verbesserung anhand von neuen Daten erforderlich macht.

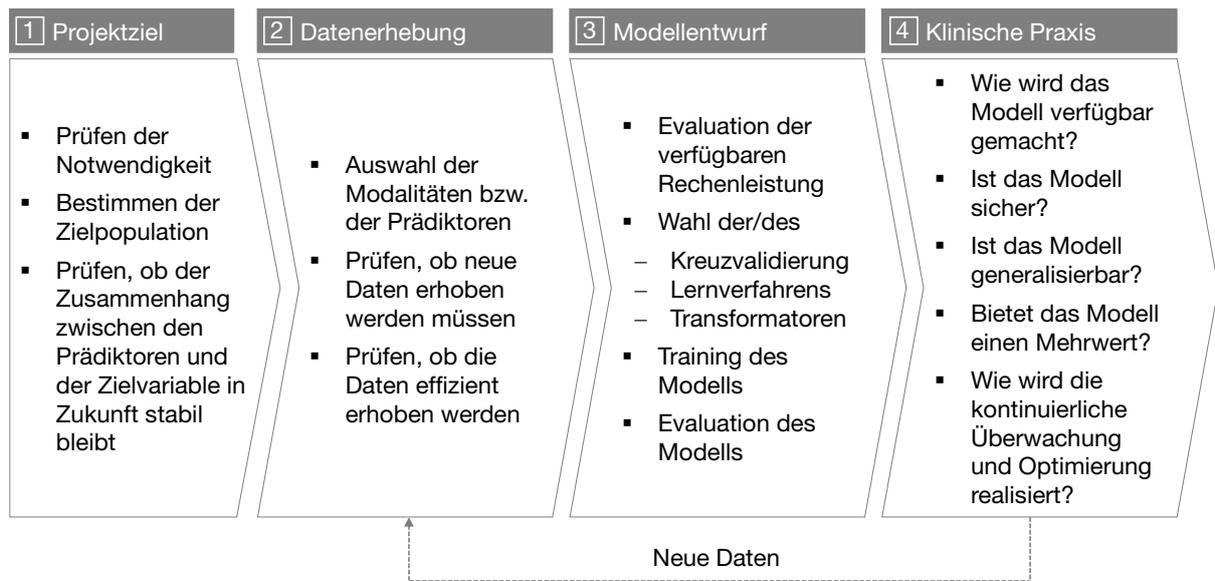


Abbildung 1-1: Entwicklungsprozess eines prädiktiven Modells (in Anlehnung an (17,25))

Anmerkungen: Entsprechend der Empfehlungen von Experten gliedert sich der Prozess zur Entwicklung eines prädiktiven Modells u. a. in die dargestellten Schritte. Das finale Modell sollte fortwährend mittels neuer Daten optimiert werden.

Den dabei entwickelten prädiktiven Modellen liegt in der Regel ein schwer nachvollziehbarer Algorithmus zugrunde (17). Die Nutzung technischer Verfahren, um ein Verständnis der relevantesten Parameter und der signifikantesten Prädiktoren zu erlangen, ermöglicht es Experten, fehlerhafte, triviale und unsichere Algorithmen bzw. Algorithmen mit systemischem Bias zu identifizieren (17). Nichtsdestotrotz sollte die Wahl der Prädiktoren schon im Entwicklungsprozess kritisch geprüft werden, anstatt eine retrospektive Bewertung vorzunehmen (17,58).

Vor allem die Bewertung der klinischen Anwendbarkeit eines prädiktiven Modells gilt explizit nicht für Forschungsarbeiten, deren Ziel das bessere Verständnis von ursächlichen Mechanismen oder der theoretische Nachweis der Funktionsfähigkeit neuer Konzepte (engl.: proof-of-concept) ist (17), wie es in dieser Arbeit der Fall ist.

1.4 Zusammenfassung und Hypothesen

Durch die hohe Lebenszeitprävalenz und den rezidivierenden Verlauf handelt es sich bei der MDD um eine Krankheit mit hoher Relevanz (s. Kapitel 1.1.2). Die pathophysiologischen Mechanismen, die dem heterogenen Krankheitsbild und dem episodenhaften Verlauf der MDD zu Grunde liegen, sind allerdings noch nicht abschließend geklärt (s. Kapitel 1.1.3). Deshalb sind unterschiedlichste Modalitäten Gegenstand der psychiatrischen Forschung (s.

Kapitel 1.2). Verfahren des maschinellen Lernens bieten – trotz aller bisherigen Hürden und Schwierigkeiten (s. Kapitel 1.3.3) – vor allem im Hinblick auf das Verständnis komplexer Zusammenhänge neue Chancen (s. Kapitel 1.3.2). Der Einsatz der Methode könnte zukünftig personalisierte Empfehlungen basierend auf individuellen Risikofaktoren ermöglichen und wäre somit hinsichtlich der Therapiewahl für den behandelnden Therapeuten von großer Bedeutung (s. Kapitel 1.1.4.).

Erarbeitung der Fragestellung dieser Studie

Ziel dieser Arbeit ist der Einsatz von maschinellem Lernen, um das Verständnis über mögliche Prädiktoren von individuellen Krankheitsverläufen zu stärken und somit einen Grundstein für einen optimierten Therapieeinsatz und einen mildereren Krankheitsverlauf zu legen. Um eine aussagekräftige und tiefgründige Analyse zu ermöglichen, fand die Beschränkung auf zwei Modalitäten (neuropsychologische Testdaten und strukturelle MRT-Daten) statt, deren Wahl im folgenden Abschnitt begründet wird:

In der Literatur konnte anhand von beiden Modalitäten nachgewiesen werden, dass Unterschiede zwischen nicht psychiatrisch erkrankten Probanden und Patienten mit einer MDD bestehen (s. Kapitel 1.2.1 und 1.2.2). Auch ein Zusammenhang mit der Depressionsschwere wird diskutiert. Zusätzlich kann von einem prädiktiven Wert beider Modalitäten hinsichtlich des Krankheitsverlaufes ausgegangen werden. Bisherige Studien implizieren z. B., dass die zukünftige Beeinträchtigung des sozialen Funktionsniveaus von Patienten mit neu eingetretener depressiver Erkrankung (83) bzw. das Risiko einer erneuten stationären Behandlung von MDD-Patienten innerhalb von zwei Jahren (31) mit den gewählten Modalitäten (als Teil eines multimodalen Modells) korreliert sind (s. Kapitel 1.3.2).

Es handelt sich dabei zudem sowohl um objektive Daten (strukturelle MRT-Daten) als auch um eine Modalität, die die Mitarbeit und kognitive Einsatzbereitschaft der Probanden voraussetzt (neuropsychologische Testung). Die Daten beider Modalitäten können zudem im klinischen Alltag ohne erheblichen Mehraufwand gewonnen werden, was im Hinblick auf die Expertenempfehlungen zur Entwicklung von prädiktiven Modellen (s. Kapitel 1.3.4) relevant ist. Gleichzeitig werden die Zusammenhänge zwischen kognitiven Beeinträchtigungen (s. Kapitel 1.2.1) bzw. strukturellen Hirnveränderungen (s. Kapitel 1.2.2) und der MDD in der Literatur kontrovers diskutiert, sodass weiterer Forschungsbedarf besteht.

Konkretisierung und Formulierung der Hypothesen

Zur Konkretisierung der Fragestellung dieser Arbeit wurden die folgenden Hypothesen aufgestellt:

1. Patienten mit einer MDD und gesunde Kontrollen lassen sich anhand der kognitiven Leistung bzw. der Gehirnstruktur voneinander unterscheiden.

In der Literatur wird für beide Modalitäten konsistent beschrieben, dass ein Gruppenunterschied besteht. Aufbauend auf diesem Mechanismus, ist es Ziel dieser Arbeit, zu prüfen, ob dieser Unterschied auch in den vorliegenden Daten vorhanden ist und ob dadurch eine Prädiktion der Gruppenzugehörigkeit auf Individualebene möglich ist.

2. Einzelne Krankheitscharakteristika der MDD lassen sich anhand der kognitiven Leistung bzw. der Gehirnstruktur präzisieren.

Die Klärung dieser Hypothese ist der Kern dieser explorativen Arbeit. Ob und bezüglich welcher Krankheitscharakteristika der MDD Korrelationen mit der kognitiven Leistung oder der Gehirnstruktur bestehen, ist nicht eindeutig geklärt. Dementsprechend orientierte sich die Auswahl der Krankheitscharakteristika, die in dieser Arbeit betrachtet werden, an unterschiedlichen in der Literatur beschriebenen Zusammenhängen, die einerseits fraglich und andererseits relevant sind.

Hinsichtlich der langfristigen Krankheitslast scheint der rezidivierende Verlauf der MDD eine maßgebliche Rolle zu spielen. Als ein zu analysierendes Charakteristikum wurde dementsprechend der zeitliche Krankheitsverlauf gewählt. Als weitere wichtige Maße der Krankheitslast wurden zudem die Depressionsschwere und das Funktionsniveau gewählt. In der Literatur wird ein Zusammenhang zwischen diesen Krankheitscharakteristika und den betrachteten Modalitäten diskutiert. Die explizite Wahl der Zielvariablen basierte neben der Literaturrecherche auch auf der Messbarkeit der Charakteristika und wird im Rahmen der Beschreibung der Studienmethodik detailliert erarbeitet (s. Kapitel 2) und in Kapitel 2.4 präzisiert.

Zur Überprüfung der Hypothesen wurden in dieser Arbeit prädiktive Analysen durchgeführt, d. h. die Modellgüten wurden anhand von Testdaten evaluiert, die den Modellen im Entwicklungsprozess unbekannt waren (s. Kapitel 1.3.1).

2 Daten und Methoden

In diesem Kapitel werden zunächst die Methoden der Datenerhebung vorgestellt (s. Kapitel 2.1). Darauf aufbauend wird die Stichprobenwahl (s. Kapitel 2.2) und die technische Umsetzung der Machine Learning-Modelle (s. Kapitel 2.3) erläutert. Abschließend werden die Analysen, die zur Klärung der Fragestellung dieser Arbeit durchgeführt wurden, beschrieben (s. Kapitel 2.4).

2.1 Die FOR 2107-Studie

Diese Arbeit ist auf Basis von Daten aus der von der Deutschen Forschungsgemeinschaft geförderten Forschergruppe 2107 (FOR 2107) entstanden. Aufgabe dieser Gruppe war bzw. ist es, die (Neuro-)Biologie affektiver Störungen (insbesondere der MDD und der bipolaren Störung) zu erforschen (37).

Im Rahmen des Teilprojekts 1 der FOR 2107 Studie wurde unter Leitung der Klinik für Psychiatrie und Psychotherapie am Universitätsklinikum Marburg sowie des Instituts für Translationale Psychiatrie an der Klinik für Psychische Gesundheit am Universitätsklinikum Münster eine große Kohortenstudie ins Leben gerufen. Erklärtes Ziel war der Einschluss von insgesamt mindestens 2500 Probanden, bestehend aus Patienten ($n = 1000$, davon MDD $n = 700$ und bipolare Störung $n = 300$), gesunden Probanden mit Umweltrisiko und/oder genetischem Risiko ($n = 500$) sowie einer gesunden Kontrollgruppe ohne bekannte Risikofaktoren ($n = 1000$). Zudem sollte die Stichprobe um Patienten mit psychotischen Störungen ($n = 300$) ergänzt werden (37).

Dabei wurden sämtliche Probanden im Rahmen des longitudinalen Studiendesigns zum Zeitpunkt des ersten Studieneinschlusses (Baseline, erster Einschluss am 11.09.2014) sowie nach Möglichkeit im Rahmen eines 2-Jahres- und eines 5-Jahres-Follow-Ups phänotypisiert. Zum Zeitpunkt der Erstellung dieser Arbeit wurden weiterhin neue Probanden rekrutiert, sodass die Studie noch nicht abgeschlossen war. Die Phänotypisierung, die alle Probanden im Rahmen der Studie standardmäßig durchliefen, umfasste gleichermaßen für alle Studienzeitpunkte eine multimodale MRT-Bildgebung, die Erhebung psychometrischer und klinischer Informationen anhand von Interviews und Fragebögen, die Durchführung einer neuropsychologischen Testung sowie die Analyse von Biomaterialien (Blut, Stuhl, Urin, Haare,

Speichel). Die Phänotypisierung fand dabei parallel in zwei Studienzentren an den Universitäten Münster und Marburg in Deutschland statt. Die Studie wurde von den jeweiligen Ethikkommissionen der Medizinischen Fakultäten der Universitäten Marburg (Datum: 07.02.2014, Aktenzeichen: Studie 07/14) und Münster (Datum: 01.08.2014, Aktenzeichen: 2014-422-b-S) bewilligt (s. Anhang 8.13 und 8.14).

Im Folgenden werden die Ein- und Ausschlusskriterien sowie der Studienablauf näher erläutert. Dabei liegt ein besonderer Fokus auf den für diese Arbeit genutzten Daten, sodass hier nicht relevante Inhalte der FOR 2107-Studie nicht näher beschrieben werden.

2.1.1 Ein-/ Ausschlusskriterien

Für einen Einschluss in die FOR 2107-Studie (37) mussten die Probanden zwischen 18 und 65 Jahren alt sowie westeuropäischer Herkunft sein und einen verbalen Intelligenzquotienten (IQ) von mind. 80 aufweisen. Zudem konnten nur Probanden eingeschlossen werden, bei denen keine MRT-Kontraindikationen (z. B. ferromagnetische Implantate, Herzschrittmacher, Klaustrophobie, Schwangerschaft usw.) vorlagen. Ausgeschlossen wurden Probanden mit schwerwiegenden neurologischen oder anderen somatischen Vorerkrankungen (z. B. Enzephalopathie, Multiple Sklerose, Demenz, Epilepsie, Krebserkrankungen, Autoimmunerkrankungen, chronisch-entzündliche Erkrankungen, kardiovaskuläre Erkrankungen usw.) sowie Probanden mit Vorerkrankungen, die im Zusammenhang mit substanzassoziierten Störungen (Missbrauch bzw. Abhängigkeit) standen. Eine Liste aller somatischen Erkrankungen, die zum Ausschluss führten, findet sich im Anhang (s. Anhang 8.1). Auch eine Rot-Grün- oder Farbblindheit führte zum Ausschluss.

Probanden wurden in die MDD-Gruppe eingeschlossen, wenn die Diagnose einer MDD im Strukturierten Klinischen Interview Achse-I (SKID-I) gestellt wurde (entsprechend der DSM-IV-Kriterien (296.xx)). Dabei war das Vorliegen einer Dysthymie oder Zykllothymia ohne das Vorliegen einer MDD kein hinreichendes Einschlusskriterium (als Komorbiditäten waren diese Erkrankungen allerdings erlaubt). Die Einnahme von Benzodiazepinen und Z-Substanzen führte zum Ausschluss (die letzte Einnahme musste > 3 Halbwertszeiten zurückliegen).

Gesunde Kontrollprobanden (eng. healthy controls, HC) wurden eingeschlossen, wenn keine zurückliegende oder akute SKID-I-Diagnose einer MDD oder einer bipolaren Störung vorlag (auch das Vorliegen einer Dysthymie oder Zyklotomia führten zum Ausschluss). Zudem wurde vorausgesetzt, dass zum Zeitpunkt der Untersuchung keine Einnahme von psychotropen

Medikamenten oder Benzodiazepinen erfolgte. Im Rahmen der FOR 2107-Studie wurde dabei auch erhoben, ob bei gesunden Probanden ein genetisches oder ein Umweltrisiko für eine psychiatrische Erkrankung bestand. Für diese Arbeit spielten diese Risikofaktoren keine Rolle und es galten die in diesem Abschnitt beschriebenen Kriterien.

2.1.2 Rekrutierung und Studienablauf

Die Rekrutierung der Probanden geschah auf freiwilliger Basis über Aushänge und Zeitungsanzeigen an den Standorten Münster und Marburg und ihrer Umgebung (s. Anhang 8.4 und 8.5). Die Rekrutierung von hospitalisierten Probanden wurde an den Psychiatrischen Kliniken der jeweiligen Universitätsklinik sowie an weiteren psychiatrischen Kliniken (u. a. LWL-Klinik Münster) in der Umgebung dieser Standorte nach einer Vorauswahl durch die jeweiligen Stationsärzte initiiert. Interessenten wurden im nächsten Schritt telefonisch oder im stationären Bereich auch persönlich kontaktiert und die Eignung zum Studieneinschluss anhand eines Screening-Fragebogens (s. Anhang 8.6) geprüft. Die Aufwandsentschädigung betrug für eine vollständige Teilnahme 70 Euro (ursprünglich 50 Euro). Zudem bekamen die Probanden eine elektronische Kopie der strukturellen MRT-Daten.

Bei Studieneinschluss wurden Termine für die MRT-Untersuchung, die Bioprobenentnahme sowie das klinische Interview mitsamt der neuropsychologischen Testung vereinbart. Im Rahmen des ersten Termins wurden die Probanden ausführlich über den Studienablauf und entsprechende Risiken aufgeklärt. Im Anschluss wurde das Einverständnis der Probanden für die verschiedenen Studienteile eingeholt (s. Anhang 8.7 und 8.8). Unmittelbar vor der MRT-Untersuchung wurde die MRT-Eignung durch einen medizinisch-technischen Radiologieassistenten bzw. einen geschulten Messleiter oder einen Radiologen validiert und schriftlich festgehalten (s. Anhang 8.9). Neben der Durchführung der einzelnen Vor-Ort-Studienteile bekamen die Probanden zusätzlich jeweils ein Fragebogenpaket sowie Material zur Gewinnung einer Stuhlprobe ausgehändigt, mit der Bitte, die entsprechenden Materialien innerhalb von 2 Wochen zu bearbeiten und zurückzusenden.

2.1.3 Erhebung der psychometrischen Daten und Diagnosestellung

Alle Probanden wurden durch geschulte Doktoranden/Masteranden anhand des Strukturierten Klinischen Interviews Achse-I (SKID-I), das sich an den Kriterien des DSM-IV orientiert (88), befragt. Die auf diese Weise ermittelte Forschungsdiagnose diente im weiteren Verlauf zur Einteilung der Probanden in die finalen Gruppen. Dabei wurde auch erhoben, ob MDD-Patienten eine akute depressive Episode hatten oder teil- bzw. vollremittiert waren.

Neben dem SKID-I umfasste die Phänotypisierung eine Reihe von Fragebögen, die teils im Rahmen eines Interviews (Fremdbeurteilung) und teils von den Probanden selbst (Selbstbeurteilung) bearbeitet wurden. In Tabelle 2-1 sind alle für diese Arbeit relevanten Fragebögen aufgelistet.

Tabelle 2-1: Übersicht über die eingesetzten Fragebögen

Psychometrischer Bereich	Eingesetzter Fragebogen	Abkürzung	Referenz
Funktionsniveau ^a	Global Assessment of Functioning	GAF	(89)
Klinischer Verlauf ^a	Checkliste klinischer Verlauf		
Depressionsschwere ^a	Hamilton Depressions-Skala	HAMD	(90)
Depressionsschwere ^b	Beck-Depressions-Inventar	BDI	(91,92)
Gesundheitsbez. Lebensqualität ^b	Short Form 36-Fragebogen	SF-36	(93,94)
Demographische Daten ^b	Soziodemografischer Bogen		

^aFremdbeurteilung, ^bSelbstbeurteilung

Global Assessment of Functioning (GAF) (angelehnt an (89))

Zur globalen Erfassung des Funktionsniveaus wurde die GAF-Skala (Bestandteil des DSM-IV) eingesetzt. Dabei wird die psychische, soziale und berufliche Funktionsfähigkeit beurteilt. Beeinträchtigungen durch körperliche Einschränkungen werden explizit nicht miteinbezogen. Die GAF-Skala bezieht sich dabei auf den aktuellen Zeitpunkt und umfasst einen Wertebereich $\in [1;100]$, wobei der Wert 100 einer optimalen Leistungsfähigkeit entspricht und jeder ganzzahlige Punktwert vergeben werden kann. Zur Unterstützung der bewertenden Person wird die Skala in 10 Funktionsniveaus, die jeweils einen Bereich von 10 Punkten umfassen, unterteilt. Zu jedem dieser Niveaus steht eine Liste beispielhafter Symptome und möglicher Funktionseinschränkungen bereit. Der Beurteiler hat die Aufgabe, das Funktionsniveau unter Berücksichtigung dieser Hilfestellung einzuschätzen. Sind Symptome vorhanden, die innerhalb verschiedener Funktionsniveaus gelistet sind, ist das schlechtere Niveau zu wählen.

Klinischer Verlauf und Hospitalisierung

Zur Beurteilung des klinischen Verlaufs der eingeschlossenen MDD-Patienten wurde u. a. die Anzahl und Dauer der depressiven Episoden, der Zeitpunkt des ersten Auftretens der Erkrankung sowie Informationen über (teil-)stationäre Klinikaufenthalte erhoben. Die Angaben

wurden (mit Ausnahme bei dem Vorliegen einer akuten Episode) retrospektiv erhoben. In der Regel wurden die Informationen im Anschluss an das SKID-I durch den identischen Testleiter abgefragt, sodass die Plausibilität der Angaben zweifach geprüft werden konnte.

Hamilton Depressions-Skala (HAMD) (angelehnt an (90))

Bei der in dieser Studie eingesetzten deutschen Adaption der HAMD (engl. Bezeichnung: Hamilton rating scale for depression) handelt es sich um ein Instrument zur Fremdbeurteilung des Schweregrades von depressiven Symptomen. Dabei werden typische Symptome einer depressiven Störung (z. B. depressive Stimmung, Leistungsverlust, Veränderung des Appetits/Gewichts oder der Libido, Ein- und Durchschlafstörungen, Antriebverlust usw.) im Rahmen eines drei- bis fünfstufigen Antwortformats abgefragt. Neben der Abfrage, ob ein Symptom besteht oder nicht, wird somit ebenfalls die Symptomschwere bewertet. In dieser Arbeit wurde ein strukturierter Interviewleitfaden angewendet (SIGH-ADS) (95) und der Test anschließend in seiner klassischen Form mit 21 Items ausgewertet. Als Bezugsrahmen wurden die letzten 14 Tage vor dem Interviewtermin gewählt.

Zur Auswertung wird ein Summenwert gebildet, dessen valider Wertebereich unter Anwendung der in dieser Studie genutzten Version im Bereich von 0 bis 64 liegt. Die Schwere der depressiven Symptomatik kann mittels des Summenwertes in drei Bereiche eingeteilt werden. Werte zwischen 0-10 entsprechen keiner bzw. einer leichten Depression, Werte von 10-20 deuten auf eine mittelschwere Depression hin wohingegen ein Summenwert > 25 einer schweren depressiven Episode entspricht. Da die eigentliche Schwere einer Depression vor allem mit einigen wenigen Items der Skala korreliert zu sein scheint, werden in der Literatur auch andere Einteilungsmöglichkeiten diskutiert (96).

Beck-Depressions-Inventar (BDI) (angelehnt an (91,92))

Bei dem BDI (engl.: Beck Depression Inventory) handelt es sich um ein Instrument zur Selbstbeurteilung der Schwere einer Depression. In dieser Arbeit kam eine deutsche Adaption des BDI zum Einsatz. Der Fragebogen umfasst insgesamt 21 Items, die typische Symptome einer depressiven Störung repräsentieren (z. B. depressive Stimmung, Reizbarkeit, Schlafstörungen, Libidoverlust usw.). Das Antwortformat hat jeweils 4 Stufen. Dabei wird erhoben, ob ein Symptom besteht bzw. wie stark ein Symptom ausgeprägt ist. Als Bezugsrahmen wurden hier die letzten 7 Tage vor dem Interviewtermin einschließlich des Tages der Bearbeitung gewählt. Der Fragebogen dient explizit nicht zur Diagnose einer MDD, sondern zur Erhebung der Depressionsschwere sowie zur Verlaufsbeurteilung dieser bei einer

bestehenden Diagnose. Im Rahmen dieser Arbeit wurde der Fragebogen bei allen Probanden zum Zweck des internen Vergleichs angewendet. Zur Auswertung wird ein Summenwert gebildet. Dieser Summenwert hat einen gültigen Wertebereich von 0 bis 63 und kann entsprechend der folgenden Intervalle interpretiert werden: 0-13: minimale Depression, 14-19: milde Depression, 20-28: moderate Depression, 29-63: schwere Depression.

Short Form 36-Fragebogen (SF-36) (angelehnt an (93,94))

Zur Selbstbeurteilung der gesundheitsbezogenen Lebensqualität (grundlegende soziale, psychische, körperliche und alltagsnahe Aspekte) kam der SF-36 zum Einsatz. Um verschiedene Dimensionen des subjektiven Gesundheitsempfindens bewerten zu können, umfasst der Test acht Subskalen (körperliche Funktionsfähigkeit, körperliche Rollenfunktion, generelles Gesundheitsempfinden, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion, psychisches Wohlbefinden, körperliche Schmerzen). Die einzelnen Subskalen haben verschiedene Antwortformate. In dieser Arbeit wurde nur die Subskala der körperlichen Funktionsfähigkeit (Beeinträchtigung der Funktion durch Veränderungen des körperlichen Gesundheitszustandes) ausgewertet. Diese Subskala beinhaltet 10 Items (z. B. mehr als einen Kilometer zu Fuß gehen, einen Treppenabsatz steigen, Einkaufstaschen heben oder tragen), wobei die Funktionsfähigkeit für jedes Item auf einer Skala von 1-3 bewertet wird. Der valide Wertebereich nach Bildung des Summenwertes beträgt für diese Subskala 0 bis 100. Je höher der Summenwert, desto höher ist das Gesundheitsempfinden/Funktionsniveau in diesem Bereich.

Soziodemografischer Bogen

Mithilfe eines Fragebogens wurden grundlegende demographische Daten wie das Geburtsdatum, das Geschlecht und der Bildungsgrad abgefragt.

2.1.4 Erhebung der kognitiven Leistung (neuropsychologische Testbatterie)

Die kognitive Leistung der Probanden wurde im Rahmen einer neuropsychologischen Testung in einem der beiden Studienzentren vor Ort untersucht. Zur Bewertung unterschiedlicher kognitiver Funktionsbereiche umfasste die Testbatterie insgesamt neun verschiedene papierbasierte Tests, die entsprechend der Reihenfolge der Durchführung in Tabelle 2-2 dargestellt sind. Es kamen sowohl Geschwindigkeitstests (engl.: speedtests) als auch Niveautests (engl.: powertests) zum Einsatz (97). Erstere erreichen eine Leistungsdifferenzierung durch die Begrenzung der Arbeitszeit, wohingegen bei letzteren schwierige Aufgaben gestellt werden, die nicht von allen Probanden richtig gelöst werden

können (97). Insgesamt wurden 55 verschiedene Variablen (s. Anhang 8.2) erhoben. Die Testung wurde durch geschulte Testleiter durchgeführt.

Tabelle 2-2: Übersicht über die neuropsychologische Testbatterie

Kognitiver Funktionsbereich	Eingesetzter Test	Abk.	Referenz
Merkfähigkeit	Verbaler Lern- und Merkfähigkeitstest	VLMT	(98)
Exekutive Funktionen/ Verarbeitungsgeschw.	Pfadfindertest A (engl.: Trail-Making-Test)	TMT-A	(99)
	Pfadfindertest B (engl.: Trail-Making-Test)	TMT-B	(99)
Visuelles Arbeitsgedächtnis	Blockspanne		(100)
Verbales Arbeitsgedächtnis	Buchstaben-Zahlen-Test	BZT	(101)
Exekutive Funktionen/ Verarbeitungsgeschw.	Zahlen-Symbol-Test	ZST	(102)
Aufmerksamkeit/ Konzentrationsfähigkeit	Test d2	d2	(103)
Semantische Verarbeitung	Regensburger Wortflüssigkeits-Test	RWT	(104)
Verbalintelligenz	Mehrfachwahl-Wortschatz-Intelligenztest	MWT-B	(105)

Verbaler Lern- und Merkfähigkeitstest (VLMT) (angelehnt an (98))

Um die Lern- und Merkfähigkeit der Probanden zu bewerten, wurde der VLMT eingesetzt. Es handelt sich dabei um eine deutsche Adaption des Rey Auditory Verbal-Test (106). Das Testmaterial beinhaltet zwei Wortlisten (A und B) mit jeweils 15 semantisch unabhängigen Wörtern sowie eine Wiedererkennungsliste, die neben den Wörtern der Listen A und B auch 20 weitere semantisch bzw. phonematisch ähnliche Wörter enthält.

Zunächst werden die in der Liste A enthaltenen Wörter fünfmal in einem gleichmäßigen Rhythmus durch den Testleiter vorgelesen. Die Aufgabe der Testperson besteht darin, nach jedem Durchgang möglichst viele Wörter wiederzugeben. Die Reihenfolge der Reproduktion ist dabei unerheblich. Neben der Anzahl der korrekt wiedergegebenen Worte (Reproduktionsleistung) wird auch protokolliert, ob einzelne Wörter mehrfach genannt wurden (Perseverationen) und/oder Worte genannt wurden, die nicht aus der Liste A stammen (Falsch Positive). Im nächsten Schritt wird die Wortliste B (Interferenzliste) einmal vorgelesen und die Anzahl der korrekt reproduzierten Wörter dieser Liste festgehalten. Direkt im Anschluss wird die Testperson darum gebeten, die Wörter aus der Wortliste A ohne erneute Darbietung durch

den Testleiter wiederzugeben (engl. Bezeichnung: Short Delay Free Recall). Neben der Summe der korrekt angegebenen Worte wird hierbei zudem protokolliert, wie viele Wörter aus der hierbei nicht abgefragten Liste B genannt wurden (Interferenzen).

Nach dem ersten Teil des VLMT wurden in der vorliegenden Studie zunächst andere Teile der Testbatterie durchgeführt (TMT-A/B, Blockspanne, BZT, ZST, d2) bevor der VLMT nach dieser etwa halbstündigen Verzögerung fortgesetzt wurde. Im zweiten Teil wird Wortliste A ohne Vorankündigung und ohne erneute Darbietung ein siebtes Mal abgefragt (engl. Bezeichnung: Long Delay Free Recall). Die Summe der richtig genannten Worte (Liste A) sowie die Anzahl der genannten Worte, die aus der Interferenzliste B stammen, werden protokolliert. Abschließend wird die Wiedererkennungsliste durch den Testleiter Wort für Wort vorgelesen. Die Aufgabe der Testperson besteht darin, zu entscheiden, ob das jeweils vorgelesene Wort aus der Wortliste A stammt oder nicht. Neben der Summe der Worte, die die Testperson richtigerweise der Liste A zugeordnet hat, wird zudem erhoben, wie viele Worte, die fälschlicherweise der Liste A zugesprochen wurden, in Wahrheit aus Liste B (Fehler Liste B) bzw. aus der Distraktionsliste (Fehler Neues Wort) stammen.

Zur Auswertung des Tests lassen sich verschiedene Summenwerte berechnen. Neben den oben dargestellten Rohwerten der einzelnen Durchläufe wurde hier die Summe der Richtigen bzw. der Perseverationen und Falsch Positiven aller fünf Durchgänge vor Interferenz, der Verlust an richtig wiedergegebenen Worten nach Interferenz bzw. nach zeitlicher Verzögerung, die Wiedererkennungsleistung nach Abzug der Fehler sowie die Summen der Interferenzen aus Durchgang sechs und sieben berechnet (s. Anhang 8.2). Die Durchführungsdauer beträgt ca. 20 bis 25 Minuten.

Pfadfinder-Test Teil A und B (TMT) (angelehnt an (99,107))

Der TMT besteht aus zwei Teilen. In Teil A wird vor allem die visuelle Such- und die motorische Bewegungsgeschwindigkeit untersucht. Die Bearbeitung des zweiten Teils (TMT-B) erfordert zusätzlich höhere kognitive Fähigkeiten (mentale Flexibilität) (108). Im ersten Teil (TMT-A) hat der Proband die Aufgabe, die Zahlen 1 bis 25 mit einem Stift in aufsteigender Reihenfolge zu verbinden. Im Teil B sind insgesamt 25 Elemente (die Zahlen 1 bis 13 und die Buchstaben A bis L) jeweils abwechselnd in numerisch aufsteigender bzw. alphabetischer Reihenfolge miteinander zu verbinden (1-A-2-B- usw.). Die Verbindungslinien der einzelnen Items sollen sich dabei nicht überschneiden. Sollte der Proband logische Fehler machen, so wird er durch den Testleiter darauf hingewiesen und muss zum letzten korrekt bearbeiteten Element

zurückkehren. Das maximale Zeitlimit für die Bearbeitung der beiden Teile beträgt jeweils 300 Sekunden. Die Durchführungsdauer für Teil A beträgt ca. 5 Minuten, die für Teil B ca. 5 bis 10 Minuten. Zur Auswertung wurden hier sowohl die Bearbeitungszeiten (in Sekunden) der Teile A und B separat protokolliert sowie die Differenz dieser beiden Werte gebildet.

Blockspanne (angelehnt an (100))

Zur Bewertung des visuellen Arbeitsgedächtnisses wurde die Blockspanne eingesetzt (engl.: spatial span). Der Test (deutschen Adaptation) ist in der revidierten Fassung der Wechsler Memory Scale enthalten. Bei der Blockspanne handelt es sich um ein Brett, auf dem neun Blöcke unregelmäßig angeordnet sind. Im Rahmen der Testung tippt der Testleiter unterschiedliche Block-Kombinationen mit einer konstanten Darbietungsgeschwindigkeit von ca. einer Sekunde pro Block an (zunächst zwei, dann drei und im Verlauf bis zu acht verschiedene Blöcke). Die Aufgabe der Testperson ist es, die Kombinationen unmittelbar nach der Darbietung in derselben Reihenfolge (vorwärts) zu reproduzieren. Im zweiten Teil hat der Proband die Aufgabe, die Blöcke in der umgekehrten Reihenfolge als in der Darbietung (rückwärts) anzutippen. Der Test wird abgebrochen, wenn der Proband keine der jeweils zwei Folgen pro Blockanzahl richtig wiedergeben konnte. Zur Auswertung wurde die Anzahl an korrekt reproduzierten Block-Kombinationen (separat für beide Teile) und die Summe aus diesen beiden Testteilen protokolliert.

Buchstaben-Zahlen-Test (BZT) (angelehnt an (101))

Der BZT, der hier zur Bewertung des verbalen Arbeitsgedächtnisses angewandt wurde, ist in der deutschsprachigen Adaption der Wechsler Adult Intelligence Scale enthalten. Die Aufgabe der Testperson besteht darin, vom Testleiter vorgelesene Kombinationen aus Buchstaben und Zahlen zu sortieren, sodass zunächst die Zahlen in numerisch aufsteigender Reihenfolge und im Anschluss die Buchstaben in alphabetischer Reihenfolge wiedergegeben werden. Insgesamt enthält der Test 24 verschiedene Kombinationen, wobei die Anzahl der Buchstaben und Zahlen im Verlauf gesteigert wird (jeweils vier Kombinationen à 2, 3, 4, 5, 6 bzw. 7 Elementen). Der Test wird abgebrochen, wenn der Proband alle vier Kombinationen einer bestimmten Anzahl an Elementen falsch reproduziert hat. Die Durchführung dauert ca. 5 bis 10 Minuten. Zur Auswertung für diese Studie wurde die Gesamtzahl an richtig bearbeiteten Kombinationen protokolliert.

Zahlen-Symbol-Test (ZST) (angelehnt an (102))

Zur Beurteilung der Verarbeitungsgeschwindigkeit wurde u. a. der ZST durchgeführt. Der Test ist Bestandteil des Hamburg-Wechsler-Intelligenztests für Erwachsene, bei dem es sich um eine deutschsprachige Adaption der Wechsler Adult Intelligence Scale handelt. Aufgabe der Testperson ist es, den Ziffern 1 bis 9 jeweils eines von neun vorgegebenen abstrakten Symbolen zuzuordnen. Der Proband hat 90 Sekunden, um eine Serie von bis zu 93 Items zu bearbeiten. Der Test ist in der vorgegebenen Reihenfolge abzuarbeiten. Vor Beginn der Zeitmessung steht ein Übungsbereich zur Verfügung (7 Items). Zur Auswertung wurde die Anzahl an Items erhoben, denen jeweils das richtige Symbol zugeordnet wurde. Die Durchführung dauert insgesamt weniger als 5 Minuten.

Test d2 (d2) (angelehnt an (103))

Die Aufmerksamkeits- und Konzentrationsfähigkeit der Probanden wurde mittels des Tests d2 beurteilt. Es handelt sich dabei um eine deutschsprachige Adaption des Interpersonal Reactivity Index. Der Test enthält insgesamt 14 Reihen mit jeweils 47 Zeichen. Bei den Zeichen handelt es sich jeweils um den Buchstaben p bzw. d. Die Buchstaben sind entweder oberhalb und/oder unterhalb mit jeweils ein bis zwei Strichen markiert (insgesamt also ein bis vier Striche pro Buchstabe). Aufgabe des Probanden ist es, sämtliche d anzustreichen, die mit insgesamt genau zwei Strichen markiert sind (Zielobjekt). Pro Reihe hat der Proband genau 20 Sekunden Zeit. Der Testleiter zeigt den Ablauf der Zeit jeweils an. Die dem Zielobjekt visuell ähnlichen Reize (Buchstabe p bzw. andere Anzahl an Strichen) dienen als Distraktoren und sollen nicht markiert werden. Die Durchführungszeit beträgt ca. 8 Minuten.

Neben der Gesamtzahl der bearbeiteten Zeichen wurde für die Auswertung sowohl der Auslassfehler (nicht durchgestrichene Zielobjekte, Fehler 1. Art), der Verwechslungsfehler (durchgestrichene Distraktoren, Fehler 2. Art) sowie die Konzentrationsleistung (Summe der Anzahl der richtig durchgestrichenen Zielobjekte pro Zeile abzüglich des jeweiligen Fehlers 2. Art) ermittelt.

Regensburger Wortflüssigkeits-Test (RWT) (angelehnt an (104))

Die formallexikalische und semantische Wortflüssigkeit der Probanden wurde mit dem RWT getestet. Dabei kamen in dieser Studie insgesamt drei Varianten des Tests zum Einsatz. Aufgabe des Probanden ist es, innerhalb von einer Minute möglichst viele Wörter zu nennen, die einer bestimmten Kategorie angehören (1. Variante: Kategorie Tiere) bzw. mit einem bestimmten Buchstaben beginnen (2. Variante: Anfangsbuchstabe P) bzw. immer abwechselnd einer von zwei verschiedenen Kategorien angehören (3. Variante: alternierende

Kategorien Sportarten und Früchte). Genannt werden sollen nur Worte, die in der deutschen Sprache existieren (falls nicht erfüllt, wird das Wort als Fehler gewertet). Mehrfachnennungen und Nennung von verschiedenen Wörtern, die den gleichen Wortstamm enthalten (Perseverationen) sind zudem zu vermeiden. In der Variante, in der der Anfangsbuchstabe vorgegeben wird, ist zudem die Nennung von Eigennamen und geographischen Orten nicht erlaubt. Die Wiederholung von Worten in der Einzahl bzw. Mehrzahl oder die Nennung von Verben in verschiedenen Formen soll hierbei ebenfalls vermieden werden. Im Rahmen der Auswertung wird in der Variante der alternierenden Kategorien die Nennung von mehreren Worten desselben Wortstammes in verschiedenen Durchgängen nicht als Fehler gewertet.

Die Durchführungszeit beträgt ca. 10 Minuten. Im Rahmen der Auswertung für diese Studie wurden sowohl die Rohwerte (Anzahl der genannten Worte, Fehler und Perseverationen) sowie die korrigierten Werte (Anzahl genannter Worte abzüglich der jeweiligen Fehler und Perseverationen) für jede Variante separat ermittelt.

Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B) (angelehnt an (105))

Zur Beurteilung des allgemeinen verbalen Intelligenzniveaus wurde der MWT-B durchgeführt. Der Test besteht aus insgesamt 37 mit aufsteigendem Schwierigkeitsgrad angeordneten Items. Jedes Item enthält eine Wortliste mit jeweils fünf Wörtern (vier fiktive Neukonstruktionen und ein umgangs- oder wissenschaftssprachliches Wort). Aufgabe der Testperson ist es, zu kennzeichnen, welches Wort der jeweils fünf Wörter pro Zeile in der deutschen Sprache existiert. Die Durchführung des Tests dauert ca. 4 bis 6 Minuten. Für die weiteren Berechnungen in dieser Arbeit wurde die Anzahl der richtig beantworteten Items entsprechend des Manuals in den verbalen IQ umgerechnet (Wertebereich: zwischen 61 und 145).

2.1.5 Erhebung der strukturellen MRT-Daten

Die MRT-Untersuchung der Probanden im Rahmen der FOR 2107-Studie (37) setzte sich aus sechs verschiedenen Sequenzen zusammen: eine hochauflösende T1-gewichtete anatomisch-strukturelle Bildgebung, eine Diffusion-Tensor-Imaging-Sequenz sowie insgesamt vier verschiedene funktionelle MRT-Paradigmen. Die Untersuchung wurde jeweils an einem Stück durchgeführt und dauerte ca. 60 Minuten.

Vor Beginn der Messung wurde die Kommunikationsfähigkeit der Probanden mit dem Testleiter über eine Gegensprechanlage sichergestellt. Den Probanden wurde ein entsprechender Gehörschutz zur Verfügung gestellt. Um im Falle von auftretenden

Schwierigkeiten schnell reagieren zu können, wurden die Probanden über die Funktionsweise des Notfallknopfes aufgeklärt und dieser vor dem Einfahren in das MRT vor Ort geprüft. Zur Vermeidung von Bewegungsartefakten während der MRT-Messung wurden die Probanden instruiert, sich möglichst ruhig zu verhalten, und ihr Kopf mit Hilfe von Polstern weitestgehend bewegungsfrei gelagert. Für die strukturelle Messung durften die Probanden sich vor Beginn der Messung entscheiden, ob sie ihre Augen für die gesamte Dauer dieses Teils der Messung offen oder geschlossen halten möchten. Für diese Arbeit wurden ausschließlich Daten der T1-gewichteten strukturellen Bildgebung genutzt, weshalb die anderen fünf MRT-Sequenzen nicht näher beschrieben werden.

Technische Spezifikationen

Die MRT-Daten wurden an den beiden Studienzentren mittels unterschiedlicher Hard- und Software-Konfigurationen erhoben. Aufgrund der typischen englischsprachigen Bezeichnungen der technischen Parameter, werden diese hier ebenfalls aufgezählt. In Münster wurde ein 3-Tesla MRT-Scanner (Prisma, Siemens, Erlangen, Deutschland) zusammen mit einer 20-Kanal-Kopfspule (engl. Bezeichnung: head matrix Rx-coil) eingesetzt, in Marburg kam ein sehr ähnlicher 3-Tesla MRT-Scanner (Tim Trio syngo MR B17, Siemens Erlangen, Deutschland) mit einer 12-Kanal-Kopfspule zum Einsatz (37,109). Die Parameter der Pulssequenzen wurden – soweit die technischen Plattformen dies zuließen – für beide Standorte standardisiert, um die Vergleichbarkeit zu ermöglichen (37,109).

Die T1-gewichteten Daten wurden dabei mit einer MPRAGE-Sequenz (magnetization prepared rapid acquisition gradient echo) erhoben (37,109). Die genauen technischen Daten für Münster (Repetitionszeit (engl.: repetition time): 1900 ms, Echozeit (engl.: echo time): 2.28 ms, Inversionszeit (engl.: inversion time): 900 ms, Anregungswinkel (engl.: flip angle): 8°, Anzahl der sagittalen Schnittbilder (engl.: sagittal slices): 192, Schnittbildabstand (engl.: slice gap): 0 mm, Voxelgröße (engl.: voxel size): 1x1x1 mm³) und Marburg (Repetitionszeit: 1900 ms, Echozeit: 2.26 ms, Inversionszeit: 900 ms, Anregungswinkel: 9°, Anzahl der sagittalen Schnittbilder: 176, Schnittbildabstand: 0.5 mm, Voxelgröße 1x1x1 mm³) unterschieden sich dabei marginal (37,109).

Vorverarbeitung der MRT-Daten

Im weiteren Verlauf durchliefen alle MRT-Daten die Vorverarbeitungsprozedur, die in dieser Form für die FOR 2107-Studie standardisiert war (37). Für diese Arbeit wurde die kortikale und subkortikale Parcellierung der Gehirnschans (basierend auf dem Desikan-Killiany Atlas

(52,110,111) voll automatisiert mithilfe der Freesurfer Software (Version 5.3) durchgeführt (<https://surfer.nmr.mgh.harvard.edu/>) (111,112).

Dabei wurden die voreingestellten Standardparameter verwendet und ein Großteil der innerhalb der Software verfügbaren Berechnungen zur Analyse von Volumina, Oberflächengrößen und Schichtdicken durchgeführt. Die finale Anzahl der in die Analyse miteinbezogenen Variablen betrug 166 (s. Anhang 8.3).

Qualitätskontrolle

Die automatische Segmentierung wurde im Rahmen eines an das Standardprotokoll des ENIGMA-Konsortiums (<http://enigma.ini.usc.edu/protocols/imaging-protocols>) (51,52,113) angelegten Qualitätschecks durch Übereinanderlegen der errechneten Segmentlabel und der T-1 gewichteten Gehirnschichten von geschulten Experten visuell überprüft (37,114). Dabei identifizierte Hirnpathologien, die den Ausschlusskriterien entsprachen (s. Kapitel 2.1.1), führten zum nachträglichen Ausschluss der entsprechenden Probanden (37). Ebenfalls wurden Probanden, die deutliche anatomische Veränderungen ohne Krankheitswert aufwiesen oder bei denen die Bildgebung fehlerhaft verlief, in dieser Arbeit ausgeschlossen (nur für die Modalität: strukturelle MRT-Daten) (37). Bei statistischen Ausreißern (Hirnvolumen > 2.698 Standardabweichungen vom Mittelwert der Kohorte entfernt) wurde eine Korrektur durchgeführt oder die Probanden entsprechend des obigen Protokolls ausgeschlossen (51,52,113). Zusätzlich zu der nachträglichen Qualitätskontrolle wurde regelmäßig eine Phantommessung durchgeführt (114).

2.2 Finale Stichprobenwahl und Stichprobenstatistik

Wie in Kapitel 1.3.3 beschrieben, hat die Größe der Stichprobe bei der Anwendung von maschinellem Lernen einen maßgeblichen Einfluss auf die Qualität der Ergebnisse. Eine adäquat große Stichprobe lag zum Analysezeitpunkt nur für den ersten Untersuchungszeitpunkt vor, weshalb auf eine differenzierte Analyse der longitudinalen Studiendaten verzichtet wurde.

Separate Stichproben für beide Modalitäten

Grundsätzlich wurde für beide Modalitäten jeweils vorausgesetzt, dass mindestens 90 % der Prädiktoren vorlagen. Da nicht alle Probanden in die MRT Messung eingeschlossen werden konnten, ergab die separate Anwendung dieser Voraussetzung für beide Modalitäten

unterschiedlich große Stichproben. Aus Gründen der Vergleichbarkeit würde eine einheitliche Stichprobe sicherlich Vorteile bieten. Das Bilden der größten gemeinsamen Schnittmenge beider Stichproben (sprich das gleichzeitige Vorliegen von mind. 90 % der Prädiktoren beider Modalitäten) hätte aber eine deutlich kleinere Stichprobe ergeben als bei einer separaten Betrachtung möglich war. Um jeweils die maximal verfügbare Anzahl an Probanden untersuchen zu können, wurden die Modalitäten entsprechend separat betrachtet.

Innerhalb der Analyse der MDD-Patienten beider Modalitäten wurden verschiedene Zielvariablen prädiziert (s. Kapitel 2.4). Als Voraussetzung von überwachtem maschinellem Lernen gilt, dass die Zielvariablen für jeden betrachteten Probanden vorliegen müssen. Nicht für alle eingeschlossenen Probanden lagen diese Zielvariablen immer konsistent vor. Eine explizite Sichtung der Daten ergab, dass der GAF-Wert bei einer großen Anzahl an Probanden nicht erhoben wurde, wohingegen die anderen Zielvariablen weitestgehend vollständig waren.

Um auch für die Prädiktionen der unterschiedlichen Zielvariablen möglichst große Stichproben analysieren zu können, wurden innerhalb der beiden Modalitäten jeweils zwei weitere separate Stichproben gebildet. In der kleineren Stichprobe wurde das Vorliegen aller Zielvariablen vorausgesetzt, während in der größeren Stichprobe auf das Vorliegen des GAF-Wertes verzichtet wurde. Für die Kontrollgruppe war eine solche weitere Unterteilung nicht notwendig, da hier keine Zielvariablen prädiziert wurden.

Zusammengefasst bedeutet das also, dass ein Proband aus der entsprechenden Stichprobe ausgeschlossen wurde, wenn mehr als 10 % der Prädiktoren oder die untersuchten Zielvariablen nicht vorlagen. In den Analysen, in denen der GAF-Wert nicht prädiziert wurde, wurde das Vorliegen dieses Wertes nicht vorausgesetzt.

Prozess der Stichprobengenerierung

Zum Zeitpunkt dieser Arbeit lagen bei Betrachtung des ersten Studienzeitpunktes die Daten von insgesamt 2479 Probanden vor. Davon waren 1050 Probanden Teil der gesunden nicht-psychiatrisch erkrankten Kontrollgruppe während bei 987 Probanden die Hauptforschungsdiagnose MDD (anhand des SKID-I) gestellt wurde. Von dieser Stichprobe ausgehend wurden alle Probanden ausgeschlossen, bei denen klinische Gründe vorlagen oder deren direkte Vorfahren nicht aus dem geographischen Einschlussgebiet stammten.

Zudem wurden Probanden aufgrund der Einnahme von Benzodiazepinen, von Stimulanzien und von Z-Substanzen ausgeschlossen (s. Kapitel 2.1.1).

Wie in Abbildung 2-1 schematisch dargestellt, wurde das Vorliegen der entsprechenden Zielvariablen (HAMD, BDI, SF-36, klinischer Verlauf, GAF) sowie das Vorliegen demographischer Angaben (u. a. Alter, Geschlecht, Bildungsgrad) im nächsten Schritt durch eine entsprechende Filterung der Datensätze sichergestellt.

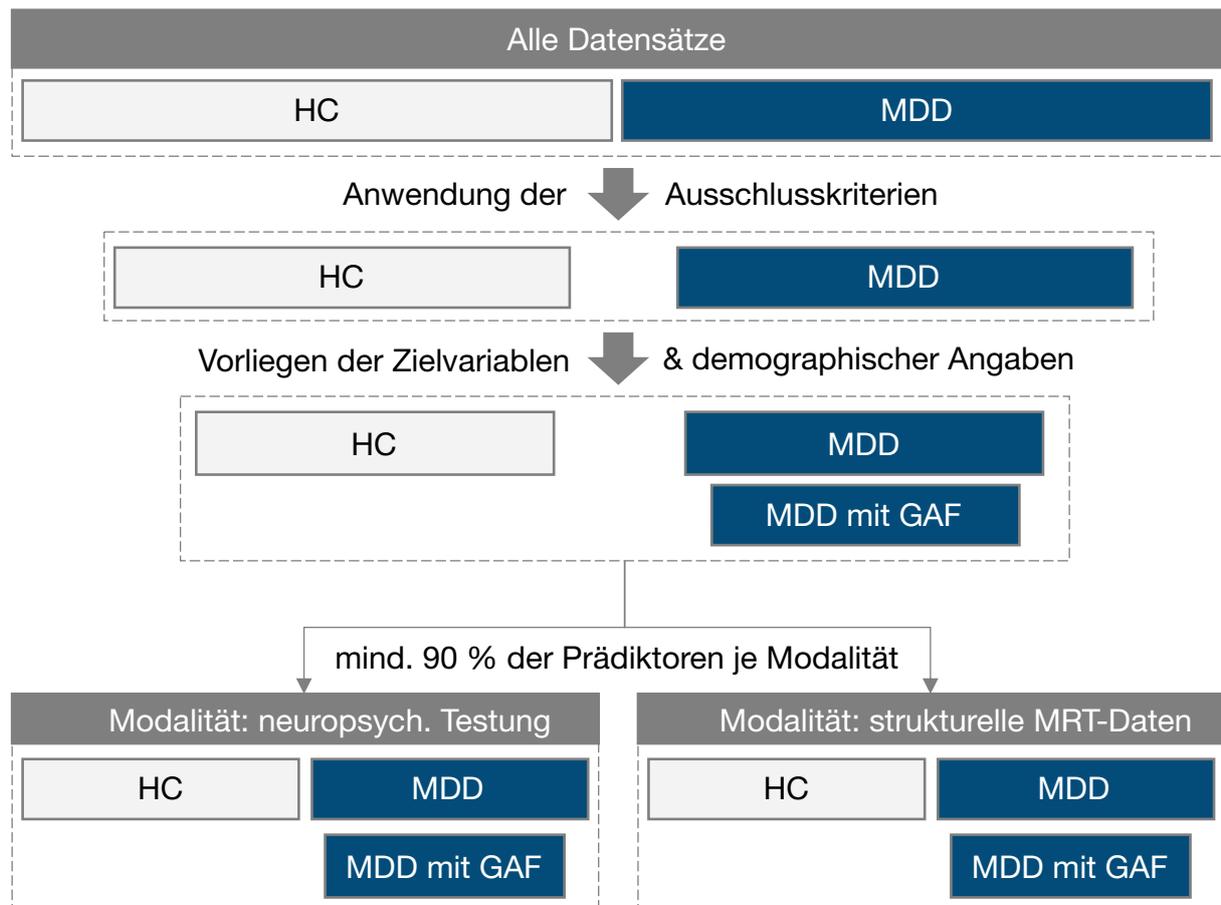


Abbildung 2-1: Prozess der Stichprobengenerierung

Anmerkungen: Dargestellt sind die einzelnen Schritte der Datenfilterung, die in dieser Arbeit zur Generierung der finalen Stichproben durchgeführt wurden. HC = gesunde Kontrollen (engl.: healthy controls, HC), MDD = Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD), GAF = Global Assessment of Functioning, MDD mit GAF = Probanden mit der Diagnose einer MDD, bei denen der GAF-Wert vollständig vorlag

Grundsätzlich wurden nur Probanden eingeschlossen, bei denen die entsprechende neuropsychologische Testung bzw. die strukturelle MRT Messung mitsamt automatisierter Segmentierung fehlerfrei verlief (letzteres war keine Voraussetzung für den Einschluss in die Stichproben der Modalität: neuropsychologische Testung), was ebenfalls über eine

entsprechende Filterung sichergestellt wurde. Im Anschluss wurden alle Probanden ausgeschlossen, bei denen weniger als 90 % der Prädiktoren vorlagen. Die daraus resultierenden Stichproben sind im Detail in den Kapiteln 2.2.1 und 2.2.2 beschrieben.

Matching der Kontrollgruppe mit der Patientengruppe

Auf eine Parallelisierung (engl.: matching) der gesunden Kontrollgruppe mit der Gruppe der MDD-Probanden anhand einschlägiger Störgrößen (Kovariaten) wie z. B. Geschlecht wurde im Hinblick auf das Ziel der Erstellung generalisierbarer Modelle explizit verzichtet. Stattdessen wurden basierend auf wichtigen Störgrößen (Geschlecht, Alter, Bildungsgrad, Studienzentrum) jeweils Subgruppen gebildet und die Analysen – neben der Gesamtgruppenanalyse – für jede dieser Subgruppen separat durchgeführt.

Die Auswahl von gesunden Kontrollprobanden, deren Merkmale hinsichtlich verschiedener Störgrößen ähnlich sind wie die der MDD-Probanden, wäre theoretisch möglich. Dazu könnte bspw. der Propensity Score berechnet und eine 1:1-Zuordnung von Kontrollen und MDD-Probanden anhand von ähnlichen Scores durchgeführt werden (bei Probanden mit identischen Propensity Scores ist die Verteilung der betrachteten Störvariablen ähnlich) (115).

Im Kontext dieser Arbeit hätte das allerdings keinen Sinn gemacht, da im Rahmen der hier angewendeten verschachtelten Kreuzvalidierung (siehe Kapitel 2.3.2) verschiedene Teilstichproben aus der Gesamtstichprobe zum Training und zur Hyperparameteroptimierung bzw. Testung des Modells gezogen wurden. Durch die zufällige Ziehung dieser Stichproben würde eine zuvor durchgeführte Parallelisierung der Gruppen nicht bestehen bleiben. Aus diesem Grund wurde auch auf Signifikanztests hinsichtlich der Verteilung einzelner Störgrößen verzichtet.

Statistische Auswertung der psychometrischen Daten

Um die Verteilung der psychometrischen Daten in den beiden Gruppen (HC vs. MDD) statistisch bewerten zu können, wurden entsprechende Signifikanztests durchgeführt. Auch für diese Analyse galt, dass die Gruppen durch die Kreuzvalidierung nicht in ihrer Ursprungsform erhalten blieben.

Dennoch konnte anhand der Tests belegt werden, dass die Ausgangsdaten entsprechende Gruppenunterschiede aufwiesen, zumal hier keine entsprechenden Subgruppen gebildet wurden. Als psychometrische Charakteristika wurden im Rahmen des Gruppenvergleiches

der Konsistenz halber dieselben betrachtet, die im weiteren Verlauf dieser Arbeit auch im Rahmen von Regressionsanalysen prädiziert wurden (HAMD, BDI, GAF, SF-36 Subskala: körperliche Funktionsfähigkeit) (s. Kapitel 2.4).

Da die Variablen entsprechend des Shapiro-Wilk-Tests auf dem 5 % Signifikanzniveau nicht normalverteilt waren, wurde der nicht-parametrische Wilcoxon-Mann-Whitney-Test (engl.: Mann-Whitney-Wilcoxon, MWW) angewendet, wobei sich alle Variablen signifikant zwischen den Gruppen unterschieden (5 % Signifikanzniveau).

Aufgrund der vergleichsweise großen Stichprobengröße war die Relevanz des Shapiro-Wilk-Tests allerdings kritisch zu betrachten (selbst kleine Abweichungen von der Normalverteilung, die u. U. keine Auswirkungen auf die Aussagekraft parametrischer Tests haben, werden bei großen Stichproben als signifikant eingestuft) (116). Um diesem Umstand Rechnung zu tragen und im Hinblick auf die statistische Power des MWW (117), wurde ebenfalls ein Welch's t-test durchgeführt, nachdem mit dem Levene-Test (5 % Signifikanzniveau) eine unterschiedliche Varianz der Variablen in beiden Gruppen nachgewiesen wurde. Erwartungsgemäß unterschieden sich die Verteilungen aller betrachteten Variablen zwischen den Gruppen auch mit diesem Test signifikant (5 % Signifikanzniveau).

Zur Bewertung der Signifikanz der Schiefe von Verteilungen kam im Rahmen der Diskussion (s. Kapitel 4) neben der grafischen Einordnung anhand von Histogrammen auch ein entsprechender Hypothesentest zum Vergleich der Schiefe mit einer entsprechenden Normalverteilung (Test of Skewness nach D'Agostino et al.) (118) zum Einsatz. Für die statistischen Tests wurde die Python Bibliothek SciPy (119) genutzt.

2.2.1 Stichprobenübersicht (Modalität: neuropsychologische Testung)

In Tabelle 2-3 sind die Stichprobencharakteristika der Kontrollgruppe und der MDD-Gruppe für die Modalität der neuropsychologischen Testung dargestellt. Das Vorliegen des GAF-Wertes innerhalb der MDD-Gruppe wurde hier nicht vorausgesetzt.

Tabelle 2-3: Stichprobentabelle 1, Modalität: neuropsychologische Testung, MDD- und HC-Gruppe

Stichprobe	MDD ^a		HC ^a		MWW
	n / M (SD)		n / M (SD)		p
Stichprobengröße	736		914		
Geschlecht					
männlich	253		329		
weiblich	483		585		
Alter	36.42	(13.00)	34.17	(12.96)	
Bildungsgrad					
Lehre, Meister, Abitur, Bachelor, Master	504		746		
Schulabschluss ^b	230		167		
kein Schulabschluss	2		1		
Bildungsjahre	13.27	(2.70)	14	(2.57)	
Psychometrische Daten					
GAF unbekannt	206		255		
GAF	65.33	(13.56)	91.11	(7.68)	< 0.001
SF-36: körperliche Funktionsfähigkeit	85.93	(17.91)	95.79	(9.85)	< 0.001
HAMD	9.14	(7.09)	1.45	(2.18)	< 0.001
BDI	17.14	(10.91)	4.11	(4.19)	< 0.001
Studienzentrum					
Marburg	323		545		
Münster	413		369		

MWW: Mann-Whitney-Wilcoxon Test, M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), p: p-Wert des MWW, GAF = Global Assessment of Functioning, SF-36 = Short Form 36-Fragebogen, HAMD = Hamilton Depressions-Skala, BDI = Beck-Depressions-Inventar, ^aProbanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD) bzw. gesunde Kontrollen (engl.: healthy controls, HC), ^bVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife

Tabelle 2-4 enthält zudem detaillierte Stichprobencharakteristika (u. a. hinsichtlich des Krankheitsverlaufes, der Medikation usw.) der MDD-Stichproben (separat für die Gruppen mit vollständigem bzw. nicht vollständigem GAF-Wert). Die Stichprobe, bei der der GAF-Wert nicht vollständig vorlag, stimmt mit der MDD-Gruppe in Tabelle 2-3 überein.

Tabelle 2-4: Stichprobentabelle 2, Modalität: neuropsychologische Testung, MDD-Gruppe

Stichprobe	MDD ^a (GAF vollständig)		MDD ^a (GAF unvollst.)	
	n / M (SD)		n / M (SD)	
Stichprobengröße	530		736	
Geschlecht				
männlich	181		253	
weiblich	349		483	
Alter	35.74 (12.75)		36.42 (13.00)	
Bildungsgrad				
Lehre, Meister, Abitur, Bachelor, Master	363		504	
Volks-, Haupt-, Realschulabschl., Fachabi	165		230	
kein Schulabschluss	2		2	
Bildungsjahre	13.2 (2.66)		13.27 (2.70)	
Psychometrische Daten				
GAF unbekannt	0		206	
GAF	65.33 (13.56)		65.33 (13.56)	
SF-36: körperliche Funktionsfähigkeit	86.50 (17.41)		85.93 (17.91)	
HAMD	8.72 (6.75)		9.14 (7.09)	
BDI	16.81 (10.60)		17.14 (10.91)	
zeitlicher Krankheitsverlauf				
einzelne Episode	185		262	
rezidivierend	345		474	
Anzahl depressiver Episoden	3.96 (6.26)		3.92 (6.38)	
(teil-)stationäre Behandlung				
mind. einmal in der Krankheitsgeschichte	348		492	
noch nie in der Krankheitsgeschichte	182		244	
derzeitig	151		218	
derzeitig nicht	379		518	
Remissionsstatus				
akut	220		312	
partielle Remission	138		190	
volle Remission	172		234	
Antidepressiva				
Medikation unbekannt	275		275	
mind. eins	149		275	
keins	106		186	
psych. Komorbidität				
unbekannt	0		201	
mind. eine	207		208	
keine	323		327	
Studienzentrum				
Marburg	240		323	
Münster	290		413	

M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), GAF = Global Assessment of Functioning, SF-36 = Short Form 36-Fragebogen, HAMD = Hamilton Depressions-Skala, BDI = Beck-Depressions-Inventar,

^aProbanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD)

2.2.2 Stichprobenübersicht (Modalität: strukturelle MRT-Daten)

Nach demselben Schema wie in Kapitel 2.2.1 sind in Tabelle 2-5 zunächst die Stichprobencharakteristika der gesamten MDD-Gruppe (das Vorliegen des GAF-Wertes wurde nicht vorausgesetzt) und der gesunden Kontrollgruppe für die Modalität der strukturellen MRT-Daten dargestellt.

Tabelle 2-5: Stichprobentabelle 3, Modalität: strukturelle MRT-Daten, MDD- und HC-Gruppe

Stichprobe	MDD ^a		HC ^a		MWW
	n / M (SD)		n / M (SD)		p
Stichprobengröße	432		615		
Geschlecht					
männlich	153		222		
weiblich	279		393		
Alter	37.13	(13.25)	32.69	(12.51)	
Bildungsgrad					
Lehre, Meister, Abitur, Bachelor, Master	291		515		
Schulabschluss ^b	140		100		
kein Schulabschluss	1		0		
Bildungsjahre	13.27	(2.75)	13.96	(2.42)	
Psychometrische Daten					
GAF unbekannt	194		235		
GAF	65.31	(14.23)	92.28	(6.44)	< 0.001
SF-36: körperliche Funktionsfähigkeit	85.10	(18.29)	96.85	(7.37)	< 0.001
HAMD	9.10	(7.27)	1.29	(1.97)	< 0.001
BDI	17.72	(11.09)	3.89	(4.13)	< 0.001
Studienzentrum					
Marburg	271		402		
Münster	161		213		

MWW: Mann-Whitney-Wilcoxon Test, M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), p: p-Wert des MWW, GAF = Global Assessment of Functioning, SF-36 = Short Form 36-Fragebogen, HAMD = Hamilton Depressions-Skala, BDI = Beck-Depressions-Inventar, ^aProbanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD) bzw. gesunde Kontrollen (engl.: healthy controls, HC), ^bVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife

In Tabelle 2-6 finden sich die detaillierten Stichprobencharakteristika der MDD-Patienten für die Modalität der strukturellen MRT-Daten. Auch hier findet sich eine separate Auflistung der Gruppen mit und ohne vollständige GAF-Werte.

Tabelle 2-6: Stichprobentabelle 4, Modalität: strukturelle MRT-Daten, MDD-Gruppe

Stichprobe	MDD ^a (GAF vollständig)		MDD ^a (GAF unvollst.)	
	n / M (SD)		n / M (SD)	
Stichprobengröße	238		432	
Geschlecht				
männlich	90		153	
weiblich	148		279	
Alter	36.5 (13.34)		37.13 (13.25)	
Bildungsgrad				
Lehre, Meister, Abitur, Bachelor, Master	156		291	
Volks-, Haupt-, Realschulabschl., Fachabi	81		140	
kein Schulabschluss	1		1	
Bildungsjahre	13.01 (2.69)		13.27 (2.75)	
Psychometrische Daten				
GAF unbekannt	0		194	
GAF	65.31 (14.23)		65.31 (14.23)	
SF-36: körperliche Funktionsfähigkeit	85.84 (17.28)		85.1 (18.29)	
HAMD	8.13 (6.58)		9.10 (7.27)	
BDI	17.46 (10.72)		17.72 (11.09)	
zeitlicher Krankheitsverlauf				
einzelne Episode	73		144	
rezidivierend	165		288	
Anzahl depressiver Episoden	4.48 (5.54)		4.23 (6.17)	
(teil-)stationäre Behandlung				
mind. einmal in der Krankheitsgeschichte	168		302	
noch nie in der Krankheitsgeschichte	70		130	
derzeitig	86		149	
derzeitig nicht	152		283	
Remissionsstatus				
akut	109		198	
partielle Remission	64		109	
volle Remission	65		125	
Antidepressiva				
Medikation unbekannt	0		0	
mind. eins	142		261	
keins	96		171	
psych. Komorbidität				
unbekannt	0		189	
mind. eine	84		85	
keine	154		158	
Studienzentrum				
Marburg	191		271	
Münster	47		161	

M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), GAF = Global Assessment of Functioning, SF-36 = Short Form 36-Fragebogen, HAMD = Hamilton Depressions-Skala, BDI = Beck-Depressions-Inventar,

^aProbanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD)

2.3 Implementierung der Methode des maschinellen Lernens

Sämtliche Analysen wurden mit der Software PyCharm (2020.1 Professional Edition, JetBrains s.r.o.) realisiert. Als Interpreter kam Python in der Version 3.7.6 zum Einsatz.

Vor der Implementierung der eigentlichen Machine Learning-Modelle wurden die Daten entsprechend aufbereitet. Dieser Schritt umfasste vor allem die Stichprobengenerierung und die inferenzstatistischen Analysen (s. Kapitel 2.2). Hierfür wurden hauptsächlich die Python-Bibliotheken SciPy (119) sowie pandas (120) genutzt. Zur visuellen Darstellung der Ergebnisse kamen die Bibliotheken Matplotlib (121) und Seaborn (122) zum Einsatz. Die Entwicklung und Implementierung der Machine Learning-Modelle ist in den folgenden Unterkapiteln beschrieben.

2.3.1 Erläuterung der PHOTONAI API und der Pipeline-Architektur

Zur Erstellung der Machine Learning-Modelle innerhalb dieser Arbeit wurde PHOTONAI (Version: 1.1.0, Entwicklerversion) genutzt (64,123). Die Software wurde im Medical Machine Learning Lab am Institut für Translationale Psychiatrie der Universität Münster entwickelt (64). Es handelt sich um eine in Python 3 implementierte Toolbox zum Entwerfen und Optimieren von Machine Learning-Projekten (64).

Die Software automatisiert diverse repetitive Schritte der Modellentwicklung, wie z. B. die Unterteilung in Trainings- und Testsets und die Hyperparameteroptimierung, und ermöglicht so die Fokussierung auf wichtige Designentscheidungen (64). Die Software ist unter der „GNU General Public License v3.0“ lizenziert und frei verfügbar (<https://www.photon-ai.com/>) (64).

PHOTONAI greift auf bekannte Data Science und Machine Learning-Bibliotheken wie z. B. pandas (120) und NumPy (124) zurück (64). Im Bereich des überwachten maschinellen Lernens wird überwiegend scikit-learn (125) eingesetzt (64).

Das Design eines Machine Learning-Modells realisiert PHOTONAI anhand einer prozessorientierten Pipeline (ähnlich wie in scikit-learn (125)) (64). Die projektspezifische Pipeline kann über die Auswahl verschiedener Bausteine entworfen werden (64). Diese Bausteine entsprechen einzelnen Prozessschritten. Dabei handelt es sich bspw. um einzelne Vorverarbeitungsschritte oder Lernverfahren (64).

In Abbildung 2-2 ist beispielhaft ein simples Pipeline-Design unter Einsatz von zwei Transformatoren (Imputation und Skalierung) sowie eines einzelnen Lernverfahrens dargestellt (64). Basierend auf den Prädiktoren (X) und dem wahren Label (y_{wahr}) der Zielvariable errechnet das beispielhafte Modell eine Prädiktion ($y_{\text{prädiziert}}$).



Abbildung 2-2: simple Pipeline-Architektur in PHOTONAI (angelehnt an (64))

Anmerkungen: Dargestellt ist eine simple Machine Learning-Pipeline mit zwei Transformatoren (Imputation/Skalierung) und einer SVM, die basierend auf den Eingangsdaten bestimmte Zielvariablen prädiziert. SVM = Support Vector Machine, X = Prädiktoren, y_{wahr} = wahrer Wert der Zielvariable, $y_{\text{prädiziert}}$ = durch das Modell prädizierter Wert der Zielvariable

Datenlecks, wie sie in Kapitel 1.3.3 beschrieben sind, werden durch die Pipeline-Architektur von PHOTONAI vermieden (64). Die Vorteile dieser weniger fehleranfälligen, Pipeline-basierten Vereinfachung des Prozesses der Modell- und Hyperparameteroptimierung sind in der Literatur beschrieben (17).

2.3.2 Erläuterung der eingesetzten Kreuzvalidierung

Wie in Kapitel 1.3.1 beschrieben, ist es üblich, die Performance eines Machine Learning-Modells anhand eines Kreuzvalidierungsschemas zu evaluieren.

In der psychiatrischen Forschung ist die Leave-one-out-Kreuzvalidierung (LOOCV) populär (17,61). Sie ermöglicht das wiederholte Training des Modells an der gesamten Stichprobe abzüglich genau eines Datenpunktes, welcher dann zum Testen des jeweiligen Modells herangezogen wird (17,61). Der große Umfang des Trainingssets ($n-1$) ist hinsichtlich der kleinen verfügbaren Stichproben eine reizvolle Eigenschaft (17). In der Literatur wurde allerdings belegt, dass die LOOCV zu instabilen Einschätzungen der prädiktiven Genauigkeiten führen kann (61).

Stattdessen wird eine wiederholte zufällige Subgruppenbildung empfohlen, was bspw. durch die k -fache (engl.: k -fold) Kreuzvalidierung realisiert werden kann (17,126,127). Dabei sollen die generierten Subgruppen jeweils aus größeren Blocks mit korrelierten Samples bestehen anstatt aus einzelnen individuellen Datenpunkten, wie es bei der LOOCV der Fall ist (61). Die k -fache Kreuzvalidierung bietet zudem den Vorteil, dass sie wiederholt in einer

verschachtelten Struktur durchgeführt werden kann, sodass sich jeweils nicht überschneidende Gruppen für die einzelnen Schritte des Trainings, der Validierung (Hyperparameteroptimierung) und der Evaluation generiert werden können (17).

Führt man die Hyperparameteroptimierung und Evaluation bei einer nicht-verschachtelten Kreuzvalidierung stattdessen an demselben Testset durch, so kann das zu einer überoptimistischen Performancebewertung führen, da in diesem Fall Informationen aus dem Testset zur Hyperparameterwahl genutzt werden (Datenleck) (128).

Weitere empfohlene Kreuzvalidierungsmethoden wie z. B. die Leave-group-out Kreuzvalidierung, bei der Daten aus verschiedenen Quellen separat betrachtet werden (17), sind für diese Arbeit nicht primär relevant, da die Daten an den beiden Studienzentren nach demselben Protokoll erhoben wurden und jeweils eine ähnliche Population repräsentieren.

Verschachtelte k-fache Kreuzvalidierung

Der folgende Abschnitt ist angelehnt an Cearns et al. (17): Im Rahmen der verschachtelten Kreuzvalidierung wird die k-fache Kreuzvalidierung wiederholt durchgeführt. Das ist notwendig, um wiederholt verschiedene, sich nicht überschneidende Samples für den Trainingsprozess, die Hyperparameteroptimierung und letztendlich die Evaluation des Modells zu generieren.

Zunächst findet die äußere k-fache Kreuzvalidierung statt (s. Abbildung 2-3). Dazu werden sämtliche zur Verfügung stehenden Samples in k Gruppen (auch Anzahl an Faltungen genannt) unterteilt. In k Durchläufen der Modellbewertung wird dann jeweils eine dieser Gruppen (erst Gruppe 1, dann Gruppe 2 usw.) als Testset bereitgehalten, wohingegen die anderen k-1 Gruppen jeweils an die innere k-fache Kreuzvalidierung übergeben werden. Dieser Schritt entspricht der wiederholten Aufteilung der Daten in ein Trainings- und ein Testset.

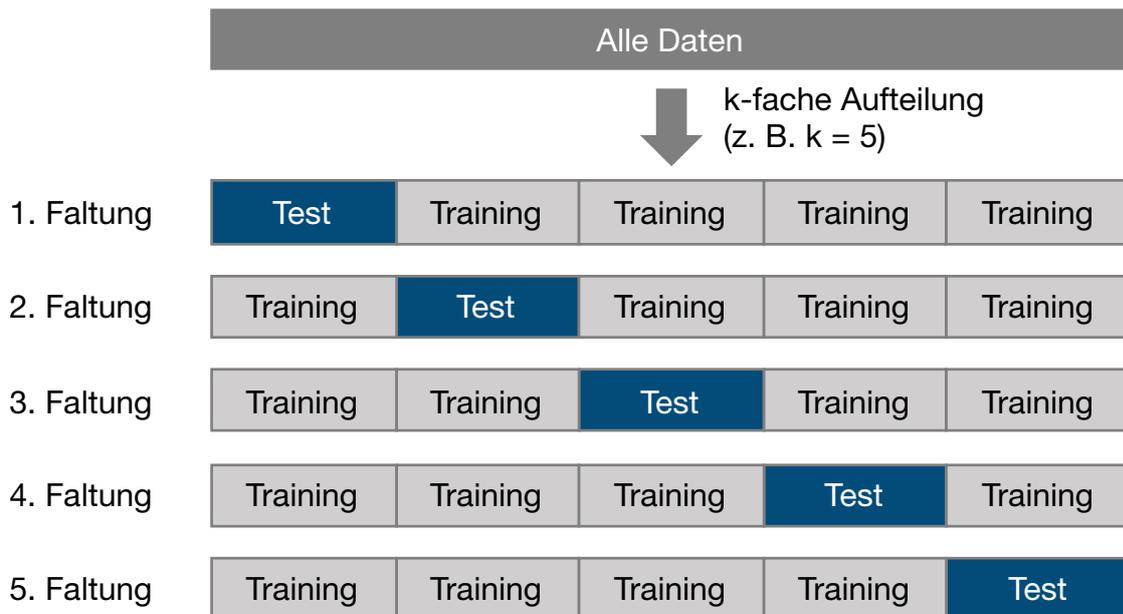


Abbildung 2-3: Prinzip der *k*-fachen äußeren Kreuzvalidierung (in Anlehnung an (17,63))

Anmerkungen: Im Rahmen der *k*-fachen Kreuzvalidierung werden ausgehend von den vorliegenden Daten verschiedene Test- und Trainingsdatensätze generiert. Als Anzahl der Faltungen wurde beispielhaft $k = 5$ gewählt. Test = eine Gruppe an Samples, die für die entsprechende Faltung das Testset bildet, Training = eine Gruppe an Samples, die dem Trainingsset zugeordnet wird, wobei alle $k-1$ Trainingsgruppen zusammen das Trainingsset für die entsprechende Faltung bilden

Innerhalb der äußeren *k*-fachen Kreuzvalidierung ist die innere *k*-fache Kreuzvalidierung eingebettet. Sie übernimmt jeweils die $k-1$ Gruppen eines jeden Durchlaufs der äußeren Kreuzvalidierung und teilt diese Daten erneut in *k* Gruppen auf. Die Anzahl der Gruppen (*k*) der inneren Kreuzvalidierung kann sich von der der äußeren Kreuzvalidierung unterscheiden.

In den *k* Durchläufen der inneren Kreuzvalidierung dient jeweils eine der Gruppen (erst die innere Gruppe 1, dann die innere Gruppe 2 usw.) als Validierungsset zur Wahl der Hyperparameter, wobei die anderen $k-1$ Gruppen jeweils als Trainingsset eingesetzt werden. Die Daten, die aus der äußeren Kreuzvalidierung übergeben wurden, werden also wiederholt in ein Trainings- und ein Validierungsset eingeteilt. Nach Auswahl der optimalen Parameter wird das Modell anhand des Testsets der äußeren Kreuzvalidierung evaluiert (s. Abbildung 2-4).

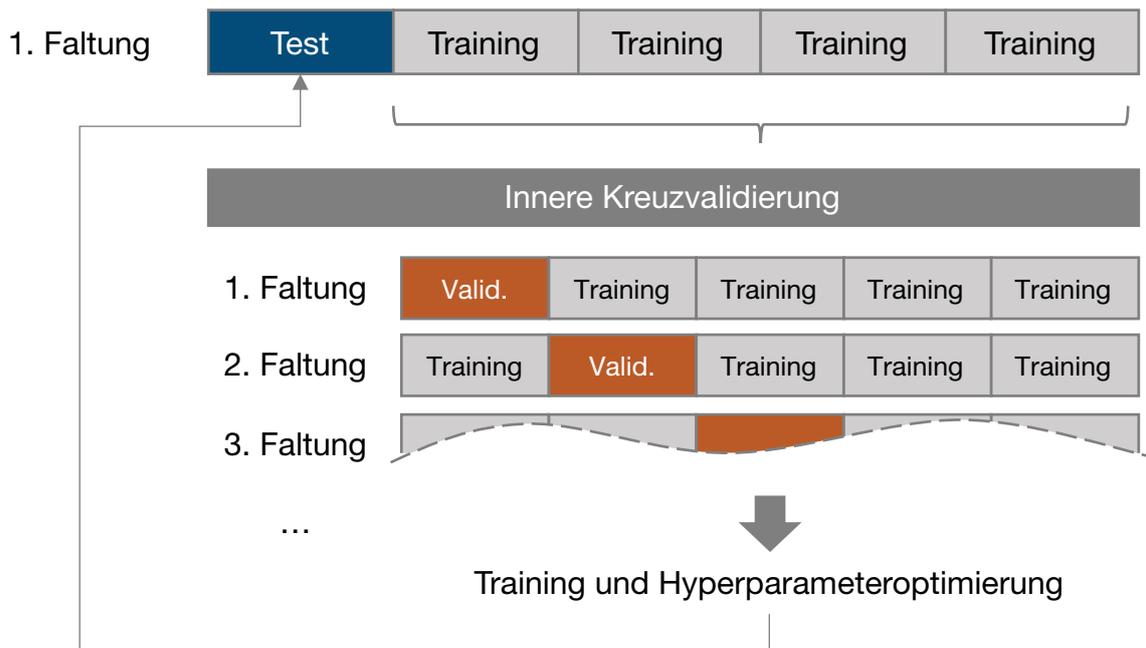


Abbildung 2-4: Prinzip der k -fachen inneren Kreuzvalidierung (in Anlehnung an (17,63))

Anmerkungen: Als Anzahl der Faltungen wurde beispielhaft $k = 5$ gewählt. Die innere Kreuzvalidierung ist hier für die erste Faltung der äußeren Kreuzvalidierung dargestellt. Ausgehend von diesen Daten werden verschiedene Trainings- und Validierungsdatensätze gebildet. Test = eine Gruppe an Samples, die für die entsprechende Faltung das Testset bildet, Training = eine Gruppe an Samples, die dem Trainingsset zugeordnet wird, wobei alle $k-1$ Trainingsgruppen zusammen das Trainingsset für die entsprechende Faltung bilden, Valid. = eine Gruppe an Samples, die für die entsprechende Faltung das Validierungsset bilden

Bei der Wahl der jeweiligen Anzahl der Gruppen (k) ist zu beachten, dass hohe k einerseits zu großen Überschneidungen zwischen den Trainingsdatensätzen und andererseits zu kleinen, u. U. nicht mehr für die vorliegenden Daten repräsentativen, Testdatensätzen führen können, was zu einer großen Varianz der Ergebnisse der einzelnen Faltungen und zu einer instabilen Schätzung der prädiktiven Genauigkeit führen kann (61,126,127). Niedrige k -Werte wiederum können zu kleinen, u. U. instabilen, Trainingsdatensätzen führen, was die Modellgenauigkeit ebenfalls beeinträchtigen kann (126).

In der Literatur konnte bezüglich dieses Zusammenhangs zwischen Bias und Varianz (engl.: bias-variance-tradeoff, s. Kapitel 1.3.1) experimentell belegt werden, dass eine 10-fache Kreuzvalidierung einen guten Kompromiss darstellt (126).

Ob eine wiederholte Anwendung der Kreuzvalidierung sinnvoll ist, um die Varianz der Ergebnisse weiter zu reduzieren, wird in der Literatur diskutiert (17,126,129) und hängt von der zur Verfügung stehenden Rechenleistung, der Varianz der Daten, der Korrelationsstärke zwischen den Prädiktoren und Zielvariablen sowie der Stabilität der Performance-Evaluation

innerhalb des Kreuzvalidierungsschemas ab (17). Es bleibt zu beachten, dass auch die verschachtelte Kreuzvalidierung zu einer systematischen Überschätzung der Ergebnisse führen kann (17).

Entsprechend der Literaturrecherche wurde die äußere Kreuzvalidierung in dieser Arbeit jeweils mit $k = 10$ Faltungen durchgeführt (126). Um die Gruppengröße nicht zu stark zu verringern, wurde für die innere Kreuzvalidierung $k = 5$ gewählt (17).

Stratifizierte Kreuzvalidierung

In dieser Arbeit wurden die Datenpunkte vor der Durchführung der beiden Teile der verschachtelten Kreuzvalidierung jeweils zufällig gemischt. Bei Klassifikationsanalysen wurde zudem eine stratifizierte Version der Kreuzvalidierung angewendet (scikit-learn (125,130): StratifiedKFold), um der Tatsache Rechnung zu tragen, dass nicht in allen betrachteten Analysen die Anzahl der Datenpunkte pro Klasse gleichmäßig verteilt war und eine rein zufällige Auswahl der Trainings- und Testsets somit für eine deutliche Unterrepräsentation einer der Klassen hätte führen können. Die Anwendung stratifizierter Kreuzvalidierung wird in der Literatur empfohlen (126) und löst dieses Problem, indem in jeder Gruppe jeweils das ursprüngliche Verhältnis der beiden Klassen zueinander näherungsweise erhalten bleibt (125,130).

Um auch für Regressionsanalysen sicherzustellen, dass die Verteilung der Zielvariable in den jeweiligen Faltungen näherungsweise der Verteilung in der gesamten Stichprobe entspricht, wurde die in PHOTONAI implementierte stratifizierte k -fache Kreuzvalidierung für kontinuierliche Zielvariablen angewendet (PHOTONAI (64): StratifiedKFoldRegression). Dabei werden die Samples zunächst entsprechend der Zielvariable der Größe nach sortiert und dann schrittweise genau eines aus jeweils k aufeinander folgenden Samples zufällig einer der k Faltungen zugeordnet (64).

Voraussetzung für die Durchführung einer Analyse war das Vorliegen von mind. 10 Probanden pro Klasse, womit das Vorliegen von mind. einem Probanden pro Klasse in jedem Testset sichergestellt wurde. War das nicht der Fall, wurde die entsprechende Analyse nicht durchgeführt.

PHOTONAI ermöglicht den automatisierten Einsatz einer verschachtelten Kreuzvalidierung (64). Das Training sowie die Hyperparameteroptimierung findet anhand der festgelegten

Referenz-Metrik innerhalb der inneren Kreuzvalidierung statt. Das so jeweils optimierte Modell wird im Rahmen der äußeren Kreuzvalidierung evaluiert und im finalen Schritt die Ergebnisse der äußeren Faltungen arithmetisch gemittelt (64).

2.3.3 Transformatoren, Lernverfahren, Hyperparameter und Metriken

Im Folgenden wird die Auswahl der Pipelineelemente konkretisiert. Dabei werden die gewählten Transformatoren, Lernverfahren und Hyperparameter beschrieben und zudem dargestellt, wie der Prozess der Modellentwicklung innerhalb des in Kapitel 2.3.2 vorgestellten Kreuzvalidierungsschemas gestaltet wurde.

Transformatoren

Im hier angewendeten Machine Learning-Modell wurden folgende Transformationen durchgeführt:

Für Klassifikationsanalysen wurde zunächst ein Encoder (scikit-learn (125,131): LabelEncoder) angewandt, der die zu prädizierenden Klassen mit 0 bzw. 1 kodiert. Bei sämtlichen Klassifikationsanalysen in dieser Arbeit wurden immer genau zwei verschiedene Klassen (binär) voneinander unterschieden. Durch die Anwendung dieses Encoders wurde sichergestellt, dass auch Klassen, die in den Originaldaten anders benannt waren (bspw. die Diagnose, die als Buchstabenfolge in den Daten spezifiziert war: MDD oder HC) vom Modell identifiziert und klassifiziert werden konnten. Die eingesetzten Lernverfahren setzten eine binäre Kodierung der Klassenzugehörigkeit voraus.

Zudem wurden fehlende Werte im Trainingsset ersetzt. Dabei wurde eine technisch einfache Funktion (scikit-learn (125,132): SimpleImputer) genutzt, welche die fehlenden Daten zunächst identifiziert und dann den prädiktorspezifischen Mittelwert (für jeden Prädiktor spaltenweise ermittelt) einsetzt. Sämtliche Prädiktoren lagen als Zahlenwerte vor. Wie in Kapitel 2.2 beschrieben, wurde in der Wahl der Stichprobe vorausgesetzt, dass mind. 90 % der Prädiktoren jeweils vorhanden waren, d.h. es wurden maximal < 10 % der Prädiktoren pro Proband ersetzt.

Im nächsten Schritt wurden die Prädiktoren im Trainingsset skaliert (scikit-learn (125,133): StandardScaler), sodass sie den Mittelwert 0 und die Standardabweichung 1 aufwiesen. Dazu wurde von jedem Datenpunkt der prädiktorspezifische Mittelwert (spaltenweise für jeden

Prädiktor separat berechnet) abgezogen und dieser Wert anschließend durch die prädiktorspezifische Standardabweichung geteilt (s. Formel 2-1) (125,133).

$$x_{\text{skaliert}} = \frac{x - \mu}{\sigma}$$

Formel 2-1: Skalierung der Prädiktoren (angelehnt an (125,133))

x: unskalierter Datenpunkt (repräsentiert einen der m Prädiktoren für eines der n Samples)

μ: Mittelwert (für den m-ten Prädiktor anhand aller n Samples im Trainingsset spaltenweise berechnet)

σ: Standardabweichung (für den m-ten Prädiktor anhand aller n Samples im Trainingsset spaltenweise berechnet)

x_{skaliert}: entsprechend der obigen Formel skaliertes Datenpunkt

Die Mittelwerte und Standardabweichungen, die so für jeden einzelnen Prädiktor des jeweiligen Trainingssets ermittelt wurden, dienten unverändert als Grundlage für die entsprechenden Transformationen (Imputation bzw. Skalierung) am Testset, um Datenlecks zwischen Trainings- und Testdaten zu vermeiden (125,133). Vor allem im Hinblick auf die Funktionsweise der eingesetzten Support Vector Machine (abhängig von der Skala) ist eine Skalierung der Trainings- und Testdaten notwendig (125,134).

Lernverfahren

Wie in Kapitel 1.3.4 beschrieben, belegt die Theorie, dass kein einzelner Algorithmus für alle denkbaren Probleme die optimale Performance haben kann (25,87). Aus diesem Grund und da zuvor kein Wissen über die Zusammenhänge in den Daten vorhanden war, kamen in dieser Arbeit verschiedene Arten von Lernverfahren zum Einsatz.

Die Wahl fiel dabei auf die Support Vector Machine (SVM) und den Random Forest. Die Begründung der Lernverfahren-Auswahl folgt im weiteren Verlauf dieses Abschnittes.

Die PHOTONAI-Software ermöglicht die Anwendung und den Vergleich verschiedener Lernverfahren innerhalb einer einzelnen Pipeline. Über den Einsatz eines „Switch“-Elementes (Schalter) wird automatisch geprüft, welches der zu testenden Lernverfahren die beste Performance bietet (s. Abbildung 2-5 am Beispiel einer SVM und eines Random Forest). Die übrigen Elemente der Pipeline (z. B. Transformatoren) müssen nur einmal spezifiziert werden (64).

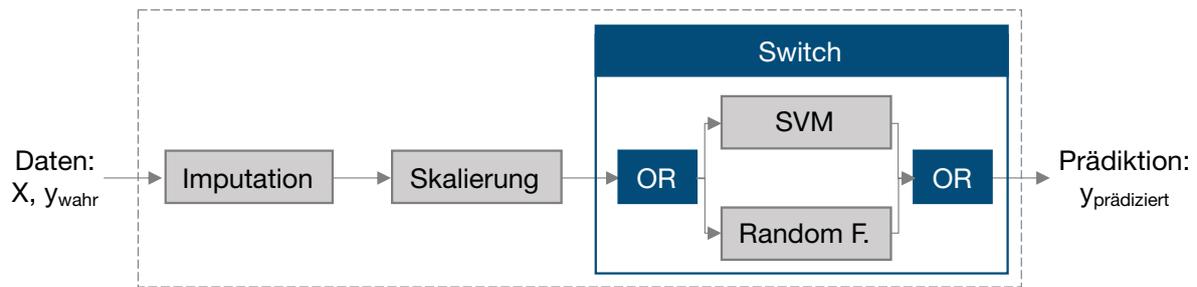


Abbildung 2-5: Pipelinearchitektur unter Einsatz eines „Switch“ (angelehnt an (64))

Anmerkungen: Dargestellt ist eine Machine Learning-Pipeline mit zwei Transformatoren (Imputation/Skalierung) und zwei verschiedenen Lernverfahren (SVM/Random Forest), die basierend auf den Eingangsdaten bestimmte Zielvariablen prädiziert. Der „Switch“ wird zur Evaluation der beiden Lernverfahren eingesetzt. SVM = Support Vector Machine, Random F. = Random Forest, X = Prädiktoren, y_{wahr} = wahrer Wert der Zielvariable, $y_{\text{prädiziert}}$ = durch das Modell prädizierter Wert der Zielvariable

Die SVM (scikit-learn (125,135–138): SVC für Klassifikationsanalysen bzw. SVR für Regressionsanalysen) wurde aus verschiedenen Gründen gewählt. Einerseits kam dieses Lernverfahren innerhalb der psychiatrischen Forschung schon vielfach zum Einsatz und ergab vielversprechende Ergebnisse (24,31,71,79). Andererseits wird in der Literatur beschrieben, dass die SVM auch bei der Nutzung von kleinen Datensets (wie es hier der Fall war) valide Ergebnisse liefern kann (139). Die SVM basiert auf einer verlässlichen theoretischen Basis und kann auch für hochdimensionale Daten eingesetzt werden (24). Neben der Anwendung für Klassifikations- bzw. Regressionsanalysen basierend auf einer linearen Funktion (lineare Entscheidungsgrenzen) bietet die SVM durch die Anwendung von Kernelfunktionen zudem die Möglichkeit, nicht-lineare Entscheidungsgrenzen zu identifizieren (24).

Auch der Random Forest ist im Bereich der psychiatrischen Literatur in der Vergangenheit eingesetzt worden (77,86). Dabei werden Ensemble-Methoden angewendet (140). Durch das Erstellen einer Vielzahl von einzelnen Entscheidungsbäumen (engl. Bezeichnung: ensemble learning), basierend auf verschiedenen Subsamples und der anschließenden Mittelung der Ergebnisse, wird die Prädiktionsgenauigkeit verbessert und zudem eine bessere Generalisierbarkeit des Modells sichergestellt (140–142).

Je nach Spezifikation des eingesetzten Random Forest können wiederholt zufällig gezogene Subsamples (engl. Bezeichnung: bootstrapping) als Basis für die einzelnen Entscheidungsbäume genutzt werden (140,142). Auch die Anzahl an miteinbezogenen Prädiktoren für die verschiedenen Entscheidungsbäume kann zufällig variiert werden (engl. Bezeichnung: random subspace method) (140,141,143). Beide Randomisierungen führen zu einer Varianzreduktion innerhalb des Ensembles (140,141). Das Verfahren der Kombination

von Entscheidungsbäumen basierend auf verschiedenen (durch Bootstrapping generierten) Subsamples wird im Englischen auch *bootstrap aggregating* (Abkürzung: *bagging*) genannt (142). Aufgrund der im Gegensatz zur SVM anderen technischen Struktur wurde dieses Ensemble-Lernverfahren gewählt. Auch hier kamen zwei verschiedene Formen für Klassifikations- und Regressionsanalysen zum Einsatz (*scikit-learn* (125,140,144–146): *RandomForestClassifier*, *RandomForestRegressor*).

Hyperparameter

Da die optimalen Hyperparameter mangels Vorwissens über die Prädiktoren vor Beginn der Analyse nicht bekannt waren, wurden verschiedene Hyperparameterkombinationen getestet. Zur Optimierung der Hyperparameter (Auswahl der optimalen Hyperparameter aus dem Hyperparameterraum) wurde eine Rastersuche (engl.: *grid search*) durchgeführt. Das bedeutet, das Machine Learning-Modell wurde separat mit allen der möglichen Hyperparameterkombinationen innerhalb der inneren Kreuzvalidierung trainiert (anhand des Trainingssets) sowie getestet (anhand des Validierungssets) und das Modell (samt Hyperparameter) mit der jeweils besten Performance gewählt. Eine solche Rastersuche ist in PHOTONAI implementiert (64).

Bei der Auswahl der zu testenden Hyperparameter wurde ein Kompromiss zwischen einem möglichst großen Hyperparameterraum und einem im Hinblick auf die zur Verfügung stehende Rechenleistung adäquaten Rechenaufwand angestrebt. Die hierbei getesteten Hyperparameter sind im Verlauf dieses Kapitels beschrieben.

Für die Support Vector Machine wurde der Hyperparameter $C \in \{0.001; 0.1; 1; 3; 10\}$ (Regularisierungsparameter: bestimmt die Höhe der Bestrafung des Algorithmus für falsche Prädiktionen (136)) gewählt. Durch diese Wahl wurde ein breiter Bereich an Regularisierungsparametern abgedeckt und durch die Schrittweite der Rechenaufwand zugleich auf ein sinnvolles Maß beschränkt.

Um sowohl lineare als auch nicht-lineare Entscheidungsgrenzen abbilden zu können (147), wurden zwei verschiedene Kernel $\in \{\text{linear}; \text{rbf}\}$ (125,137,138) gewählt. Dabei müssen bei der Anwendung des rbf-Kernels (engl.: *radial basis function*) relativ wenig Hyperparameter deklariert werden (128), was die Komplexität verringert und die Wahl dieses nicht-linearen Kernels motiviert hat. Die maximale Anzahl an Iterationen wurde im Hinblick auf die Rechenzeit auf 10^7 festgelegt.

Der Koeffizient gamma für den rbf-Kernel (definiert das Einflussgebiet eines einzelnen Trainings-Samples (125,137,138)) wurde entsprechend der Anzahl an Prädiktoren sowie der Varianz des Trainingssets skaliert ($\text{gamma} = \text{scale} (125,137,138)$), um die Wahl jeweils sinnvoll an die vorliegenden Daten anzupassen. Dazu wurde entsprechend der scikit-learn-Implementation der Kehrwert aus dem Produkt der Anzahl an Prädiktoren und der Varianz der Trainingsdaten gebildet (s. Formel 2-2) (125,137,138).

$$\text{gamma}_{\text{„scale“}} = \frac{1}{n_{\text{features}} \times \sigma^2}$$

Formel 2-2: Berechnung des gamma-Koeffizienten (= scale) (angelehnt an (125,137,138))

n_{features} : Anzahl der eingesetzten Prädiktoren

σ^2 : Varianz des Trainingssets

Für die Regressionsanalysen wurde Epsilon (definiert einen Toleranzbereich für die Abweichung der Prädiktionen von den wahren Werten, innerhalb dem der Algorithmus nicht für Abweichungen bestraft wird) auf 0.1 festgelegt ($\text{epsilon} = 0.1$), was der scikit-learn-StandardEinstellung entspricht (125,138).

Für die Entscheidungsbäume innerhalb eines Random Forest wurde jeweils die minimale Anzahl an Samples pro Blatt (engl.: minimum number of samples per leaf) $\in \{1; 10; 0.1\}$ gewählt (125,145,146). Dieser Parameter beeinflusst die Menge der Aufteilungen des Entscheidungsbaumes anhand der Vorgabe einer Mindestanzahl an Samples, die pro endständigem Blatt vorliegen müssen (125,145,146). Der Wert 0.1 ist dabei eine relative Angabe und steht für mind. 10 % der Samples (125,145,146).

Für die maximale Tiefe der Entscheidungsbäume (engl.: maximum depth) wurden Werte $\in \{2; 4; 10; \text{None}\}$ gewählt (125,145,146). „None“ bedeutet, dass die Tiefe in diesem Fall nicht limitiert wurde (die minimale Anzahl an Samples pro Blatt blieb als Kriterium allerdings weiter erhalten) (125,145,146).

Die Anzahl der Entscheidungsbäume (engl.: number of estimators) wurde im Hinblick auf die Anzahl der Prädiktoren und die zur Verfügung stehende Rechenleistung auf 100 festgelegt ($\text{number of estimators} = 100$) (125,145,146). Die Anzahl an Entscheidungsbäumen, die notwendig ist, um gute prädiktive Genauigkeiten zu erreichen, hängt dabei im Allgemeinen von der Anzahl der Prädiktoren ab (148).

Zur Bewertung der Qualität einer Aufteilung innerhalb eines Entscheidungsbaumes wurden die Kriterien (engl. Bezeichnung: criterion) verwendet, die standardmäßig in scikit-learn definiert sind (criterion = „mse“ (Regression) bzw. „gini“ (Klassifikation)) (125,145,146).

Abgesehen von den hier spezifizierten Hyperparametern wurden die Lernverfahren jeweils mit den in der scikit-learn-Bibliothek spezifizierten Standardeinstellungen angewendet (125,137,138,145,146).

Optimierungsmetriken

Im Rahmen des Optimierungsprozesses der Parameter und Hyperparameter eines Machine Learning-Modells ist es notwendig, zu definieren, welcher Fehler des Modells minimiert werden soll (64).

In den Regressionsanalysen wurde jeweils der mittlere absolute Fehler (engl.: mean absolute error, MAE) – also das arithmetische Mittel der Abweichung der prädizierten Werte von den jeweiligen wahren Werten – minimiert (125,149).

Für Klassifikationsanalysen wurde jeweils die korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC) optimiert. Die Metrik nimmt Werte zwischen 0 und 1 an und lässt auch bei einer unausgeglichenen Klassenverteilung eine Bewertung der Genauigkeit der Treffsicherheit pro Klasse (125,150–152) zu. Die Metrik berechnet sich im Fall von zwei Klassen durch das arithmetische Mittel von Sensitivität und Spezifität (153). Entsprechend der Literatur sind solche korrigierten Metriken für unausgeglichene (engl.: unbalanced) Klassenverteilungen zur Evaluierung bspw. gegenüber der absoluten Klassifikationsgenauigkeit (die höhere Werte annehmen und somit fehlinterpretiert werden könnte) zu bevorzugen (24). Im Rahmen der Regressionsanalysen wurde zudem ergänzend die Pearson-Korrelation berechnet.

Einordnung der Ergebnisse (Dummy)

Um die im Rahmen der Evaluation ermittelte Performance der Analysen bewerten zu können, wurde eine Vergleichsmetrik eingeführt. Der hier eingesetzte sog. Dummy (entsprechend der scikit-learn-Dokumentation) wurde so initialisiert, dass er für jeden Datenpunkt die Klasse prädiziert hat, die am häufigsten vorkam (bei Klassifikationsanalysen) (125,154). Dabei wurde die Vorhersage bei ungleich verteilten Klassenhäufigkeiten unter Beibehaltung der Verteilungshäufigkeiten entsprechend angepasst (125,154), sodass der Wert jeweils 0.5

betrug. Bei Regressionsanalysen wurde jeweils der Mittelwert der Zielvariable als Referenz betrachtet (125,155). Bei den Metriken, die zur Evaluation der Modelle im folgenden Ergebnisteil angegeben sind, handelt es sich (wie in Kapitel 2.3.2 angedeutet) jeweils um das arithmetische Mittel der Ergebnisse aller 10 äußeren Faltungen.

Permutationstests

Um die statistische Signifikanz der Ergebnisse sicherzustellen, wurden – wie in der Literatur empfohlen (156,157) – Permutationstests durchgeführt. Dabei wurden die Analysen mit zufällig permutierten Zielvariablen/Labels (unter Beibehaltung der ursprünglichen Verteilungshäufigkeiten) erneut berechnet (64). Die Anzahl an Permutationen wurde im Hinblick auf die Rechendauer auf $n = 1000$ festgelegt. Dieser Wert wurde ebenfalls in vorherigen Arbeiten im Bereich der psychiatrischen Forschung verwendet (71,78,83,158).

Da die Klassifikationsanalysen zur Klärung der 1. Hypothese im Hinblick auf die Prüfung der Reproduzierbarkeit der in der Literatur konsistent beschriebenen Gruppenunterschiede zwischen gesunden Kontrollen und MDD-Patienten besonders relevant waren, wurden dafür grundsätzlich Permutationstests durchgeführt (s. Kapitel 2.4). Aufgrund der hohen involvierten computerbezogenen Ressourcen wurden die Permutationstests konsistent nur für die Analysen anhand der gesamten Stichproben (die Modelle sollten für die Gesamtgruppen valide sein), nicht aber für die Subgruppenanalysen durchgeführt.

Für die explorativen Analysen zur Klärung der 2. Hypothese wurde zudem ein arbiträrer Grenzwert festgelegt, um zu entscheiden, wann ein Permutationstest durchgeführt wird und wann nicht. Dieser Grenzwert orientierte sich an der klinischen Relevanz der Ergebnisse und wurde als überschritten betrachtet, wenn die Prädiktion eines Modells eine um mindestens 7.5 % bessere Performance als der Dummy erreichte.

Ein Ergebnis wurde als statistisch signifikant eingestuft, wenn in dem dazugehörigen Permutationstest ein p-Wert (Anteil der Permutationen, bei denen ein besseres Ergebnis als bei der Analyse mit den wahren Labels erreicht wurde (78)) von ≤ 0.05 erreicht wurde. In PHOTONAI ist eine Funktion zum Durchführen von Permutationstests implementiert (64).

Die für diese Studie verwendeten Skripte, in denen die PHOTONAI-Modelle spezifiziert sind, finden sich im Anhang beispielhaft für eine Klassifikations- und eine Regressionsanalyse (s. Anhang 8.10).

2.4 Beschreibung der durchgeführten Analysen

Im Hinblick auf die in Kapitel 1.4 vorgestellten Hypothesen wurden folgende Analysen unter Anwendung der in Kapitel 2.3 vorgestellten Methoden durchgeführt:

Analysen der Gesamtstichproben

Wie in Abbildung 2-6 schematisch dargestellt, wurden die beiden Modalitäten grundsätzlich differenziert voneinander betrachtet. Zunächst wurde zur Klärung der ersten Hypothese geprüft, ob eine Unterscheidung zwischen Probanden mit der Hauptdiagnose einer MDD und psychiatrisch gesunden Kontrollen anhand der Prädiktoren möglich ist.

Analysen innerhalb der MDD-Gruppen

Zur Klärung der zweiten Hypothese wurde der Zusammenhang zwischen den Modalitäten und einzelnen Krankheitscharakteristika innerhalb der MDD-Gruppen geprüft. Da die Ausprägung kognitiver Defizite bzw. struktureller Hirnveränderungen bei der MDD unklar ist (s. Kapitel 1.2) und entsprechend der in Kapitel 1.4 vorgestellten explorativen Art dieser Arbeit, fiel die Wahl der Zielvariablen entsprechend umfangreich aus. Die eingesetzten psychometrischen Instrumente zur Quantifizierung der Krankheitscharakteristika wurden bereits in Kapitel 2.1.3. vorgestellt. Pro Modalität wurden jeweils zwei Klassifikationsanalysen und vier Regressionsanalysen durchgeführt.

Im Rahmen der beiden Klassifikationsanalysen wurden der zeitliche Krankheitsverlauf (rezidivierender Verlauf vs. einzelne Episode) und der Hospitalisierungsstatus (Proband zum Zeitpunkt des Studieneinschlusses in (teil-)stationärer Behandlung oder nicht) als Zielvariablen betrachtet. Der zeitliche Krankheitsverlauf stellt in Bezug auf den in Kapitel 1.1 vorgestellten rezidivierenden Verlauf der MDD einen relevanten Parameter dar und war durch die hier gewählte Unterscheidung von zwei Verläufen gut differenzierbar. Bei dem Hospitalisierungsstatus handelt es sich um ein klinisch gut messbares Äquivalent der Krankheitsintensität bzw. der Depressionsschwere.

Im Rahmen der Regressionsanalysen wurde ebenfalls die Depressionsschwere (durch HAMD- und BDI-Werte quantifiziert) als Zielvariable gewählt. Dabei handelt es sich einerseits um ein Fremd- und andererseits um ein Selbstbeurteilungsinstrument, sodass die sehr subjektive Einschätzung der Krankheitsschwere aus zwei verschiedenen Perspektiven betrachtet werden konnte. Zusätzlich wurden das allgemeine Funktionsniveau (GAF) und die

körperliche Funktionsfähigkeit (entsprechend der Subskala des SF-36) als Zielvariablen gewählt. Im Gegensatz zur Abfrage der Symptomschwere anhand einschlägiger Diagnosekriterien (wie es bei der Hamilton-Skala und dem BDI der Fall ist) konnten dadurch zusätzlich soziale, physische und berufliche Funktionsaspekte berücksichtigt werden. Wie in Kapitel 2.1.3 beschrieben, werden körperliche Einschränkungen in der GAF-Skala nicht unmittelbar mitbewertet. Dass die MDD Auswirkungen auf die allgemeine und die körperliche Funktionsfähigkeit von Betroffenen haben kann, wurde allerdings in Kapitel 1.1.1 erläutert, sodass dies hier differenziert betrachtet wurde.

Durch die Auswahl dieser Zielvariablen wurden sowohl in der Literatur einschlägig diskutierte (teils subjektive) psychometrische Instrumente (siehe Kapitel 1.2) als auch klinisch gut messbare „harte“ Marker betrachtet.

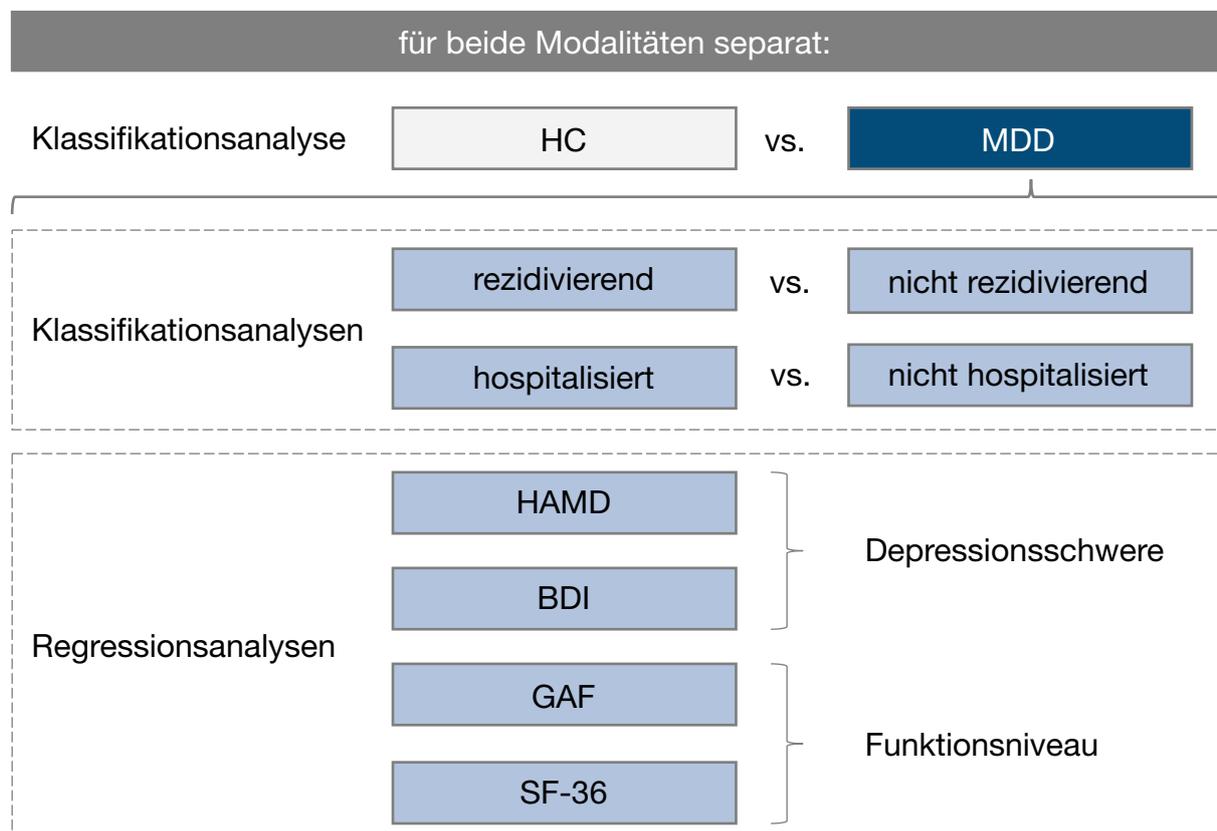


Abbildung 2-6: Übersicht über die durchgeführten Analysen

Anmerkungen: Dargestellt sind die verschiedenen Klassifikations- und Regressionsanalysen, die zur Klärung der Hypothesen in dieser Arbeit für beide Modalitäten jeweils separat durchgeführt wurden. HC = gesunde Kontrollen (engl.: healthy controls, HC), MDD = Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD), HAMD = Hamilton Depressions-Skala, BDI = Beck-Depressions-Inventar, GAF = Global Assessment of Functioning, SF-36 = Short Form 36-Fragebogen

Subgruppenanalysen der Gesamtstichproben

Wie in Kapitel 2.2 erläutert, wurde auf ein Matching der Kontrollgruppen und der MDD-Gruppen explizit verzichtet. Stattdessen wurden sämtliche Analysen für verschiedene Subgruppen (Geschlecht, Alter, Bildungsgrad, Studienzentrum) durchgeführt. Bezüglich des Alters wurde eine Unterteilung in drei Gruppen vorgenommen (< 30 , ≥ 30 und ≤ 55 , > 55 Jahren), welche einerseits im Hinblick auf die Altersverteilung der Stichproben als auch im Hinblick auf die Leistungsfähigkeit unterschiedlicher Altersgruppen gewählt wurde. Der Bildungsgrad wurde ebenfalls in drei Gruppen unterteilt (1. Gruppe: Lehre, Meister, allgemeine Hochschulreife, Bachelor- bzw. Masterabschluss, 2. Gruppe: Volks-, Haupt- oder Realschulabschluss bzw. Fachhochschulreife, 3. Gruppe: kein Schulabschluss). Da die dritte Gruppe (kein Schulabschluss) in allen Stichproben nur max. zwei Probanden umfasste (s. Kapitel 2.2), wurde diese Subgruppe in den Analysen jeweils ausgeklammert.

Subgruppenanalysen innerhalb der MDD-Gruppen

Im Rahmen der in der zweiten Phase innerhalb der MDD-Gruppen durchgeführten Analysen wurden zusätzlich sechs weitere Subgruppen differenziert betrachtet. Das geschah, um im Detail analysieren zu können, ob die untersuchten Zielvariablen (s. Kapitel 2.1.3) in besonderer Weise mit einzelnen Krankheitscharakteristika assoziiert sind und um Unterschiede in den Verteilungen der einzelnen Charakteristika berücksichtigen zu können.

Im Detail wurden (konsistent mit den Zielvariablen) der zeitliche Krankheitsverlauf (rezidivierender Verlauf vs. einzelne Episode) und der Hospitalisierungsstatus (Proband zum Zeitpunkt des Studieneinschlusses in (teil-) stationärer Behandlung oder nicht) unterschieden. Zusätzlich zum Hospitalisierungsstatus wurde zudem differenziert, ob Probanden in der Vorgeschichte mind. einen auf die MDD bezogenen (teil-)stationären Aufenthalt hatten (inkl. des Zeitpunktes des Studieneinschlusses) oder nicht (Hospitalisierungsgrad).

Neben diesen gut messbaren Markern wurde zudem der Remissionsstatus zum Zeitpunkt der Studienteilnahme (im Rahmen des SKID-I ermittelt, s. Kapitel 2.1.3) in drei getrennten Gruppen betrachtet (akute Episode, partielle bzw. volle Remission). Außerdem wurde unterschieden, ob die Probanden zum Zeitpunkt des Studieneinschlusses mind. ein Antidepressivum einnahmen oder nicht und, ob Probanden zusätzlich zu der MDD eine psychiatrische Komorbidität aufwiesen. Diese beiden letzteren Informationen waren nicht für alle Probanden bekannt.

3 Ergebnisse

Nach einer einleitenden Beschreibung der Stichproben (s. Kapitel 3.1) werden im Folgenden die Ergebnisse der in Kapitel 2.4 erläuterten Analysen dargestellt. Für die beiden betrachteten Modalitäten wurden dabei (konsistent mit der Gliederung der Stichproben in Kapitel 2.2) getrennte Kapitel (s. Kapitel 3.2 und 3.3) angelegt.

Wie in Kapitel 2.3.2 beschrieben, wurde für die Durchführung der Berechnungen jeweils vorausgesetzt, dass mindestens 10 Probanden pro Klasse repräsentiert waren. War das nicht der Fall, wurden die entsprechenden Ergebnisfelder mit Nullen kodiert. Wenn nicht explizit anders formuliert, beziehen sich die im Fließtext beschriebenen Ergebnisse auf die Evaluation der Modelle anhand der Testdaten.

Der Übersichtlichkeit halber sind die umfangreichen Tabellen, die im Rahmen dieser Arbeit zur Darstellung der Ergebnisse erstellt wurden, zum Teil nicht in diesem Kapitel, sondern stattdessen im Anhang dargestellt (s. Anhang 8.11 und 8.12). Die Entscheidung darüber, ob eine Tabelle im Anhang oder in diesem Kapitel dargestellt ist, richtete sich nach deren Relevanz. Konsistent mit der in Kapitel 2.3.3 beschriebenen Entscheidung darüber, wann ein Permutationstest durchgeführt wurde und wann nicht, sind auch an dieser Stelle die Ergebnisse der Analysen zu Gruppenunterschieden zwischen gesunden Kontrollen und MDD-Patienten grundsätzlich aufgeführt. Die übrigen Ergebnistabellen sind in diesem Kapitel dargestellt, wenn die jeweiligen Zielwerte der entsprechenden Analysen über alle Subgruppen hinweg weitestgehend konsistent besser als anhand des Dummys prädiziert werden konnten.

3.1 Beschreibung der Stichproben

Entsprechend des in Kapitel 2.2 beschriebenen Prozesses der Stichprobengenerierung, konnten für die Analysen anhand der Daten der neuropsychologischen Testung insgesamt $n = 1650$ Probanden und für die Analysen anhand der strukturellen MRT-Daten insgesamt $n = 1047$ eingeschlossen werden. Die expliziten Stichprobencharakteristika können den Kapiteln 2.2.1 und 2.2.2 entnommen werden (s. Tabelle 2-3, Tabelle 2-4, Tabelle 2-5 und Tabelle 2-6).

Demographische Daten

Ein statistischer Vergleich einzelner demographischer Charakteristika (Alter, Geschlecht, Bildungsgrad und Studienzentrum) zwischen den jeweiligen gesunden Kontrollgruppen und den entsprechenden MDD-Gruppen entfiel in dieser Arbeit durch die Bildung entsprechender Subgruppen (s. Erläuterungen in Kapitel 2.2 und Kapitel 2.4). Aus diesem Grund erfolgt an dieser Stelle lediglich eine orientierende Beschreibung von Unterschieden in den Verteilungen. Je nach betrachteter Stichprobe betrug das Durchschnittsalter ca. 32 bis 37 Jahre. Für beide Modalitäten konnten anteilig mehr gesunde Kontrollen als MDD-Patienten (Verhältnis ca. 1.2 – 1.4 : 1) sowie mehr weibliche als männliche Probanden (Verhältnis ca. 1.8 – 1.9 : 1) eingeschlossen werden. Auch die Verteilung der Anzahl der Einschlüsse unterschied sich zwischen den beiden Studienzentren, wobei je nach Stichprobe teils mehr Probanden in Marburg und teils mehr in Münster eingeschlossen wurden. Insgesamt wies die überwiegende Mehrheit der eingeschlossenen Probanden einen über den reinen Schulabschluss hinausgehenden Bildungsgrad auf.

Psychometrische Daten

Die Stichprobe, die für die Analysen anhand der strukturellen MRT-Daten generiert wurde, enthielt weitestgehend Probanden, die auch in die Stichprobe für die Analysen anhand der neuropsychologischen Testdaten eingeschlossen wurden. Durch die zusätzliche Filterung in Bezug auf die MRT-Daten, war erstere lediglich weniger umfangreich. Die betrachteten psychometrischen Charakteristika (HAMD, BDI, GAF, SF-36 Subskala: körperliche Funktionsfähigkeit) unterschieden sich im Gruppenvergleich signifikant (5 % Signifikanzniveau) für beide Modalitäten gleichermaßen, was in Anbetracht der Überschneidungen zwischen den Stichproben zu erwarten war.

Zur qualitativen Einordnung der psychometrischen Charakteristiken sind die entsprechenden Werte im Folgenden beschrieben, wobei das erste Wertepaar jeweils für die Modalität der neuropsychologischen Testung und das zweite für die Modalität der strukturellen MRT-Daten gilt. Im Mittel wiesen die Probanden der MDD-Gruppe ca. 9.14 (\pm 7.09) bzw. 9.10 (\pm 7.27) Punkte in der Hamilton-Skala und 17.14 (\pm 10.91) bzw. 17.72 (\pm 11.09) Punkte im BDI auf, was einer milden bis mittelschweren Depression entsprach. Die im Mittel sehr niedrigen HAMD-Werte von 1.45 (\pm 2.18) bzw. 1.29 (\pm 1.97) und BDI-Werte von 4.11 (\pm 4.19) bzw. 3.89 (\pm 4.13) der Kontrollgruppe mit einer gleichermaßen niedrigeren Standardabweichung entsprachen keiner oder allenfalls einer minimalen Depression.

Ein ähnliches Bild ergab sich bei Betrachtung der Skalen, die das Funktionsniveau quantifizieren. Entsprechend der GAF-Werte von 65.33 (\pm 13.56) bzw. 65.31 (\pm 14.23) wiesen die MDD-Patienten ein im Vergleich zu den Normalprobanden (Punktwerte: 91.11 (\pm 7.68) bzw. 92.28 (\pm 6.44)) deutlich eingeschränktes globales Funktionsniveau auf. Im Hinblick auf die Maximalpunktzahl von 100 bei der Bewertung der GAF-Skala wies die Kontrollgruppe eine annähernd volle Funktionsfähigkeit auf. Der SF-36-Wert (Subskala: körperliche Funktionsfähigkeit) wies vor allem innerhalb der MDD-Gruppe vergleichsweise große Standardabweichungen auf. Auch hier deuteten die Werte auf eine insgesamt niedrigere körperliche Funktionsfähigkeit der MDD-Patienten (85.93 (\pm 17.91) bzw. 85.10 (\pm 18.29) Punkte) im Vergleich zur Kontrollgruppe hin (95.79 (\pm 9.85) bzw. 96.85 (\pm 7.37) Punkte).

Daten zum Verlauf der Erkrankung der MDD-Patienten

Auch hinsichtlich dieser Kriterien wurden Subgruppen gebildet, um Unterschiede in den Verteilungen zu berücksichtigen (s. Kapitel 2.4). Zur Einordnung des durchschnittlichen Krankheitsverlaufes der eingeschlossenen MDD-Patienten werden die Verteilungen im Folgenden dennoch knapp erläutert. Für beide Modalitäten wies die Mehrzahl der eingeschlossenen MDD-Patienten einen rezidivierenden Verlauf und nicht nur eine einzelne depressive Episode auf (Verhältnis ca. 1.9 – 2.3 : 1). Die Mehrzahl der Probanden war zudem mind. einmal in (teil-)stationärer psychiatrischer Behandlung in der Krankheitsgeschichte (Verhältnis ca. 1.9 – 2.4 : 1), wies zum Zeitpunkt des Studieneinschlusses eine akute depressive Episode und keine partielle oder volle Remission auf (Verhältnis je ca. 1.2 – 1.8 : 1), war aber zu diesem Zeitpunkt nicht in (teil-)stationärer Behandlung (Verhältnis ca. 1.8 – 2.5 : 1). Die Mehrzahl der Probanden nahm mind. ein Antidepressivum ein (Verhältnis ca. 1.4 – 1.5 : 1) und wies keine psychiatrische Komorbidität auf (Verhältnis ca. 1.6 – 1.8 : 1), wobei diese beiden Informationen nicht für alle Probanden vorlagen.

3.2 Analysen zur Modalität: neuropsychologische Testung

3.2.1 Klassifikationsanalyse: MDD vs. HC

Wie in Tabelle 3-1 beschrieben, konnte bei der Unterscheidung der MDD-Patienten ($n = 736$) von gesunden Kontrollen (HC) ($n = 914$) anhand der neuropsychologischen Testdaten innerhalb der kompletten für diese Analyse betrachteten Stichprobe ($n = 1650$) eine korrigierte Klassifikationsgenauigkeit von 59.62 % (\pm 4.06 %) erreicht werden. Die erreichte mittlere Spezifität betrug 74.27 % (\pm 7.60 %), die mittlere Sensitivität 44.97 % (\pm 5.00 %). Innerhalb des Trainings betrug die korrigierte Klassifikationsgenauigkeit für die gesamte Stichprobe

69.59 % (± 8.04 %). Die Maschine mit der optimalen Performance war ein Random Forest (maximale Tiefe = None (= keine Beschränkung), minimale Anzahl an Samples pro Blatt = 10).

Auch innerhalb der Subgruppen konnten weitestgehend ähnliche Klassifikationsgenauigkeiten erreicht werden. In der mittleren Altersgruppe ($n = 718$), fiel die Genauigkeit mit ca. 57.80 % (± 6.52 %) etwas geringer aus. Die Genauigkeiten innerhalb der Bildungsniveau-Subgruppen (55.98 % (± 1.97 %) bzw. 53.73 % (± 5.44 %)) blieben jedoch deutlich hinter der Performance der Gesamtgruppen-Analyse zurück. In der Gruppe mit den Probanden, die zum Zeitpunkt der Untersuchung älter als 55 Jahre waren ($n = 159$), fiel die Standardabweichung der korrigierten Klassifikationsgenauigkeit mit 13.01 % vergleichsweise (in Relation zu den Werten der anderen Subgruppen) hoch aus.

Tabelle 3-1: Klassifikationsanalyse: MDD vs. HC, Modalität: neuropsychologische Testung

	MDD ^a		HC ^a	Dummy Test	Test	Perf. ^b
	n	n	n	BAC M (SD)	BAC M (SD)	BAC _{test} - BAC _{dum.}
komplette Stichprobe	1650	736	914	0.50 (0.00)	0.60 (0.04)	0.10
Geschlecht						
männlich	582	253	329	0.50 (0.00)	0.60 (0.05)	0.10
weiblich	1068	483	585	0.50 (0.00)	0.60 (0.02)	0.10
Alter						
> 55	159	77	82	0.50 (0.00)	0.60 (0.13)	0.10
≥ 30 und ≤ 55	718	352	366	0.50 (0.00)	0.58 (0.07)	0.08
< 30	773	307	466	0.50 (0.00)	0.61 (0.04)	0.11
höchster Bildungsgrad						
Lehre, Abi, Meister, Uni ^c	1250	504	746	0.50 (0.00)	0.56 (0.02)	0.06
Schulabschluss ^d	397	230	167	0.50 (0.00)	0.54 (0.05)	0.04
Studienzentrum						
Marburg	868	323	545	0.50 (0.00)	0.59 (0.04)	0.09
Münster	782	413	369	0.50 (0.00)	0.60 (0.04)	0.10

Bei den Metriken (M (SD)) handelt es sich um gemittelte Werte der entsprechenden Metriken aller 10 äußeren Faltungen. BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD) bzw. gesunde Kontrollen (engl.: healthy controls, HC), ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife

Die Durchführung eines Permutationstests für die Analyse anhand der kompletten Stichprobe bestätigte, dass die Ergebnisse der Gesamtgruppenanalyse statistisch signifikant waren (s.

Abbildung 3-1). Dabei wurde innerhalb keiner einzelnen Permutation eine gleichwertige oder bessere gewichtete Klassifikationsgenauigkeit als bei dem originalen Modell ohne permutierte Label erreicht (p-Wert = 0.00).

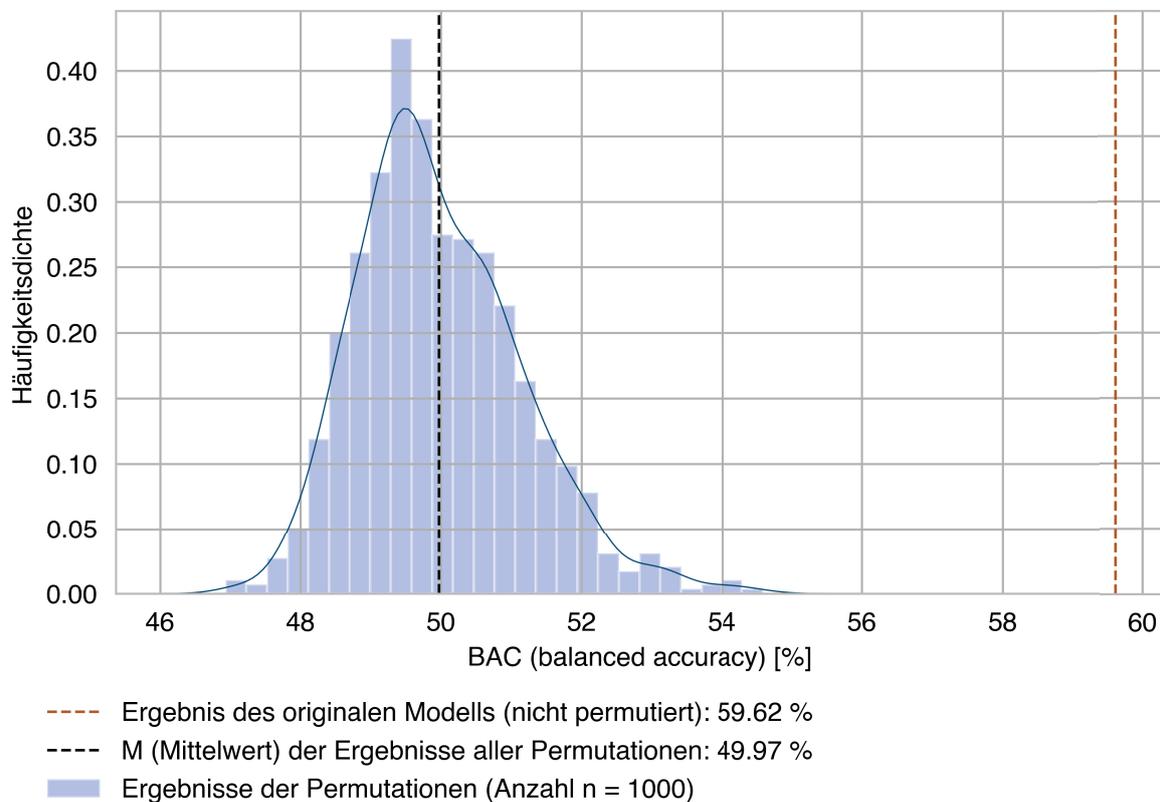


Abbildung 3-1: Permutationstest, Klassifikationsanalyse: MDD vs. HC, Modalität: neuropsych. Testung

Anmerkungen: Zur Sicherung der statistischen Signifikanz der Ergebnisse (bei der Analyse der kompletten Stichprobe) wurde ein Permutationstest durchgeführt, wobei die Analyse insgesamt 1000 Mal mit jeweils zufällig permutierten Labels durchgeführt wurde (p-Wert = 0.00). HC = gesunde Kontrollen (engl.: healthy controls, HC), MDD = Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD)

3.2.2 Klassifikationsanalyse: rezidivierender Verlauf vs. einzelne Episode

In Tabelle 8-7 im Anhang sind die Ergebnisse der Klassifikationsanalyse zur Unterscheidung verschiedener zeitlicher Krankheitsverläufe (rezidivierender Verlauf vs. einzelne Episode) innerhalb der MDD-Gruppe (n = 736) beschrieben. Die korrigierte Klassifikationsgenauigkeit betrug 50.02 % (\pm 3.53 %) und war somit kaum besser als die Vergleichsmetrik (der Dummy).

Die höchsten Genauigkeiten konnten bei der Betrachtung einzelner Subgruppen erreicht werden. In der Subgruppe mit den Probanden, die zum Zeitpunkt der Untersuchung voll remittiert waren (n = 234), ließ sich die Klassifikation mit einer korrigierten Genauigkeit von 55.60 % (\pm 4.16 %) durchführen. Bei Betrachtung der Probanden ohne psychiatrische

Komorbidität (n = 327), konnte sogar eine korrigierte Genauigkeit von 58.89 % (± 6.04 %) erreicht werden, wobei die Information für n = 201 Probanden nicht bekannt war (s. Tabelle 2-4). Da die Ergebnisse innerhalb der anderen Subgruppen teils etwas besser bzw. ähnlich gut und teils etwas schlechter als die Dummy-Ergebnisse ausfielen, konnte hier keine eindeutige Assoziation festgestellt werden.

3.2.3 Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert

Auch die Klassifikationsanalyse zur Unterscheidung, ob Probanden zum Zeitpunkt der Studienteilnahme in (teil-)stationärer Behandlung waren oder nicht, ergab in der gesamten MDD-Gruppe (n = 736) mit einer korrigierten Klassifikationsgenauigkeit von 52 % (± 3.53 %) kaum bessere Werte als anhand des Dummys (s. Tabelle 8-8 im Anhang).

Die Analyse der einzelnen Subgruppen ergab in den meisten Fällen eine etwas bessere oder mit dem Dummy vergleichbare Genauigkeit. In den Subgruppen mit Probanden > 55 Jahren (n = 77) bzw. mittlerem Schulabschluss (n = 230) bzw. mind. einmaliger (teil-)stationärer Behandlung in der Krankengeschichte (n = 492) und in der Subgruppe mit den Probanden, die zum Studienzeitpunkt mind. ein Antidepressivum einnahmen (n = 275), konnten mit korrigierten Genauigkeiten von 55.02 % (± 20.82 %), 56.92 % (± 11.98 %), 54.74 % (± 4.40 %) und 55.93 % (± 7.32 %) die höchsten Werte erreicht werden. Zusammenfassend ließ sich der Hospitalisierungsstatus aber auch in den Subgruppen kaum besser als anhand des Dummys präzisieren.

3.2.4 Regressionsanalyse: HAMD-Wert

Innerhalb der MDD-Gruppe konnte die Depressionsschwere (quantifiziert mittels der Hamilton-Skala) nicht eindeutig anhand der Leistung in der neuropsychologischen Testung präzisiert werden.

Wie aus Tabelle 8-9 im Anhang ersichtlich, ergab der mittels Regressionsanalyse errechnete Mittelwert der absoluten Fehler für die Gesamtgruppe (n = 736) mit 5.82 (± 0.12) eine kaum geringere Abweichung vom wahren Wert (im Interview ermittelt) als die anhand des Dummys ermittelte Vergleichsmetrik mit 5.88 (± 0.09). Somit war die Genauigkeit des Modells nur knapp 1 % besser als der Dummy. Die Pearson-Korrelation für die gesamte Stichprobe betrug im Mittel $r = 0.08$ (± 0.08).

Innerhalb der einzelnen Subgruppen wurden teils positive (z. B. Subgruppe der weiblichen Probanden, $n = 483$, $r = 0.12 \pm 0.18$) und teils negative Korrelationen (z. B. Subgruppe mit Probanden in partieller Remission, $n = 190$, $r = -0.19 \pm 0.19$) ermittelt. Auch hinsichtlich der Mittelwerte der absoluten Fehler stellte sich in den Subgruppen-Analysen ein inkonsistentes Bild dar. Teils konnten im Vergleich zum Dummy etwas geringere Fehler erreicht werden (z. B. Subgruppe der weiblichen Probanden, $n = 483$, ca. 1 % besser) und teils fiel der Fehler etwas größer aus (z. B. Subgruppe der Probanden mit rezidivierendem Krankheitsverlauf, $n = 474$, ca. 2 % schlechter).

3.2.5 Regressionsanalyse: BDI-Wert

Ähnlich wie schon bei der Regressionsanalyse des HAMD-Wertes konnte auch der BDI-Wert nicht eindeutig anhand der Ergebnisse der neuropsychologischen Testung prädiziert werden. Wie aus Tabelle 8-10 im Anhang ersichtlich, konnte der Mittelwert des absoluten Fehlers im Rahmen der Analyse der kompletten MDD-Stichprobe ($n = 736$) im Vergleich zum Dummy kaum reduziert werden. Dabei wurde eine gering positive Korrelation ermittelt ($r = 0.09 (\pm 0.14)$).

Im Rahmen der Subgruppenanalysen wurden teils positive (z. B. Subgruppe der weiblichen Probanden, $n = 483$, $r = 0.15 \pm 0.09$) und teils negative Korrelationen (z. B. Subgruppe der Probanden mit psychiatrischer Komorbidität, $n = 208$, $r = -0.08 \pm 0.18$) sowie teils etwas geringere (z. B. Subgruppe der weiblichen Probanden, $n = 483$, ca. 1 % besser) und teils etwas größere mittlere absolute Fehler (z. B. Subgruppe der Probanden < 30 Jahre, $n = 307$, ca. 2 % schlechter) als anhand des Dummys ermittelt.

3.2.6 Regressionsanalyse: GAF-Wert

In Tabelle 8-11 im Anhang sind die Ergebnisse der Analyse des Zusammenhangs zwischen dem Funktionsniveau (quantifiziert mittels des GAF-Wertes) und den neuropsychologischen Testdaten innerhalb der MDD-Gruppe beschrieben. Für die komplette Stichprobe ($n = 530$) war der mittlere absolute Fehler des Modells kaum geringer als der Vergleichswert. Die Pearson-Korrelation betrug $0.2 (\pm 0.12)$.

Auch wenn die Korrelation im Rahmen der Betrachtung der einzelnen Subgruppen weitgehend positiv blieb, so fanden sich Werte zwischen $r = 0.00 (\pm 0.34)$ (in der Subgruppe mit Probanden in partieller Remission ($n = 138$)) und $r = 0.37 (\pm 0.52)$ (in der Subgruppe der Probanden, die zum Zeitpunkt der Untersuchung älter als 55 Jahre waren ($n = 46$)).

Dementsprechend ließ sich keine eindeutige Korrelation zwischen den Prädiktoren und der Zielvariable feststellen. Hinsichtlich des mittleren absoluten Fehlers wurden in der Analyse ebenfalls teils etwas geringere (z. B. Subgruppe weiblichen Probanden, $n = 349$, ca. 3 % besser) und teils etwas höhere Werte (z. B. Subgruppe der Probanden < 30 Jahre, $n = 224$, ca. 6 % schlechter) als die Vergleichswerte ermittelt.

3.2.7 Regressionsanalyse: SF-36-Wert (Subskala: körperliche Funktionsfähigkeit)

Die Prädiktion der Werte des SF-36 (Subskala: körperliche Funktionsfähigkeit) gelang für die komplette Stichprobe innerhalb der MDD-Gruppe ($n = 736$) mit einer ca. 9 % besseren Genauigkeit als anhand des Dummys (s. Tabelle 3-2). Der durch das Modell ermittelte mittlere absolute Fehler betrug dabei $12.18 (\pm 0.39)$ Punkte, wohingegen der anhand des Dummys ermittelte Fehler bei $13.34 (\pm 0.14)$ Punkten lag. Dabei wurde eine positive Pearson-Korrelation ermittelt ($r = 0.24 (\pm 0.14)$). Innerhalb des Trainings wurde ein mittlerer absoluter Fehler von $11.08 (\pm 0.48)$ und eine Korrelation von $r = 0.43 (\pm 0.06)$ erreicht. Die beste Maschine war eine SVM ($C = 1$, Kernel = rbf).

Auch innerhalb der Subgruppen konnte mittels der prädiktiven Modelle konsistent eine höhere Performance (geringerer mittlerer absoluter Fehler) als anhand des Dummys erreicht werden. Dabei schwankte die Performanceverbesserung im Bereich von 1 bis 13 %. Insbesondere in der Subgruppe der Probanden, die sich zum Zeitpunkt der Studienteilnahme in (teil-)stationärer Behandlung befanden ($n = 218$), konnte eine Verbesserung von nur 1 % erreicht werden. In der Subgruppe der Probanden, die noch nie in (teil-)stationärer Behandlung waren ($n = 244$) sowie der Gruppe der Probanden, die zum Zeitpunkt der Untersuchung kein Antidepressivum einnahmen ($n = 186$), konnten die größten Verbesserungen im Vergleich zum Dummy erreicht werden (jeweils ca. 13 %).

Die Pearson-Korrelation war in allen Subgruppen durchweg positiv, wobei sie innerhalb der Subgruppe mit den Probanden, die zum Zeitpunkt der Untersuchung genau eine depressive Episode ($n = 262$) aufwiesen mit $r = 0.27 (\pm 0.20)$ sowie in der Subgruppe der Probanden, die zum Zeitpunkt der Untersuchung eine akute Episode aufwiesen ($n = 312$) mit $r = 0.27 (\pm 0.16)$ die größten Beträge annahm. In der Subgruppe der Probanden, die zum Zeitpunkt der Studie in voller Remission waren ($n = 234$), nahm die Pearson-Korrelation mit $r = 0.07 (\pm 0.18)$ den geringsten Wert an.

Tabelle 3-2: Regressionsanalyse: SF-36-Wert, Modalität: neuropsychologische Testung

	n	SF-36		Dummy Test		Test		Perf. ^a		Test	
		M (SD)		MAE M (SD)		MAE M (SD)		$\frac{MAE_{test}}{MAE_{dum.}}$	r	M (SD)	
komplette Stichprobe	736	85.93 (17.91)		13.45 (0.14)		12.18 (0.39)		0.91	0.24	(0.14)	
Geschlecht											
männlich	253	88.66 (15.56)		11.51 (0.37)		10.65 (0.87)		0.93	0.22	(0.14)	
weiblich	483	84.50 (18.88)		14.30 (0.23)		13.10 (0.49)		0.92	0.24	(0.12)	
Alter											
> 55	77	78.18 (20.60)		16.15 (1.40)		15.33 (2.46)		0.95	0.23	(0.27)	
≥ 30 und ≤ 55	352	83.14 (20.23)		15.82 (0.28)		14.50 (0.83)		0.92	0.25	(0.18)	
< 30	307	91.07 (12.03)		8.92 (0.47)		7.98 (0.38)		0.89	0.09	(0.18)	
höchster Bildungsgrad											
Lehre, Abi, Meister, Uni ^b	504	88.36 (15.99)		11.66 (0.25)		10.46 (0.34)		0.90	0.19	(0.17)	
Schulabschluss ^c	230	80.83 (20.50)		16.10 (0.44)		15.51 (1.12)		0.96	0.14	(0.20)	
Zeitl. Krankheitsverlauf											
rezidivierend	474	84.26 (18.62)		14.27 (0.22)		13.16 (0.66)		0.92	0.22	(0.17)	
einzelne Episode	262	88.95 (16.15)		11.35 (0.54)		10.18 (1.76)		0.90	0.27	(0.20)	
Hospitalisierungsgrad^d											
mind. einmal	492	83.14 (18.94)		14.68 (0.28)		13.46 (0.47)		0.92	0.22	(0.08)	
noch nie	244	91.56 (14.06)		9.39 (0.49)		8.18 (0.61)		0.87	0.14	(0.17)	
Hospitalisierungsstatus^d											
hospitalisiert	218	81.67 (18.85)		14.74 (0.46)		14.53 (1.50)		0.99	0.19	(0.32)	
nicht hospitalisiert	518	87.72 (17.20)		12.50 (0.27)		10.98 (0.53)		0.88	0.25	(0.13)	
Remissionsstatus											
akut	312	80.90 (20.86)		16.17 (0.27)		14.56 (1.01)		0.90	0.27	(0.16)	
partielle Remission	190	86.24 (16.56)		12.88 (0.56)		12.23 (0.59)		0.95	0.16	(0.24)	
volle Remission	234	92.39 (11.51)		7.96 (0.48)		7.08 (0.45)		0.89	0.07	(0.18)	
Antidepressiva											
mind. eins	275	81.64 (19.84)		15.52 (0.31)		13.99 (1.00)		0.90	0.25	(0.15)	
keins	186	89.41 (15.23)		11.20 (0.55)		9.75 (0.59)		0.87	0.17	(0.19)	
Psych. Komorbidität											
mind. eine	208	83.44 (20.22)		15.16 (0.67)		14.06 (1.23)		0.93	0.25	(0.22)	
keine	327	88.44 (15.08)		11.69 (0.34)		10.59 (0.72)		0.91	0.24	(0.16)	
Studienzentrum											
Marburg	323	83.45 (19.38)		14.91 (0.31)		13.66 (0.58)		0.92	0.21	(0.17)	
Münster	413	87.87 (16.43)		12.08 (0.39)		10.79 (0.40)		0.89	0.24	(0.10)	

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, SF-36 = Short Form 36-Fragebogen (Subskala: körperliche Funktionsfähigkeit), ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlungen

Für die Analyse anhand der kompletten MDD-Stichprobe wurde im Permutationstest (s. Abbildung 3-2) für keine einzelne Permutation ein gleichwertiger oder niedrigerer absoluter Fehler als anhand des Modells ohne permutierte Label ermittelt, sodass das Ergebnis dieser Analyse als statistisch signifikant eingestuft werden kann (p -Wert = 0.00).

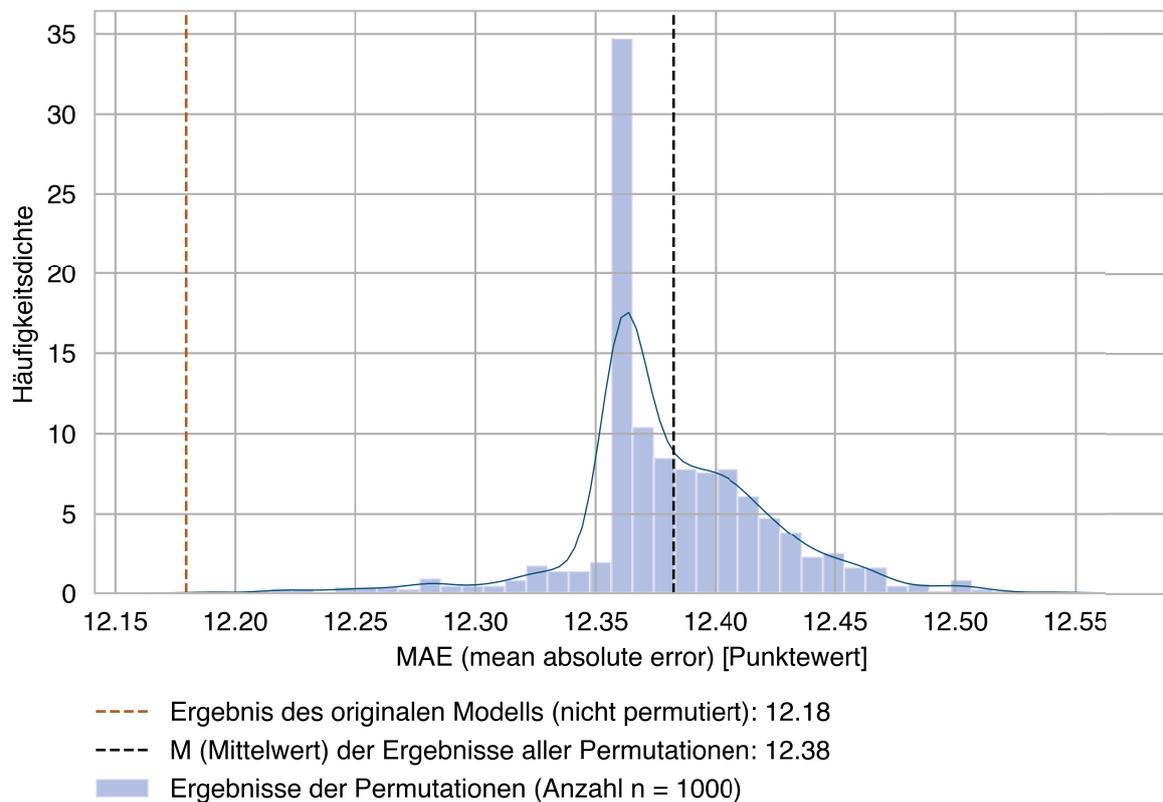


Abbildung 3-2: Permutationstest, Regressionsanalyse: SF-36-Wert (körperliche Funktionsfähigkeit), Modalität: neuropsychologische Testung

Anmerkungen: Zur Sicherung der statistischen Signifikanz der Ergebnisse (bei der Analyse der kompletten Stichprobe) wurde ein Permutationstest durchgeführt, wobei die Analyse insgesamt 1000 Mal mit jeweils zufällig permutierten Zielwerten durchgeführt wurde (p -Wert = 0.00). SF-36 = Short Form 36-Fragebogen

3.3 Analysen zur Modalität: strukturelle MRT-Daten

3.3.1 Klassifikationsanalyse: MDD vs. HC

Die Unterscheidung von MDD-Patienten ($n = 432$) und gesunden Kontrollen ($n = 615$) anhand der strukturellen MRT-Daten gelang im Rahmen der Analyse der kompletten Stichprobe ($n = 1047$) mit einer korrigierten Klassifikationsgenauigkeit von 55.34 % (± 3.57 %) (s. Tabelle 3-3). Die mittlere Sensitivität betrug dabei 24.52 % (± 5.45 %) und die mittlere Spezifität 86.15 % (± 8.01 %). Im Trainingsbereich konnte eine korrigierte Klassifikationsgenauigkeit von 79.13

% (± 15.26 %) erreicht werden. Das Lernverfahren mit der besten Performance war dabei ein Random Forest (maximale Tiefe = 4, minimale Anzahl an Samples pro Blatt = 10).

Auch in sämtlichen untersuchten Subgruppen – mit Ausnahme der Gruppe der Probanden mit mittlerem Bildungsgrad ($n = 240$), in der die korrigierte Klassifikationsgenauigkeit 46.29 % (± 12.42 %) betrug – war die korrigierte Klassifikationsgenauigkeit besser als der jeweilige Vergleichswert (Dummy). Die anhand dieser Metrik ermittelte Performancezunahme im Vergleich zum Dummy fiel mit 2 bis 7 % allerdings eher gering aus. Die höchste Genauigkeit konnte dabei mit 57.07 % (± 6.77 %) in der Gruppe der in Münster eingeschlossenen Probanden ($n = 374$) erreicht werden.

Tabelle 3-3: Klassifikationsanalyse: MDD vs. HC, Modalität: strukturelle MRT-Daten

	MDD ^a		HC ^a	Dummy Test		Test	Perf. ^b
	n	n	n	BAC M (SD)		BAC M (SD)	BAC _{test} - BAC _{dum.}
komplette Stichprobe	1047	432	615	0.50 (0.00)		0.55 (0.04)	0.05
Geschlecht							
männlich	375	153	222	0.50 (0.00)		0.56 (0.07)	0.06
weiblich	672	279	393	0.50 (0.00)		0.55 (0.03)	0.05
Alter							
> 55	99	51	48	0.50 (0.00)		0.52 (0.11)	0.02
≥ 30 und ≤ 55	413	201	212	0.50 (0.00)		0.56 (0.05)	0.06
< 30	535	180	355	0.50 (0.00)		0.52 (0.06)	0.02
höchster Bildungsgrad							
Lehre, Abi, Meister, Uni ^c	806	291	515	0.50 (0.00)		0.52 (0.04)	0.02
Schulabschluss ^d	240	140	100	0.50 (0.00)		0.46 (0.12)	-0.04
Studienzentrum							
Marburg	673	271	402	0.50 (0.00)		0.53 (0.04)	0.03
Münster	374	161	213	0.50 (0.00)		0.57 (0.07)	0.07

Bei den Metriken (M (SD)) handelt es sich um gemittelte Werte der entsprechenden Metriken aller 10 äußeren Faltungen. BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD) bzw. gesunde Kontrollen (engl.: healthy controls, HC), ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife

Die Durchführung eines Permutationstests (s. Abbildung 3-3) für die Analyse anhand der kompletten Stichprobe belegte die statistische Signifikanz der Ergebnisse. Bei zwei Permutationen wurde dabei eine gleichwertige oder bessere gewichtete

Klassifikationsgenauigkeit erreicht als bei dem Modell ohne permutierte Label (p -Wert = 0.002).

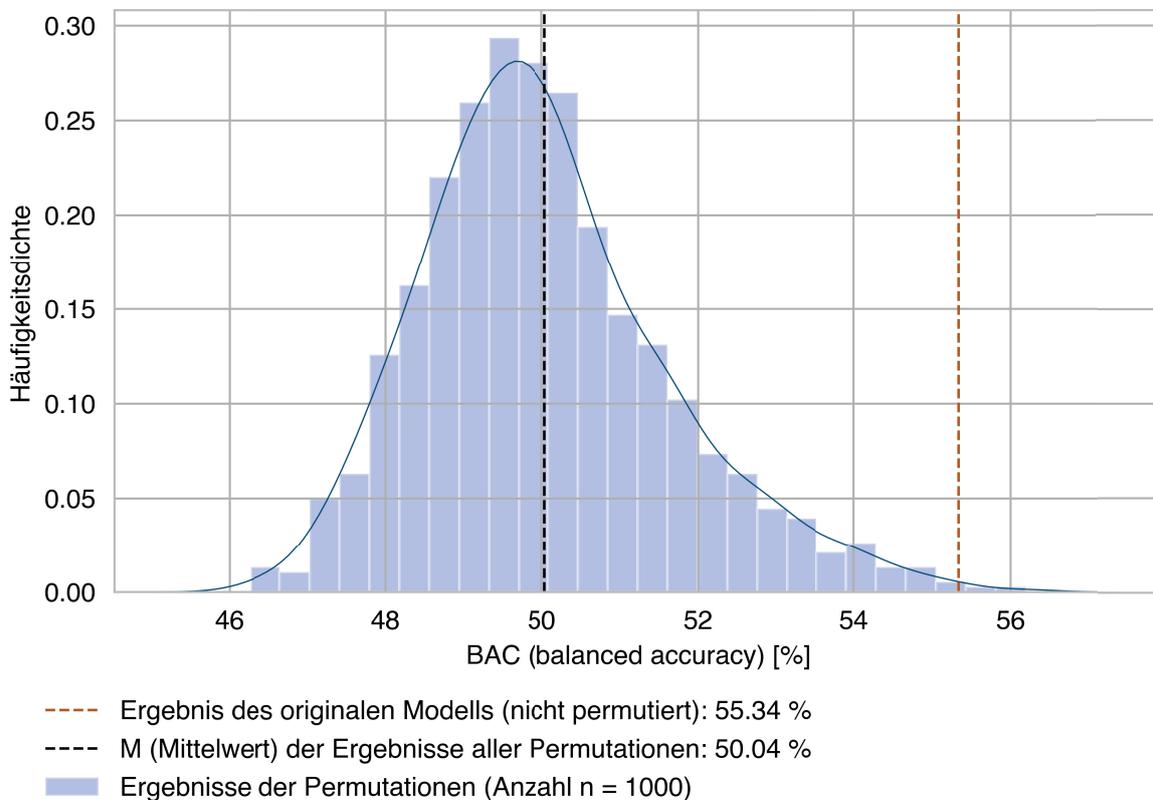


Abbildung 3-3: Permutationstest, Klassifikationsanalyse: MDD vs. HC, Modalität: strukturelle MRT-Daten

Anmerkungen: Zur Sicherung der statistischen Signifikanz der Ergebnisse (bei der Analyse der kompletten Stichprobe) wurde ein Permutationstest durchgeführt, wobei die Analyse insgesamt 1000 Mal mit jeweils zufällig permutierten Labels durchgeführt wurde (p -Wert = 0.002). HC = gesunde Kontrollen (engl.: healthy controls, HC), MDD = Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD)

3.3.2 Klassifikationsanalyse: rezidivierender Verlauf vs. einzelne Episode

Wie in Tabelle 8-12 im Anhang dargestellt, gelang die Unterscheidung der verschiedenen zeitlichen Krankheitsverläufe (rezidivierender Verlauf vs. einzelne Episode) innerhalb der kompletten MDD-Gruppe ($n = 432$) anhand der strukturellen MRT-Daten mit einer korrigierten Klassifikationsgenauigkeit von 55.08 % (± 7.45 %).

Innerhalb der Subgruppen unterschied sich die korrigierte Klassifikationsgenauigkeit teils fast gar nicht von den Vergleichswerten, die anhand des Dummys ermittelt wurden (z. B. in der Gruppe der Probanden, die in Marburg eingeschlossen wurden ($n = 271$), in der sie 49.82 % (± 6.40 %) betrug). Teils war die Klassifikationsgenauigkeit allerdings auch deutlich besser als

der entsprechende Vergleichswert. Am deutlichsten traf das auf die Gruppe der Probanden zu, die zum Zeitpunkt des Studieneinschlusses älter als 55 Jahre waren ($n = 51$). In dieser Gruppe konnte eine Genauigkeit von 71.25 % (± 21.69 %) erreicht werden, allerdings umfasste diese Gruppe nur rund 12 % der Probanden der gesamten Bezugsstichprobe. Durch die hohe Varianz der Ergebnisse innerhalb der Subgruppen konnte von keiner verlässlichen Klassifizierbarkeit der verschiedenen zeitlichen Krankheitsverläufe ausgegangen werden.

3.3.3 Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert

Innerhalb der gesamten MDD-Gruppe ($n = 432$) ließ sich der Hospitalisierungsstatus (Proband zum Zeitpunkt der Studienteilnahme in (teil-)stationärer Behandlung oder nicht) anhand der strukturellen MRT-Daten mit einer korrigierten Klassifikationsgenauigkeit von ca. 54.33 % (± 4.45 %) vorhersagen (s. Tabelle 8-13 im Anhang).

Die Analyse der Subgruppen zeigte allerdings, dass sich das nicht konsistent über alle Subgruppen reproduzieren ließ. Innerhalb der Subgruppen wurden teils schlechtere Ergebnisse als anhand des Dummys erreicht (z. B. in der Gruppe der Probanden, die in Marburg eingeschlossen wurden, $n = 271$, korrigierte Klassifikationsgenauigkeit: 46.63 % ± 6.15 %) und teils bis zu 10 % bessere Ergebnisse (z. B. in der Gruppe der Probanden, die zum Zeitpunkt des Studieneinschlusses eine akute Episode hatten, $n = 198$, korrigierte Klassifikationsgenauigkeit: 60.08 % ± 10.91 %). Durch die hohe Varianz innerhalb der Subgruppen, konnte auch hier nicht von einer stabilen Klassifikationsgüte ausgegangen werden.

3.3.4 Regressionsanalyse: HAMD-Wert

Die Prädiktion der Depressionsschwere (quantifiziert durch die Hamilton-Skala) anhand der strukturellen MRT-Daten ergab in der gesamten MDD-Gruppe ($n = 432$) einen kaum geringeren mittleren absoluten Fehler (ca. 1 % besser) als der anhand des Dummys ermittelte Vergleichswert (s. Tabelle 8-14 im Anhang). Die Pearson-Korrelation wich mit $r = -0.05$ (± 0.22) kaum von 0 ab. Somit wurde keine relevante Korrelation festgestellt.

Die Betrachtung der einzelnen Subgruppen ergab inkonsistente Zusammenhänge. Teilweise wurden negative (z. B. Subgruppe der männlichen Probanden, $n = 153$, $r = -0.22 \pm 0.15$) und positive Korrelationen (z. B. Subgruppe der Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-)stationärer Behandlung waren, $n = 130$, $r = 0.34 \pm 0.25$) ermittelt. Auch der

Betrag der mittleren absoluten Fehler der Prädiktionen fiel teils kleiner (z. B. Subgruppe der Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-)stationärer Behandlung waren, $n = 130$, ca. 7 % besser) und teils größer (z. B. Subgruppe der Probanden mit mind. einer psychiatrischen Komorbidität, $n = 85$, ca. 5 % schlechter) als die Vergleichswerte aus.

3.3.5 Regressionsanalyse: BDI-Wert

Ähnlich wie der HAMD-Wert konnte auch der BDI-Wert (und somit ein weiteres Maß der Depressionsschwere) bei der Analyse der gesamten MDD-Gruppe ($n = 432$) basierend auf den Prädiktoren kaum besser (bezogen auf den mittleren absoluten Fehler) als anhand des Dummys vorhergesagt werden (s. Tabelle 8-15 im Anhang). Für die komplette Stichprobe wurde eine geringe Korrelation zwischen den Prädiktoren und der Zielvariable festgestellt ($r = 0.01 (\pm 0.13)$).

Auch die Betrachtung der einzelnen Subgruppen ließ nicht auf einen stabilen Zusammenhang zwischen den strukturellen MRT-Daten und der Depressionsschwere schließen. Es wurden sowohl positive (z. B. Subgruppe der Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-)stationärer Behandlung waren, $n = 130$, $r = 0.15 \pm 0.23$) als auch negative (z. B. Subgruppe der Probanden in partieller Remission, $n = 109$, $r = -0.20 \pm 0.21$) Pearson-Korrelationen festgestellt. Teils konnten mittels des Modells geringere mittlere absolute Fehler als durch den Dummy erreicht werden (z. B. Subgruppe der Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-) stationärer Behandlung waren, $n = 130$, ca. 3 % besser), teils wurden höhere ermittelt (z. B. Subgruppe der Probanden in voller Remission, $n = 125$, ca. 2 % schlechter).

3.3.6 Regressionsanalyse: GAF-Wert

Wie in Tabelle 8-16 im Anhang dargestellt, lies sich GAF-Wert bei der Analyse der kompletten MDD-Stichprobe ($n = 238$) kaum anhand der strukturellen MRT-Daten prädizieren. Der mittlere absolute Fehler des Modells war nur geringfügig kleiner (ca. 1 % besser) als der Vergleichswert, wobei die Pearson-Korrelation bei $r = 0.00 (\pm 0.17)$ lag.

Auch in den Subgruppen konnte durch die hohe Varianz der Korrelationen von $r = -0.34 \pm 0.29$ (in der Gruppe der Probanden in partieller Remission, $n = 64$) bis zu $r = 0.53 \pm 0.26$ (in der Gruppe der Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-)stationärer Behandlung waren, $n = 70$) kein eindeutiger Zusammenhang festgestellt werden. Der mittlere absolute Fehler fiel in den Subgruppen ebenfalls teils geringer (z. B. Subgruppe der

Probanden, die zum Untersuchungszeitpunkt noch nie in (teil-)stationärer Behandlung waren, $n = 70$, ca. 16 % besser) und teils größer (z. B. Subgruppe der Probanden, die in Münster eingeschlossen wurden, $n = 47$, ca. 12 % schlechter) aus als die anhand des Dummys ermittelten Vergleichswerte.

3.3.7 Regressionsanalyse: SF-36-Wert (Subskala: körperliche Funktionsfähigkeit)

Die Prädiktion des SF-36-Wertes (Subskala: körperliche Funktionsfähigkeit) gelang für die gesamte MDD-Gruppe ($n = 432$) mit etwa 13 % besserer Genauigkeit als anhand des Dummys (s. Tabelle 3-4). Das Modell erreichte einen mittleren absoluten Fehler von 12.08 (± 0.67) Punkten, wobei der Vergleichswert 13.90 (± 0.17) Punkte betrug. Für die gesamte MDD-Stichprobe wurde eine Pearson-Korrelation von $r = 0.35$ (± 0.10) ermittelt. Innerhalb des Trainings konnte für die gesamte Stichprobe ein mittlerer absoluter Fehler von 8.62 (± 1.32) Punkten und eine Pearson-Korrelation von $r = 0.28$ (± 0.11) erreicht werden. Die optimale Maschine war eine SVM ($C = 10$, Kernel = rbf).

Mit Ausnahme der Subgruppe der Probanden, die zum Zeitpunkt der Studienteilnahme älter als 55 Jahre waren ($n = 51$), konnte mittels maschinellen Lernens in allen Subgruppen eine bessere Performance (bezogen auf den mittleren absoluten Fehler) als anhand des Dummys erreicht werden. Die Performanceverbesserung schwankte dabei zwischen 8 % und 15 %, wobei die größte Verbesserung in der Gruppe der Probanden, die in Münster eingeschlossen wurden ($n = 161$), erreicht wurde.

Auch die Pearson-Korrelation war mit Ausnahme der Subgruppe der Probanden, die zum Studienzeitpunkt jünger als 30 Jahre waren ($n = 180$, $r = -0.05$ (± 0.17)), durchweg positiv. Die stärkste Korrelation wurde mit $r = 0.48$ (± 0.25) innerhalb der Subgruppe der Probanden, die zum Zeitpunkt der Studienteilnahme keine psychiatrische Komorbidität aufwiesen, festgestellt.

Tabelle 3-4: Regressionsanalyse: SF-36-Wert, Modalität: strukturelle MRT-Daten

	n	SF-36		Dummy Test		Test		Perf. ^a	Test	
		M (SD)		MAE M (SD)		MAE M (SD)		$\frac{MAE_{test}}{MAE_{dum.}}$	r	M (SD)
komplette Stichprobe	432	85.10 (18.29)		13.90 (0.17)		12.08 (0.67)		0.87	0.35	(0.10)
Geschlecht										
männlich	153	87.55 (16.78)		12.42 (0.24)		10.88 (0.79)		0.88	0.33	(0.21)
weiblich	279	83.76 (18.96)		14.70 (0.36)		13.07 (0.90)		0.89	0.36	(0.20)
Alter										
> 55	51	78.33 (22.13)		17.26 (2.42)		17.63 (2.05)		1.02	0.16	(0.55)
≥ 30 und ≤ 55	201	81.47 (20.46)		16.46 (0.54)		14.89 (0.83)		0.90	0.25	(0.15)
< 30	180	91.08 (11.70)		8.64 (0.69)		7.66 (0.64)		0.89	-0.05	(0.17)
höchster Bildungsgrad										
Lehre, Abi, Meister, Uni ^b	291	87.63 (16.41)		12.23 (0.32)		10.61 (0.47)		0.87	0.24	(0.24)
Schulabschluss ^c	140	80.14 (20.64)		16.12 (0.68)		14.84 (1.33)		0.92	0.31	(0.16)
zeitl. Krankheitsverlauf										
rezidivierend	288	83.56 (18.96)		14.73 (0.38)		12.82 (0.98)		0.87	0.38	(0.14)
einzelne Episode	144	88.19 (16.50)		11.99 (0.77)		10.49 (0.92)		0.87	0.20	(0.19)
Hospitalisierungsgrad^d										
mind. einmal	302	82.86 (18.83)		14.83 (0.23)		13.35 (0.81)		0.90	0.30	(0.15)
noch nie	130	90.31 (15.85)		10.43 (1.45)		8.88 (1.47)		0.85	0.34	(0.28)
Hospitalisierungsstatus^d										
hospitalisiert	149	81.91 (18.71)		15.01 (0.41)		13.59 (0.86)		0.90	0.30	(0.22)
nicht hospitalisiert	283	86.78 (17.87)		13.12 (0.41)		11.02 (0.66)		0.84	0.38	(0.19)
Remissionsstatus										
akut	198	79.95 (20.75)		16.38 (0.41)		14.23 (1.04)		0.87	0.44	(0.16)
partielle Remission	109	85.37 (17.29)		13.40 (0.93)		12.03 (1.30)		0.90	0.19	(0.34)
volle Remission	125	93.04 (10.60)		6.98 (1.17)		6.22 (1.22)		0.89	0.23	(0.25)
Antidepressiva										
mind. eins	261	81.92 (19.80)		15.50 (0.39)		13.19 (1.01)		0.85	0.42	(0.16)
keins	171	89.97 (14.47)		10.57 (0.79)		9.29 (0.53)		0.88	0.12	(0.28)
Psych. Komorbidität										
mind. eine	85	80.88 (20.82)		16.53 (1.51)		14.96 (1.80)		0.90	0.30	(0.23)
keine	158	88.51 (14.45)		11.41 (0.37)		9.80 (1.14)		0.86	0.48	(0.25)
Studienzentrum										
Marburg	271	83.71 (19.18)		14.72 (0.46)		13.22 (0.72)		0.90	0.34	(0.11)
Münster	161	87.45 (16.47)		12.39 (0.71)		10.50 (0.97)		0.85	0.24	(0.21)

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, SF-36 = Short Form 36-Fragebogen (Subskala: körperliche Funktionsfähigkeit), ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlung

Die Überprüfung der Ergebnisse der Analyse der kompletten MDD-Stichprobe anhand eines Permutationstests (s. Abbildung 3-4) belegte die statistische Signifikanz. Dabei wurde für keine einzelne Permutation ein gleichwertiger oder niedrigerer absoluter Fehler als anhand des originalen Modells ermittelt (p-Wert = 0.00).

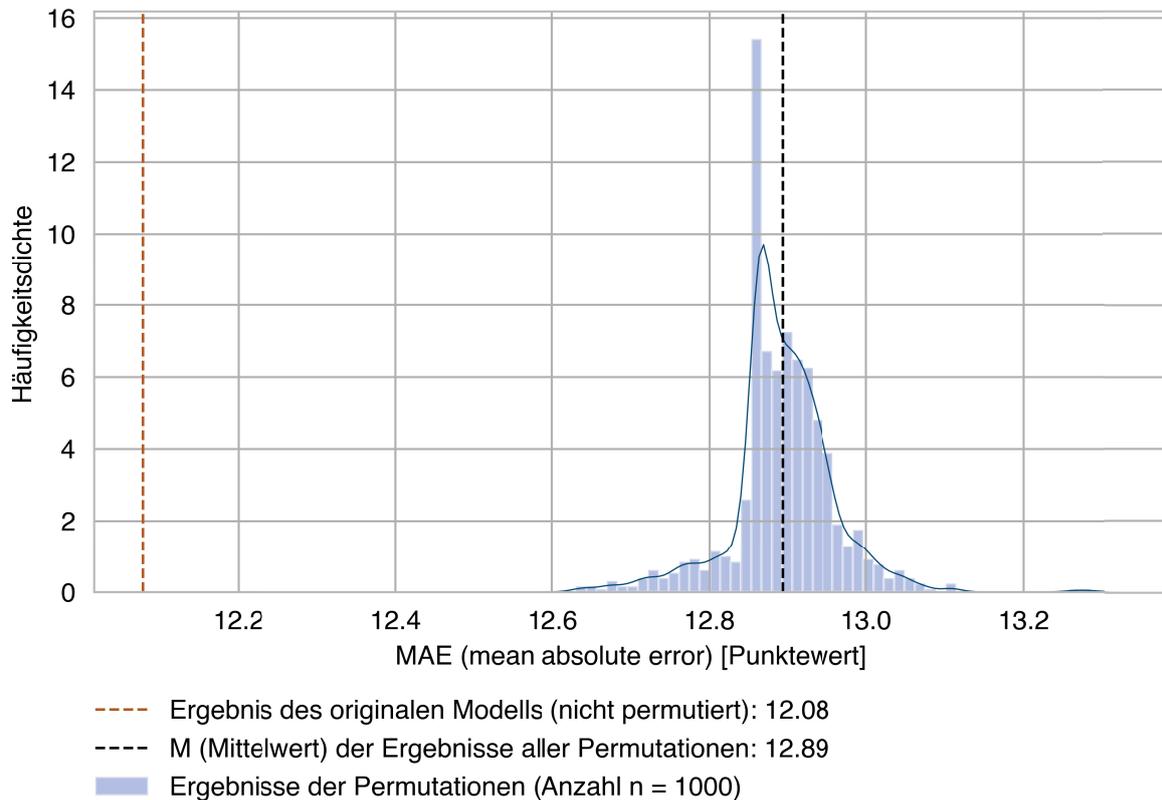


Abbildung 3-4: Permutationstest, Regressionsanalyse: SF-36-Wert (körperliche Funktionsfähigkeit), Modalität: strukturelle MRT-Daten

Anmerkungen: Zur Sicherung der statistischen Signifikanz der Ergebnisse (bei der Analyse der kompletten Stichprobe) wurde ein Permutationstest durchgeführt, wobei die Analyse insgesamt 1000 Mal mit jeweils zufällig permutierten Zielwerten durchgeführt wurde (p-Wert = 0.00). SF-36 = Short Form 36-Fragebogen

4 Diskussion

Die Ergebnisse unterschieden sich prinzipiell kaum zwischen den beiden Modalitäten, sodass sie hier nicht separat für die Modalitäten, sondern stattdessen in Bezug auf die Hypothesen (s. Kapitel 1.4) zusammengefasst (s. Kapitel 4.1) und diskutiert (s. Kapitel 4.2, 4.3) werden. Zuletzt werden die Methoden diskutiert (s. Kapitel 4.4 und 4.5).

4.1 Zusammenfassung der Ergebnisse

Klassifikationsanalysen: MDD vs. HC (1. Hypothese)

Die Gruppenunterscheidung (MDD-Patienten vs. gesunde Kontrollen) gelang anhand der Modalität der neuropsychologischen Testdaten ($n = 1650$) mit einer korrigierten Klassifikationsgenauigkeit von 59.62 % (± 4.06 %). Auch in den betrachteten Subgruppen ließ sich diese Klassifikation konsistent mit ca. 4 - 6 % besserer Genauigkeit als anhand des Dummys durchführen. Anhand der strukturellen MRT-Daten ($n = 1047$) konnte hinsichtlich der Gruppenunterscheidung eine korrigierte Klassifikationsgenauigkeit von 55.34 % (± 3.57 %) erreicht werden. Mit Ausnahme der Gruppe mit mittlerem Bildungsabschluss (Volks-, Haupt- und Realschulabschluss bzw. Fachhochschulreife) konnte in allen untersuchten Subgruppen ebenfalls eine höhere Genauigkeit (2 - 7 %) als anhand des Dummys erreicht werden.

Die in Kapitel 1.4 aufgestellte Hypothese, dass sich MDD-Patienten und gesunde Kontrollen anhand der kognitiven Leistung bzw. der Gehirnstruktur unterscheiden lassen, konnte also anhand der vorliegenden Stichprobe prinzipiell für die Modalität der neuropsychologischen Testung (s. Kapitel 3.2.1) und – mit Einschränkungen – auch für die Modalität der strukturellen MRT-Daten (s. Kapitel 3.3.1) bestätigt werden.

Analysen zur Prädiktion einzelner Krankheitscharakteristika (2. Hypothese)

Innerhalb der MDD-Stichprobe konnte anhand beider Modalitäten weder der zeitliche Krankheitsverlauf (rezidivierender Verlauf vs. einzelne Episode), der Hospitalisierungsstatus (Proband zum Zeitpunkt der Studienteilnahme in (teil-) stationärer Behandlung oder nicht), die Depressionsschwere (HAMD, BDI) noch das globale Funktionsniveau (GAF) prädiziert werden. Bezüglich der körperlichen Funktionsfähigkeit ließ sich allerdings ein stabiler Zusammenhang zwischen den Prädiktoren (beide Modalitäten) und der Zielvariable (SF-36 Subskala: körperliche Funktionsfähigkeit) feststellen.

In der gesamten MDD-Gruppe konnte der mittlere absolute Fehler der Prädiktion des SF-36 (Subskala: körperliche Funktionsfähigkeit) anhand der neuropsychologischen Testdaten ($n = 736$) um 9 % gegenüber dem Vergleichswert reduziert werden (s. Kapitel 3.2.7). Der durch das Modell ermittelte mittlere absolute Fehler betrug dabei 12.18 (± 0.39) Punkte, wohingegen der anhand des Dummys ermittelte Fehler bei 13.34 (± 0.14) Punkten lag.

Die Pearson-Korrelation betrug dabei $r = 0.24$ (± 0.14). In allen Subgruppen – mit Ausnahme der Probanden, die sich zum Zeitpunkt der Untersuchung in (teil-)stationärer Behandlung befanden – konnte das Modell eine um 4 - 13 % bessere Performance als der Dummy erreichen. Dabei wurden für alle Subgruppen positive Pearson-Korrelationen festgestellt.

Auch anhand der Modalität der strukturellen MRT-Daten ($n = 432$) konnte diese Zielvariable innerhalb der gesamten MDD-Stichprobe bei einer Pearson-Korrelation von $r = 0.35$ (± 0.10) mit einer ca. 13 % besseren Genauigkeit als durch den Dummy prädiziert werden (s. Kapitel 3.3.7). Das Modell erreichte einen mittleren absoluten Fehler von 12.08 (± 0.67) Punkten, wobei der Vergleichswert 13.90 (± 0.17) Punkte betrug.

Ähnliche Performanceverbesserungen gegenüber dem Dummy (8 - 15 %) konnten für alle Subgruppen mit Ausnahme der Probanden, die zum Zeitpunkt des Studieneinschlusses älter als 55 Jahre waren, erreicht werden. Abgesehen von der Gruppe der Probanden, die zum Studienzeitpunkt jünger als 30 Jahre waren, wurde für jede der betrachteten Subgruppen eine positive Pearson-Korrelation ermittelt.

Die in Kapitel 1.4 aufgestellte zweite Hypothese, dass sich einzelne Krankheitscharakteristika der MDD anhand der kognitiven Leistungsfähigkeit bzw. der Gehirnstruktur prädizieren lassen, konnte nach den explorativen Analysen in dieser Arbeit entsprechend nur für den SF-36-Wert (Subskala: körperliche Funktionsfähigkeit) bestätigt werden.

4.2 Diskussion der Klassifikationsanalysen: MDD vs. HC

Im Folgenden werden die Ergebnisse der Analysen zur Unterscheidung von MDD-Patienten und gesunden Kontrollen anhand der betrachteten Modalitäten diskutiert. Dass die Gruppenunterscheidung anhand der Prädiktoren besser als der Zufall funktioniert hat, ist konsistent mit dem aktuellen Stand der Forschung (s. Kapitel 1.2.1 und 1.2.2). Grundsätzlich stellt sich jedoch die Frage, inwieweit die erreichten Modellgüten (maximal wurde eine

gewichtete Klassifikationsgenauigkeit von ca. 60 % erreicht) überhaupt klinisch relevant sind. Einem erfahrenen klinischen Mitarbeiter wird die entsprechende Unterscheidung sicherlich auch in kürzerer Zeit (ohne Erhebung der neuropsychologischen Testdaten bzw. der strukturellen MRT-Daten) mit höherer Präzision gelingen.

Ausprägung der kognitiven Defizite/Gehirnveränderungen bei MDD-Patienten

Wie in Kapitel 1.2.1 beschrieben, wird in der Literatur kontrovers diskutiert, wie stark die kognitive Leistung bei der Erkrankung mit einer MDD tatsächlich beeinträchtigt ist und inwieweit diese Beeinträchtigung auf die Erkrankung an sich oder aber Begleiteffekte wie bspw. motivationale Aspekte zurückzuführen ist.

Passend zu der in dieser Arbeit festgestellten geringen Trennschärfe zwischen MDD-Patienten und gesunden Kontrollen wird bspw. in einer Studie beschrieben, dass nur 25 % der dort eingeschlossenen MDD-Patienten innerhalb der neuropsychologischen Testung Werte von \geq einer Standardabweichung unterhalb des Mittelwertes der gesunden Referenzpopulation aufwiesen (159). Ein möglicher Erklärungsansatz für die hier ermittelten geringen Klassifikationsgenauigkeiten ist dementsprechend, dass die kognitiven Leistungseinschränkungen bei MDD-Patienten tatsächlich gering ausgeprägt sind.

Die geringe Klassifikationsgenauigkeit anhand der strukturellen MRT-Daten (s. Kapitel 3.3.1) steht in Einklang mit den Erkenntnissen in der aktuellen Literatur. Wie in Kapitel 1.2.2 beschrieben, können bei einer MDD zwar morphologische Veränderungen bestehen, deren genauen Intensitäten und Lokalisationen werden allerdings weiterhin kontrovers diskutiert.

Einschränkungen aufgrund des Studiendesigns

Aufgrund des Studiendesigns muss beachtet werden, dass innerhalb der MDD-Gruppe nicht nur Probanden mit einer akuten depressiven Episode eingeschlossen wurden, sondern auch solche, die zum Zeitpunkt der Untersuchung eine teil- oder vollremittierte MDD aufwiesen.

Wie in Kapitel 1.2.1 beschrieben, ist unklar, ob und in welcher Intensität kognitive Defizite auch über eine akute depressive Phase hinaus bestehen. Falls die Defizite nur im akuten Zeitraum bestehen, wäre eine Klassifikation im Remissionsstadium anhand der kognitiven Leistung nicht mehr möglich. Auch der Zusammenhang von hirmorphologischen Veränderungen und dem Remissionsstadium ist, wie in Kapitel 1.2.2 beschrieben, unklar.

Demnach manifestieren sich morphologische Veränderungen vor allem bei rezidivierenden Verläufen. Auch das wurde in den obigen Analysen nicht unterschieden.

Eine Ad-hoc-Analyse (s. Tabelle 4-1) unter Einbezug der hier angesprochenen Subgruppen ergab allerdings keine höheren Klassifikationsgenauigkeiten als in den ursprünglichen Analysen (zum Vergleich in der Tabelle ebenfalls vermerkt).

Tabelle 4-1: Klassifikationsanalyse: MDD vs. HC, beide Modalitäten, verschiedene MDD Subgruppen

	MDD ^a	HC ^a	Dummy Test	Test
	n	n	BAC M (SD)	BAC M (SD)
Modalität: neuropsychologische Testung				
gesamte MDD-Gruppe	736	914	0.50 (0.00)	0.60 (0.04)
Remissionsstatus: akut ^b	312	914	0.50 (0.00)	0.57 (0.03)
derzeitig in (teil-)stationärer Behandlung ^b	218	914	0.50 (0.00)	0.55 (0.05)
rezidivierender Krankheitsverlauf ^b	474	914	0.50 (0.00)	0.56 (0.02)
Modalität strukturelle MRT-Daten				
gesamte MDD-Gruppe	432	615	0.50 (0.00)	0.55 (0.04)
Remissionsstatus: akut ^b	198	615	0.50 (0.00)	0.51 (0.04)
derzeitig in (teil-)stationärer Behandlung ^b	149	615	0.50 (0.00)	0.51 (0.04)
rezidivierender Krankheitsverlauf	288	615	0.50 (0.00)	0.53 (0.05)

Bei den Metriken (M (SD)) handelt es sich um gemittelte Werte der entsprechenden Metriken aller 10 äußeren Faltungen. BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden mit der Diagnose einer MDD (engl.: major depressive disorder, MDD) bzw. gesunde Kontrollen (engl.: healthy controls, HC), ^bDie Krankheitscharakteristika beziehen sich jeweils nur auf die Auswahl der Probanden aus der jeweiligen MDD-Gruppe. Die gesunde Kontrollgruppe blieb jeweils in ihrem vollen Umfang bestehen.

Ein möglicher Grund dafür könnte sein, dass die resultierenden MDD Subgruppen in dieser Analyse im Verhältnis zu den gesunden Kontrollgruppen deutlich weniger Probanden enthielten als die beiden ursprünglichen MDD-Stichproben. Dadurch waren die Gruppenverhältnisse weniger ausgeglichen und es standen in der Trainings- und Testphase weniger Daten zur Verfügung.

Aufgrund der Rekrutierung freiwilliger Probanden basierend auf Zeitungsanzeigen und der Ansprache auf Station nach entsprechender Vorauswahl durch die zuständigen Stationsärzte kann zudem davon ausgegangen werden, dass schwerstdepressive Patienten nicht ausreichend in der Studie repräsentiert waren. Diese Patientengruppe war vermutlich von

vorneherein nicht in der Lage, an der Studie teilzunehmen bzw. konnte nicht von einer Teilnahme überzeugt werden. Zur Verdeutlichung dieses Umstandes ist die rechtsschiefe Verteilung (Ergebnis des Signifikanztests der Schiefe: $p < 5\%$ für beide Modalitäten) der Depressionsschwere (HAMD- und BDI-Werte als Äquivalenz) innerhalb der gesamten MDD-Stichprobe in Abbildung 4-1 dargestellt.

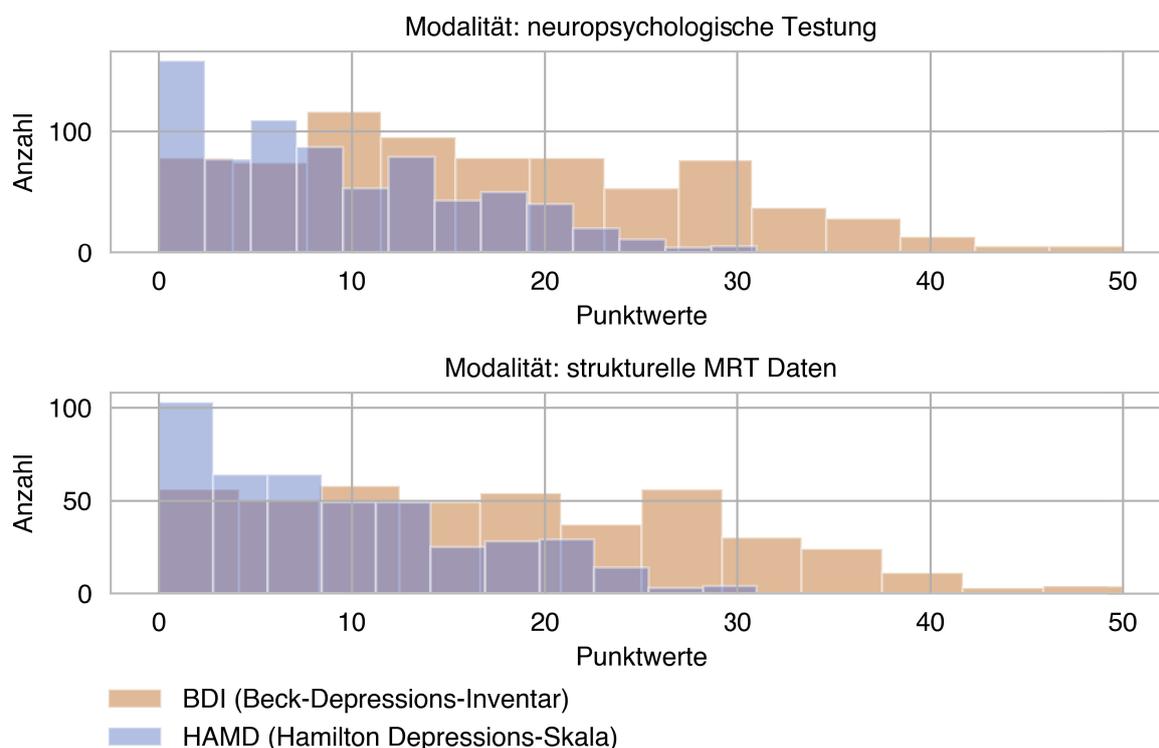


Abbildung 4-1: Verteilung der HAMD- und BDI-Werte für beide Modalitäten

Anmerkungen: Die Zielvariablen (HAMD- und BDI-Werte) wiesen jeweils eine rechtsschiefe Verteilung auf (getrennte Darstellung für beide Modalitäten).

Die wahre Verteilung der Depressionsschwere innerhalb der Population ist unbekannt. Wie in Kapitel 3.1 beschrieben, war die Depressionsschwere der eingeschlossenen MDD-Patienten mild bis mittelschwer. Innerhalb der Subgruppe der Probanden mit einer akuten depressiven Episode waren die HAMD- und BDI-Werte zwar etwas höher als für die gesamte MDD-Gruppe (s. Kapitel 2.2), entsprachen qualitativ gesehen im Mittel aber ebenfalls einer mittelschweren bzw. moderaten Depression.

Da die Analysen in dieser Arbeit vermuten lassen, dass keine direkte Korrelation zwischen der Depressionsschwere und der kognitiven Leistung bzw. strukturellen Gehirnveränderungen besteht (s. Kapitel 3), ist dieser Umstand vermutlich sekundär. Es wäre allerdings möglich,

dass die Analysen durch den Einschluss von mehr schwer depressiven Probanden andere Ergebnisse ergeben hätten.

In Kapitel 1.2.1 wurde darüber hinaus diskutiert, inwieweit bei dem Vorliegen einer MDD einzelne kognitive Funktionsbereiche singular eingeschränkt sind. Diesbezüglich wurde in dieser Arbeit keine Unterscheidung getroffen, sondern wie in Kapitel 2.1.4 beschrieben eine breite Palette an Funktionsbereichen untersucht und die Resultate gesammelt als Prädiktoren genutzt.

Subgruppenanalysen

Auffällig ist, dass die Prädiktion der Gruppenzugehörigkeit in der Subgruppe der Probanden mit Volks-, Haupt- oder Realschulabschluss bzw. Fachhochschulreife anhand der neuropsychologischen Testdaten im Vergleich zu den anderen Gruppen deutlich schlechtere gewichtete Genauigkeiten ergab (53.73 % (\pm 5.44 %)) und anhand der strukturellen MRT-Daten ein schlechteres Ergebnis als anhand des Dummys erreicht wurde (46.29 % (\pm 12.42 %)). Die beiden entsprechenden Subgruppen enthielten ca. 23-24 % der Probanden der Gesamtstichproben der jeweiligen Modalität und zudem mehr MDD-Probanden als gesunde Kontrollen, wobei es gegenüber den Gesamtstichproben näherungsweise zu einer Umkehr der Gruppenverhältnisse kam (ca. 1.4:1 statt 0.7-0.8:1 in der Gesamtgruppe).

In der Literatur wird beschrieben, dass das Bildungsniveau einen signifikanten Einfluss auf die kognitive Leistung (160,161) und die Gehirnstruktur (162,163) haben kann. Diese Einflussfaktoren, die kleineren Gruppengrößen und die veränderten Gruppenverhältnisse könnten ein Grund dafür gewesen sein, dass die Gruppenunterschiede in diesen Subgruppen nicht so ausgeprägt waren wie in den jeweiligen Gesamtstichproben.

Fazit

Dass die Unterscheidung von MDD-Patienten und gesunden Kontrollen anhand der betrachteten Modalitäten besser als anhand des Dummys funktioniert hat, weist auf einen Gruppenunterschied hin. Die weiteren Analysen hinsichtlich der Krankheitscharakteristika innerhalb der MDD-Gruppe hatten dadurch eine wissenschaftlich begründbare Basis. Es ist unklar, ob die erreichten Klassifikationsgenauigkeiten so gering waren, weil der Unterschied (hinsichtlich der Modalitäten) zwischen den Gruppen tatsächlich relativ gering war oder ob dies bspw. an der Stichprobe lag, in der schwer depressive Probanden unterrepräsentiert waren und auch teil- bzw. vollremittierte MDD-Probanden eingeschlossen wurden.

4.3 Diskussion der Prädizierbarkeit einzelner Krankheitscharakteristika

In diesem Kapitel wird diskutiert, warum sich innerhalb der MDD-Gruppen – mit Ausnahme des SF-36-Wertes (Subskala: körperliche Funktionsfähigkeit) – keine der anderen Krankheitscharakteristika anhand der betrachteten Prädiktoren vorhersagen ließen. Bezüglich des SF-36-Wertes ist fraglich, inwiefern die geringe Verbesserung der Prädiktion durch das Modell gegenüber der Vergleichsmetrik (dem Dummy) überhaupt klinisch relevant ist.

Die Struktur der Hamilton-Skala, des BDI und der GAF-Skala

Zunächst erfolgt eine Analyse von Gemeinsamkeiten und Unterschieden der psychometrischen Daten (und somit der Zielvariablen), die im Rahmen der Regressionsanalysen nicht gut prädiziert werden konnten. Dabei wird deutlich, dass sich die in den Fragebögen (HAMD, BDI, GAF) inkludierten Items sinnvollerweise mit den Diagnosekriterien einer MDD überschneiden (entsprechend DSM IV).

Wie in Kapitel 1.1.1 beschrieben, sind die beiden Hauptkriterien für die Diagnose einer MDD die gedrückte Stimmung über die meiste Zeit des Tages und der Verlust von Interesse bzw. Freude an annähernd allen Aktivitäten über die meiste Zeit des Tages. Die Nebenkriterien beinhalten bspw. Gewichtsveränderungen, Schlafstörungen, Energielosigkeit, Suizidgedanken, Entscheidungsschwierigkeiten und motorische Verlangsamung.

Innerhalb der Hamilton-Skala werden genau diese typischen MDD-Symptome (bspw. Stimmung, Appetit, Schlafveränderungen, Energieverlust usw.) explizit abgefragt. Auch im BDI, das zwar nicht zur Diagnose einer Depression genutzt werden soll, sondern die Schwere quantifiziert, werden hauptsächlich diese Symptomkomplexe abgefragt (z. B. Gewichtsveränderungen, erhöhte Müdigkeit, sozialer Rückzug usw.). Die GAF-Skala ist an das DSM IV angelehnt und weist dementsprechend ähnliche Überschneidungen auf (s. Kapitel 2.1.3).

Die Struktur des Short Form 36 Fragebogens

Im SF-36 werden hingegen weniger typische MDD Symptome, sondern alltägliche körperliche Funktionen abgefragt, wie z. B. das Steigen von einem Treppenabsatz oder das Heben und Tragen von Einkaufstaschen (s. Kapitel 2.1.3). Bei einem Durchschnittsalter von 36.42 (\pm 13.00) bzw. 37.13 (\pm 13.25) in den beiden für die Analyse des SF-36 relevanten MDD-Stichproben (s. Kapitel 2.2.1 und 2.2.2) kann tendenziell davon ausgegangen werden, dass die Kohorte rein physisch noch in einer guten Verfassung war.

Die Vor-Ort-Durchführung der Studie setzte zudem eine gewisse Mobilität voraus. Der recht hohe Durchschnittswert des SF-36-Wertes (Subskala: körperliche Funktionsfähigkeit) von 85.93 (\pm 17.91) bzw. 85.84 (\pm 17.28) in der MDD-Stichprobe bestätigt die Vermutung. Auffällig ist auch, dass der Wert in der entsprechenden gesunden Kontrollgruppe rund 10 Punkte höher ausfiel (s. Kapitel 3.1).

Fazit der Analyse der Fragebogen-Struktur

Eine mögliche Erklärung des Zusammenhangs zwischen dem SF-36-Wert und der kognitiven Leistung könnte darin liegen, dass sowohl die in dem SF-36 abgefragte eingeschränkte Alltagsfunktion als auch die veränderte kognitive Leistungsfähigkeit, wie sie durch die neuropsychologische Testung festgestellt wurde, nicht an sich mit der Symptomatik der MDD zusammenhängen, sondern mit einer (teils daraus resultierenden) geringeren Motivation bzw. mit einer tendenziell eher negativen Einstellung zum Testen/Alltagsleben begründet werden müssen. Das würde auch erklären, warum sich die Werte der symptom-basierten Fragebögen (HAMD, BDI, GAF) kaum anhand der Ergebnisse der neuropsychologischen Testung präzisieren ließen.

Die Struktur der Zielvariablen der Klassifikationsanalysen

Die beiden innerhalb der MDD-Gruppe durchgeführten Klassifikationsanalysen zur Unterscheidung von verschiedenen zeitlichen Krankheitsverläufen (rezidivierend vs. nicht rezidivierend) bzw. des Hospitalisierungsstatus (zum Zeitpunkt der Untersuchung in (teil-)stationärer Behandlung oder nicht), waren in diesem Zusammenhang „harte“ klinische Marker und gaben wenig Einblick in eventuelle motivationale Prozesse.

Legt man obenstehende Theorie zu Grunde, ergibt es Sinn, dass auch in diesen Analysen keine bessere Klassifikationsgenauigkeiten als anhand des Dummys erreicht wurden. Zudem war bei Probanden, die zum Zeitpunkt der Studienteilnahme hospitalisiert waren, unklar, in welcher Phase der Erkrankung sie sich befanden, also ob bspw. eine akute depressive Episode vorlag oder Probanden kurz vor einer Entlassung standen. Diese Information, wie sie bspw. durch die Hamilton-Skala erhoben wurde, fand hier keine Beachtung.

Die Rolle der Motivation bei der MDD

In der Literatur ist die Rolle der Motivation hinsichtlich der Leistung in der neuropsychologischen Testung ebenfalls beschrieben. So berichteten MDD-Probanden in einer Studie von größeren Ängsten vor schlechten Ergebnissen, beschwerten sich vermehrt

über negative äußere Einflüsse in der Testsituation und waren weniger motiviert als die gesunde Kontrollgruppe, wie eine retrospektive Befragung ergab (159). Entsprechend der zitierten Studie, ist die verminderte Performance von MDD-Probanden in der neuropsychologischen Testung nicht notwendigerweise ein Korrelat einer ebenfalls verringerten Funktionsfähigkeit in den Hirnarealen, die diese kognitive Funktionen repräsentieren, sondern vor allem durch andere Kovariaten beeinflusst (159).

Im Rahmen einer weiteren Studie aus dem Forschungsbereich der Schizophrenie, wurde eine robuste Korrelation zwischen der Motivation und der kognitiven Performance festgestellt (164). In einer anderen wissenschaftlichen Arbeit wird zudem beschrieben, dass Fehler bei dem Lösen einer Aufgabe innerhalb der neuropsychologischen Testung bei MDD-Probanden im Vergleich zu gesunden Kontrollprobanden mit einer statistisch signifikant höheren Wahrscheinlichkeit zu weiteren Fehlern bzw. zur Unfähigkeit beim Lösen der nachfolgenden Aufgaben geführt haben (165).

Ein Zusammenhang zwischen negativem Affekt – wie es bei der MDD der Fall ist (s. Kapitel 1.1.1) – und geringer Motivation wird in der Literatur ebenfalls beschrieben (41). Depressive Patienten reagieren demnach weniger stark auf positive Belohnungsreize, woraus eine verringerte Motivation resultieren kann (41).

Die Rolle der Motivation in einer Testsituation

In der Literatur sind verschiedenartige Ursachen für die Motivation zur Teilnahme an einer Testsituation beschrieben. So wird u. a. zwischen den Begriffen „Optimizing“ und „Satisficing“ unterschieden, wobei bei ersterem ein positiver Grund zur Teilnahme vorliegt (bspw. monetäre Gründe, das Verantwortungsbewusstsein als Studienteilnehmer oder die Erfüllung des eigenen Selbstbildes), der zu einer gründlichen Bearbeitung der Aufgaben führt (97,166).

„Satisficing“ hingegen tritt eher ein, je schwieriger die Aufgaben sind, je geringer die Leistungsfähigkeit des Probanden ist und je geringer die Motivation zur Studienteilnahme ist (bspw. aufgrund einer verpflichtenden Teilnahme) und führt zu einer oberflächlichen Beantwortung der Aufgaben oder zur Angabe von formal adäquaten Antworten, die zwar sinnvoll in den Kontext passen, aber gar nicht mehr die eigentliche Einstellung des Probanden widerspiegeln (97,166).

Ein weiteres Augenmerk muss auf die Gesamtdauer der neuropsychologischen Testung gelegt werden, da eine subjektiv als zu lang empfundene Testdauer zu einer motivational bedingten Senkung der Bearbeitungsqualität führen kann (97). Wie in Kapitel 1.1.1 beschrieben, ist eines der möglichen Symptome der MDD eine verminderte Konzentrationsfähigkeit, sodass dieser Aspekt vor allem bei depressiven Patienten ins Gewicht fallen könnte.

Strukturellen Gehirnveränderungen und die Rolle der Motivation

Im Widerspruch zu der Hypothese, dass die kognitiven Einschränkungen bei MDD-Probanden vor allem motivational bedingt sind, steht, dass der SF-36-Wert (Subskala: körperliche Funktionsfähigkeit) auch anhand der strukturellen MRT-Daten ähnlich gut prädiziert werden konnte (s. Kapitel 3.2.7 und 3.3.7). Eine mögliche Erklärung dafür wäre, dass die hirnmorphologischen Veränderungen in Bereichen liegen, die mit der Motivation assoziiert sind bzw. die den Affekt regulieren. Wie zuvor beschrieben, besteht ein Zusammenhang zwischen dem Affekt und der Motivation. Das würde ebenfalls mit dem in Kapitel 1.2.2 vorgestellten Stand der Forschung übereinstimmen, der besagt, dass morphologische Veränderungen vor allem in Bereichen der emotionalen Verarbeitung auftreten.

Im Einklang mit dieser Theorie stehen Studienergebnisse, in denen die Unterscheidung von MDD-Probanden und gesunden Kontrollen anhand von aufgabenbasierten funktionellen MRT-Daten (vor allem zur Quantifizierung emotionaler Verarbeitungsprozess, wie bspw. innerhalb der Amygdala) mit vergleichsweise hohen Genauigkeiten gelang (24,167). Ein solcher Rückschluss auf funktionelle Kausalitäten ist basierend auf den prädiktiven Modellen dieser Arbeit jedoch nicht möglich und erfordert weitere Studien zur Bestimmung der Wirkungsbeziehungen.

Eine kritische Betrachtung der Auswertbarkeit des SF-36

Eine zu dem oben beschriebenen Ansatz alternative Erklärung wäre, dass die bessere Prädizierbarkeit der SF-36-Wertes durch die Bewertung des Fragebogens bedingt ist. Wie in Kapitel 3.1 erläutert, wies der SF-36-Wert (Subskala: körperliche Funktionsfähigkeit) in dieser Studie eine im Vergleich zu den anderen psychometrischen Daten (HAMD, BDI, GAF) deutlich höhere Standardabweichung auf. Wie in Kapitel 2.1.3 beschrieben, wird der Punktwert der Subskala der körperlichen Funktionsfähigkeit grundsätzlich anhand von 10 Items bestimmt, welche vom Probanden jeweils auf einer Skala von 1-3 bewertet werden. Der daraus gebildete

Punktwert $\in [0;100]$ ist also an sich kein echter kontinuierlicher Wert. Auch der Punktwert der Hamilton-Skala und des BDI werden auf der Grundlage von einzelnen Items berechnet, allerdings sind es dort jeweils 21 statt 10 Items und das resultierende Intervall der möglichen Summenwerte ist zudem jeweils geringer.

Die Verteilung der Zielvariablen

Wie schon in Kapitel 4.2 für die HAMD- und BDI-Werte beschrieben, waren auch die GAF- und SF-36-Werte schief verteilt (Ergebnis des Signifikanztests der Schiefe: $p < 5\%$, mit Ausnahme des GAF-Wertes für die Modalität: strukturelle MRT-Daten) (s. Abbildung 4-2). Die Verteilungen der SF-36-Werte (Subskala: körperliche Funktionsfähigkeit) wiesen dabei einen deutlichen Schwerpunkt auf hohen Werten auf. Wie schon in den Kapiteln 2.2, 3.1 und 4.2 beschrieben, waren die Probanden im Durchschnitt mittelschwer depressiv und wiesen eine mittlere bis gute Funktionsfähigkeit auf. Da die einzelnen Parameter der Modelle anhand der Trainingsdaten automatisch erlernt wurden, bildet sich dieser Bias, der in den Daten vorhanden war, auch in den Machine Learning-Modellen ab und kann zu Verzerrungen der Ergebnisse führen (17).

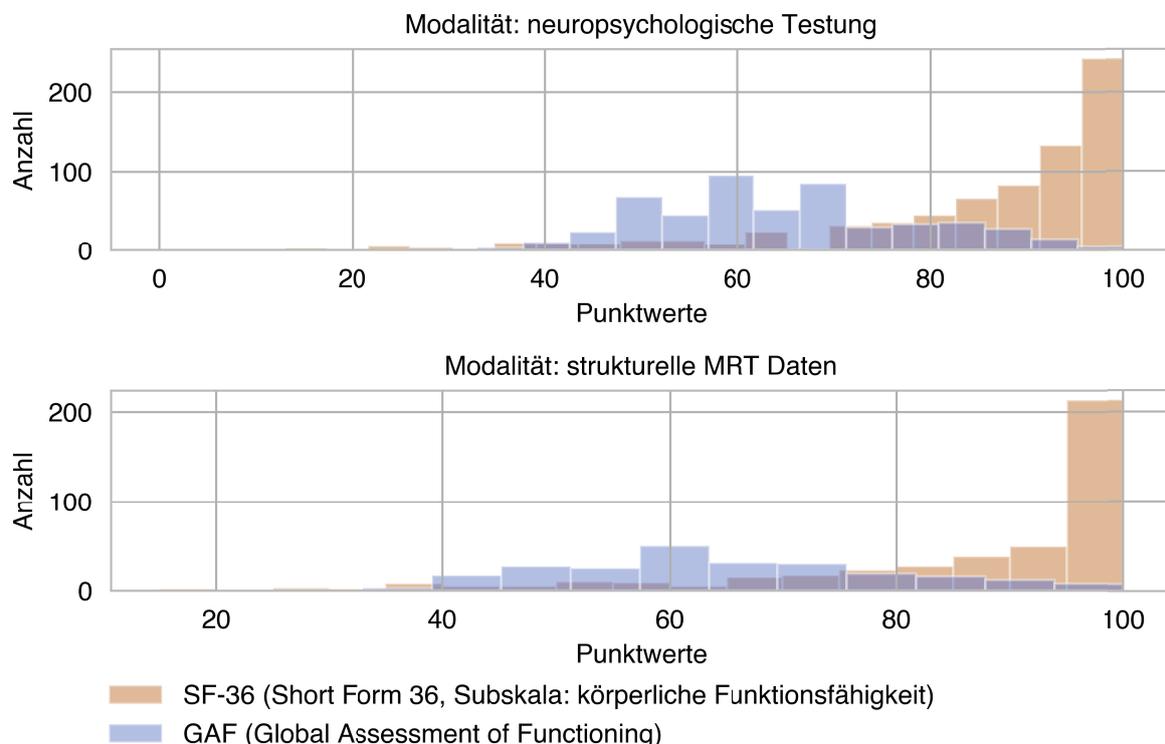


Abbildung 4-2: Verteilung der GAF- und SF-36-Werte für beide Modalitäten

Anmerkungen: Die Zielvariablen (GAF- und SF-36-Werte) wiesen, mit Ausnahme des GAF-Wertes für die Modalität der strukturellen MRT-Daten, jeweils eine schiefe Verteilung auf (getrennte Darstellung für beide Modalitäten).

Subgruppenanalysen

Bei den Analysen zur Prädiktion des SF-36-Wertes anhand der neuropsychologischen Testdaten wurde innerhalb der Subgruppe der Probanden, die sich zum Zeitpunkt der Studienteilnahme in (teil-)stationärer Behandlung befanden, lediglich eine Performanceverbesserung von ca. 1 % gegenüber dem Dummy erreicht. Die Subgruppe umfasste ca. 30 % der entsprechenden Gesamtstichprobe bei einem ca. 5 % niedrigeren SF-36-Durchschnittswert. Der Grund für die vergleichsweise niedrige Performance innerhalb dieser Subgruppe bleibt offen und erfordert weitere Analysen. Da sich alle Probanden dieser Subgruppe in (teil-)stationärer Behandlung und somit in einer ähnlichen Ausgangslage befanden, könnte die Varianz der Prädiktoren in dieser Subgruppe im Vergleich zu der Gesamtstichprobe u. U. niedriger gewesen sein.

Anhand der strukturellen MRT-Daten wurde bei der Prädiktion des SF-36-Wertes innerhalb der Subgruppe der Probanden, die zum Zeitpunkt der Studienteilnahme > 55 Jahre alt waren, ein schlechteres Ergebnis als durch den Dummy erreicht. In der Subgruppe der Probanden, die < 30 Jahre alt waren, wurde eine negative Pearson-Korrelation ermittelt. Die Gruppe der > 55-Jährigen umfasste nur ca. 12 % der Probanden (n = 51) der gesamten Stichprobe und hatte bei einem k von 10 in der äußeren Kreuzvalidierung u. U. keinen adäquaten Stichprobenumfang mehr, was ein Grund für die Abweichungen darstellen könnte. Warum in der Gruppe der < 30-jährigen Probanden eine negative Pearson-Korrelation ermittelt wurde, bleibt offen. Auffällig ist, dass der SF-36-Durchschnittswert in dieser Subgruppe vergleichsweise hoch war.

In der Literatur wird beschrieben, dass MDD-Probanden in Abhängigkeit vom Alter unterschiedliche strukturelle Gehirnveränderungen aufweisen können (s. Kapitel 1.2.2) (52) und dass der SF-36-Wert (Subskala: körperliche Funktionsfähigkeit) bei älteren Probanden mit dem Alter korreliert sein kann (168). Beide Effekte stellen potentielle Gründe für die hier ermittelte unterschiedliche Güte der Prädiktion des SF-36-Wertes in den unterschiedlichen Altersgruppen dar.

Fazit

Wie in Kapitel 1.1.3 und 1.1.4 beschrieben, ist die Ätiologie der MDD noch nicht abschließend geklärt. Wie in Kapitel 1.2 beschrieben, wird die Ausprägung und die Ursache der kognitiven Einschränkungen und der strukturellen Gehirnveränderungen bei MDD-Probanden kontrovers diskutiert. Der hier diskutierte motivationale Aspekt als (Mit-)Ursache von kognitiven

Einschränkungen kann als weiterer alternativer Erklärungsansatz betrachtet werden. Da im Rahmen dieser Arbeit die Motivation der Probanden nicht explizit abgefragt wurde, muss dieser Ansatz in weiteren Studien anhand eines abgeänderten Studiendesigns validiert werden.

4.4 Diskussion der Methodik des maschinellen Lernens

In diesem Kapitel wird die angewandte Methodik in Bezug auf die in Kapitel 1.3.4 vorgestellten Best Practices für Projekte des maschinellen Lernens knapp diskutiert.

Umsetzung der einzelnen durch Experten empfohlenen Schritte

Wie eingangs erläutert (s. Kapitel 1.4), handelt es sich hierbei um eine explorative Arbeit mit dem grundlegenden Ziel, den prädiktiven Wert zweier einzelner klinischer Modalitäten hinsichtlich verschiedener Krankheitscharakteristiken von MDD-Patienten zu analysieren. Der finale Entwicklungsschritt zur Bewertung der klinischen Anwendbarkeit der Modelle entfiel aufgrund dieser experimentellen Natur der Arbeit.

Wie empfohlen, wurde zunächst ein Analyseziel festgelegt, indem im Rahmen einer Literaturrecherche (s. Kapitel 1.1 und 1.2) geprüft wurde, welche ursächlichen Mechanismen bei der MDD weiterhin ungeklärt sind. Durch die Prüfung kontrovers diskutierter Zusammenhänge, sollte ein Beitrag zum Fortschritt der Forschung geleistet werden. Als Zielpopulation kann aufgrund der vorliegenden Daten nur die gelten, die durch die Stichprobe adäquat repräsentiert wird. Zur Diversifizierung wurden die Daten von zwei verschiedenen Standorten und Einzugsbereichen genutzt.

Die Auswahl der Modalitäten basierte auf der Literaturrecherche und der klinischen Realisierbarkeit der Datenerhebung. Im Hinblick auf die in der Literatur beschriebenen vielversprechenden Ergebnisse multimodaler Modelle (s. Kapitel 1.3.2) bleibt offen, ob eine Kombination beider Modalitäten die Genauigkeit der Modelle verbessert hätte. Die getrennte Betrachtung beider Modalitäten wurde in dieser Arbeit explizit und bewusst zur Identifikation einzelner Zusammenhänge gewählt.

Im Rahmen des technischen Modellentwurfes erfolgte ebenfalls eine ausführliche Literaturrecherche (s. Kapitel 2.3). Um eine Überanpassung der Modelle an die Trainingsdaten zu vermeiden und eine gewisse Generalisierbarkeit zu ermöglichen, wurde ein breiter

Hyperparameterraum gewählt. Es wurden verschiedene einschlägige Lernverfahren und die entsprechenden empfohlenen Transformatoren eingesetzt. Datenlecks wurden aufgrund der Pipeline-Struktur vermieden. Die hier genutzte k-fache verschachtelte Kreuzvalidierung wird ebenfalls in der aktuellen Literatur empfohlen.

Fazit

Eindeutige Einschränkungen bestehen hinsichtlich der Generalisierbarkeit der Ergebnisse auf andere Populationen. Ein unabhängiges externes Testset zur Validierung der Ergebnisse stand nicht zur Verfügung. Es bleibt – auch aufgrund der geringen klinischen Relevanz der vergleichsweise geringen Modellgüten – festzustellen, dass die hier aufgestellten Modelle weit von einer praktischen Anwendung in der Klinik entfernt sind.

4.5 Limitierungen und Stärken dieser Studie

Fragestellung

Die in dieser Arbeit analysierten Fragestellungen sind nicht alle grundsätzlich neu. Schon in der Vergangenheit gab es Arbeiten, die neuropsychologische Testdaten und strukturelle MRT-Daten als Modalitäten zur Klassifikation zwischen MDD-Patienten und gesunden Kontrollen und zur Analyse verschiedener Krankheitscharakteristika verwendet haben (s. Kapitel 1.2 und 1.3.2). Neu ist der konsequente, an Empfehlungen von Experten orientierte Einsatz der Methode des maschinellen Lernens für sämtliche Analysen und die explorative Analyse derart vieler verschiedener Fragestellungen basierend auf einem einheitlichen Datensatz.

Außerdem wurden die prädiktiven Modelle anhand von Testdaten evaluiert, wodurch die Ergebnisse realistischer einschätzbar sind als bei reinen Korrelationsstudien (s. Kapitel 1.3.1). Zudem wurden verschiedene objektive als auch subjektive Daten als Prädiktoren sowie Zielvariablen genutzt, wodurch die Mechanismen aus verschiedenen Blickwinkeln betrachtet werden konnten. Aufgrund der mangelnden Datenverfügbarkeit der Follow-Up-Erhebungen (s. Kapitel 2.1) zum Zeitpunkt der Erstellung dieser Arbeit wurde ein Querschnittsdesign gewählt. Dadurch ist keine eindeutige Interpretation der Kausalitäten und der zeitlichen Zusammenhänge der Ergebnisse möglich. Weiterhin bleibt unklar, ob es andere (nicht betrachtete) Krankheitscharakteristika gibt, die sich auf Basis der Prädiktoren vorhersagen lassen.

Stichprobe

Die Stichprobengröße in dieser Arbeit war mit bis zu 1650 eingeschlossenen Probanden vergleichsweise hoch (s. zum Vergleich: (24)). Nichtsdestotrotz müssen vielseitige Limitierungen hinsichtlich der Aussagekraft der Ergebnisse bedacht werden. Da die Probanden ausschließlich auf freiwilliger Basis (abgesehen von der Aufwandsentschädigung), unter der Beachtung detaillierter Ein- und Ausschlusskriterien (siehe Kapitel 2.1.1) und nur an den beiden Studienstandorten Münster und Marburg sowie Umgebung eingeschlossen wurden, muss von einer limitierten Generalisierbarkeit und Übertragbarkeit der Ergebnisse auf Populationen, die nicht adäquat durch die Stichprobe repräsentiert sind, ausgegangen werden (s. auch (17)). Wie schon in Kapitel 4.2 diskutiert, waren schwer depressive Probanden in dieser Studie zudem unterrepräsentiert.

Insgesamt war das Durchschnittsalter der eingeschlossenen Probanden im Vergleich zur durchschnittlichen deutschen Bevölkerung niedrig. Möglichen Verzerrungen der Ergebnisse durch Unterschiede hinsichtlich des Durchschnittsalters und des Bildungsniveaus zwischen den Gruppen und den anteilig vermehrten Einschluss von gesunden Kontrollen und weiblichen Probanden wurde durch entsprechende Subgruppenanalysen entgegengewirkt. Da die Zuteilung zu einer der beiden Gruppen durch den jeweiligen Testleiter anhand des SKID-I geschah (s. Kapitel 2.1.3), sind Ungenauigkeiten bei der Auswertung oder Erinnerungsverzerrungen seitens der Probanden hinsichtlich der teils retrospektiven Bewertung von Symptomen und somit fehlerhafte Gruppenzuordnungen nicht auszuschließen (97). Für die Datenerhebung standen standardisierte Vorgehensweisen zur Verfügung und die Interviews wurden teilweise unter Supervision durchgeführt, um eine hohe Reliabilität zu erreichen (37).

Psychometrische Daten

Die Erhebung der psychometrischen Daten anhand von Fragebögen (HAMD, BDI, GAF, SF-36) birgt neben vielen Chancen (bspw. eine vergleichsweise einfache Erhebung der Daten) auch weitläufige Einschränkungen. In der Literatur wird beschrieben, dass den kognitiven Prozessen, die letztendlich der Beantwortung der einzelnen Items durch den Probanden zu Grunde liegen, bei der Testkonstruktion häufig generell zu wenig Beachtung gewidmet wird (97). Vor allem bei einer Mehrdeutigkeit von einzelnen Items kann es zu kontextbezogenen oder willkürlichen Antworten kommen (97). Probanden beachten bei der Beantwortung von Items zudem in der Regel die soziale Erwünschtheit, um sich selbst in den Augen anderer positiver darzustellen als es tatsächlich der Fall ist (97). Die aktuelle Stimmungslage kann

zudem die retrospektive Abfrage von Ereignissen und die daraus resultierenden Antworten beeinflussen (97). Weitere typische Verzerrungseffekte ergeben sich dadurch, dass Probanden extreme Antworten oft zugunsten einer Tendenz zur Mitte vermeiden und dass eine gewisse Zustimmungstendenz (Akquieszenz) vorherrscht, wodurch mitunter einzelnen Items unabhängig von deren Inhalt zugestimmt wird (97). Die beiden anderen betrachteten Zielvariablen (rezidivierender Verlauf vs. einzelne Episode und hospitalisiert vs. nicht hospitalisiert) waren für solche Verzerrungseffekte hingegen weitgehend resistent.

Neuropsychologische Testung

In dieser Arbeit wurde eine breite Palette an kognitiven Funktionsbereichen getestet, wobei sowohl Geschwindigkeits- als auch Niveautests zum Einsatz kamen, was eine Stärke der Datenerhebung ist. Eine gewisse Beeinflussung der Probanden durch Versuchsleitereffekte kann dabei nicht ausgeschlossen werden (169). Grundsätzlich wird bei einem Leistungstest die maximale Leistung erfordert, sodass nur eine Verfälschung der Ergebnisse nach unten möglich ist (Dissimulation) (97). Motivationale Aspekte, die hinsichtlich einer Dissimulation eine Rolle spielen könnten, wurden in dieser Arbeit ausführlich diskutiert (s. Kapitel 4.3).

Als explizite Einschränkung des MWT-B muss beachtet werden, dass dieser Test im Vergleich zu anderen Intelligenztests zu einer Überschätzung des verbalen IQ führt (170). Für die Evaluation der hier erstellten Machine Learning-Modelle waren allerdings nur die relativen Verteilungen innerhalb der Stichproben und nicht die absoluten Werte relevant. Wird im Rahmen der Anwendung der Modelle ein anderer Test zur Ermittlung des verbalen IQ genutzt, so ist diese Einschränkung zu beachten und das Modell entsprechend anzupassen.

Strukturelle MRT-Daten

Ein großer Vorteil der Nutzung von strukturellen MRT-Daten liegt in ihrer objektiven und replizierbaren Natur (171,172). An beiden Studienzentren wurden ähnliche MRT-Scanner genutzt, die auf Grund ihrer Feldstärke von 3 Tesla eine hohe Qualität der Scans sicherstellten (173,174). Trotz der implementierten Qualitätskontrolle (s. Kapitel 2.1.5) kann es bei der automatischen Parcellierung der Gehirnschans durch Freesurfer zu fehlerhaften Segmentgrenzen und damit zu einer Unter- bzw. Überschätzung einzelner Volumina, Oberflächengrößen und Schichtdicken gekommen sein.

Dadurch, dass nicht die nativen Voxel als Prädiktoren genutzt wurden, sondern daraus berechnete komprimierte Datenpunkte, bleibt offen, ob die hier erfolgte Komplexitäts-

reduktion zu viel Varianz aus den Daten entfernt hat und die Ergebnisse negativ beeinflusst hat. Gleichzeitig bringt die Verwendung der Freesurfer-Software und die Anlehnung der Verarbeitungsprozesse an das ENIGMA-Protokoll (<http://enigma.ini.usc.edu/protocols/imaging-protocols>) (51,52,113) eine erhöhte Standardisierung und Vergleichbarkeit der Ergebnisse mit anderen Studien mit sich.

Methodik des maschinellen Lernens

Der Einsatz von maschinellem Lernen als Methode ermöglicht zwar das Erkennen komplexer Muster, liefert aber keine leicht interpretierbare explizite modellhafte Beschreibung der Zusammenhänge (65), was die Interpretation der Ergebnisse erschwert.

Der in den Daten vorhandene Bias (s. Kapitel 4.2 und 4.3) kann zu einer verminderten Generalisierbarkeit oder zu Fehlinterpretationen der Ergebnisse führen (17). Wie schon oben angesprochen (s. Kapitel 4.4), stand kein externes Testset zur unabhängigen Validierung der Modelle zur Verfügung. Durch den Einsatz der verschachtelten Kreuzvalidierung wurde das Modell zwar an Testdaten evaluiert, diese wurden allerdings an denselben Standorten erhoben. Hinsichtlich der Lernverfahren besteht das Risiko, dass ein anderes Lernverfahren bessere Ergebnisse als die beiden eingesetzten ergeben hätte (25,87) und dass der gewählte Hyperparameterraum nicht die optimalen Parameter abgedeckt hat.

4.6 Fazit und Ausblick

Bei der MDD handelt es sich um eine Erkrankung mit komplexer Ätiologie (s. Kapitel 1.1.3). Trotz der aktuellen wissenschaftlichen Fortschritte, die nicht zuletzt durch neue Methoden wie maschinelles Lernen begründet sind (s. Kapitel 1.3.2), bleiben weiterhin viele Fragestellungen offen. Das betrifft v. a. die Mechanismen der Krankheitsentstehung und die Erforschung der darauf aufbauenden Therapieansätze.

Basierend auf dem in der Literatur konsistent beschriebenen Mechanismus, dass Unterschiede zwischen gesunden Kontrollen und MDD-Patienten hinsichtlich der kognitiven Leistung (erhoben mittels neuropsychologischer Testung) und der Gehirnstruktur (erhoben mittels MRT-Untersuchung) bestehen, wurden in dieser Arbeit prädiktive Modelle zur Gruppenunterscheidung entwickelt.

Die klinische Relevanz der erreichten gewichteten Genauigkeiten im Bereich von 55-60 % bei der Evaluation an Testdaten, die den Modellen im Lernprozess nicht zu Verfügung standen, ist allerdings fraglich. Ob Zusammenhänge zwischen den oben genannten Prädiktoren und einzelnen Krankheitscharakteristika der Depression (klinischer Verlauf, Depressionsschwere und Funktionsniveau) bestehen, wird in der Literatur kontrovers diskutiert.

Anhand der prädiktiven Modelle zur Klärung dieser Mechanismen, konnte in der vorliegenden Arbeit lediglich eine Vorhersagbarkeit der SF-36-Werte (Subskala: körperliche Funktionsfähigkeit), also eines Äquivalentes der körperlichen Funktionsfähigkeit, festgestellt werden. In Bezug auf die Literatur lässt sich diskutieren, ob dies auf eine motivationale Komponente, die gleichermaßen durch die verwendeten Modalitäten und den SF-36 erfasst wird, in der Genese der MDD hinweist oder ob die Prädizierbarkeit des Wertes mit der Struktur sowie der Art und Weise der Auswertung des SF-36 zusammenhängt.

Grundsätzlich bleibt die Generalisierbarkeit der erstellten prädiktiven Modelle offen. Dabei ist kritisch zu betrachten, inwieweit die betrachtete Stichprobe repräsentativ war. Schwer depressive Probanden wurden bspw. nur vereinzelt eingeschlossen.

Basierend auf diesen Schlussfolgerungen müssen weitere Studien die Rolle der Motivation in Bezug auf die kognitive Leistung und die Gehirnstruktur bei Probanden mit einer MDD explizit klären. Dadurch wäre eine bessere kausale Einordnung der hier erkannten Zusammenhänge möglich. Dabei wird ein besonderes Augenmerk auf die neuropsychologische Testung, die Betrachtung ergänzender Instrumente zur Beurteilung des Funktionsniveaus (neben der GAF-Skala und dem SF-36) von Patienten mit einer MDD, spezielle Instrumente zur Quantifizierung der Motivation der Probanden in der Testsituation und ein longitudinales Studiendesign notwendig sein. Auch die Nutzung von funktionellen MRT-Paradigmen zur Analyse einzelner Funktionsbereiche wäre denkbar.

Neben den hier aufgestellten unimodalen Modellen wäre es – wie schon zuvor in der Literatur beschrieben (s. Kapitel 1.3.2) – wahrscheinlich, dass multimodale Machine Learning-Modelle eine höhere Prädiktionsgüte und somit auch eine höhere klinische Relevanz erreichen. Eine weitere Erforschung der Mechanismen, die der MDD zugrunde liegen, ist unumgänglich, um in der mittelfristigen Zukunft personalisierte Behandlungsansätze, individuelle Risiko-prädiktionen und eine verbesserte Differenzialdiagnostik zu ermöglichen (s. Kapitel 1.3.2).

Zur Klärung der Frage der Generalisierbarkeit der hier erstellten Modelle ist das Testen an externen Stichproben unerlässlich (17) (s. Kapitel 1.3.3). Grundsätzlich besteht hinsichtlich der klinischen Implementation prädiktiver Modelle noch viel Handlungsbedarf (17,19,58). Die im Gegensatz zu anderen medizinischen Disziplinen unscharfen Ursache-Wirkungs-Beziehungen in der Genese psychiatrischer Erkrankungen mögen ein Grund für die bestehende Zurückhaltung sein (17). Eine echte Weiterentwicklung aussichtsreicher neuer Forschungsansätze wird aber nur im Rahmen einer klinischen Reevaluation erreicht werden (17).

5 Danksagung

Zunächst möchte ich gerne meinen größten Dank gegenüber Univ.-Prof. Dr. rer. nat. Dipl.-Psych. Tim Hahn ausdrücken, der durch seine inhaltliche Anleitung und persönliche Förderung diese Arbeit überhaupt erst ermöglicht hat. Durch sein unerschöpfliches Expertenwissen im komplexen Bereich des maschinellen Lernens konnte ich sehr viel über diese spannende und zugleich hochaktuelle Methodik lernen und die Theorie im Bereich der psychiatrischen klinischen Forschung in der Praxis anwenden.

Neben dem wissenschaftlichen Erstbetreuer dieser Arbeit möchte ich an dieser Stelle auch gegenüber dem Institutsdirektor Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski meinen größten Dank aussprechen. Nur durch die Aufnahme in die hiesige Arbeitsgruppe im Institut für Translationale Psychiatrie an der Universitätsklinik Münster und durch die fundierte Einführung in die insbesondere MRT-basierte neurowissenschaftliche Forschung im Bereich der affektiven Erkrankungen konnten die für diese Arbeit notwendigen wissenschaftlichen Ansätze und Daten generiert werden.

Insbesondere möchte ich auch den Mitgliedern der Machine Learning-Arbeitsgruppe (Medical Machine Learning Lab) danken (vor allem Frau Ramona Leenings, M. Sc. und Herr Nils Winter, M. Sc.), die bei technischen Herausforderungen stets schnell und sehr kompetent zur Seite standen.

Darüber hinaus gilt mein Dank den Kolleginnen und Kollegen der FOR 2107-Arbeitsgruppe (vor allem meiner Studienbetreuerin Frau Hannah Lemke, M. Sc.), die mich so freundlich in das Team integriert und die Zeit vor Ort äußerst angenehm gestaltet haben.

Nicht zuletzt möchte ich mich bei sämtlichen Kolleginnen und Kollegen am Institut für Translationale Psychiatrie und allen dort mitwirkenden Kommilitonen und Kommilitoninnen sowie bei meinen Freunden und meiner Familie bedanken.

6 Lebenslauf

7 Literaturverzeichnis

1. James SL, Abate D, Abate KH et al. (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 392: 1789–1858. doi: 10.1016/S0140-6736(18)32279-7.
2. Alonso J, Lépine JP (2007) Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). *J Clin Psychiatry* 68: 3–9.
3. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS (2003) The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J Am Med Assoc* 289: 3095–3105. doi: 10.1001/jama.289.23.3095.
4. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, Grant BF (2018) Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry* 75: 336–346. doi: 10.1001/jamapsychiatry.2017.4602.
5. de Graaf R, ten Have M, van Gool C, van Dorsselaer S (2012) Prevalence of mental disorders and trends from 1996 to 2009: results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol* 47: 203–213. doi: 10.1007/s00127-010-0334-8.
6. Bromet E, Andrade LH, Hwang I et al. (2011) Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* 9: 90. doi: 10.1186/1741-7015-9-90.
7. Papakostas GI, Petersen T, Mahal Y, Mischoulon D, Nierenberg AA, Fava M (2004) Quality of life assessments in major depressive disorder: a review of the literature. *Gen Hosp Psychiatry* 26: 13–17. doi: 10.1016/j.genhosppsy.2003.07.004.
8. Vos T, Haby MM, Barendregt JJ, Kruijshaar M, Corry J, Andrews G (2004) The burden of major depression avoidable by longer-term treatment strategies. *Arch Gen Psychiatry* 61: 1097–1103. doi: 10.1001/archpsyc.61.11.1097.
9. Harris EC, Barraclough B (1998) Excess mortality of mental disorder. *Br J Psychiatry* 173: 11–53. doi: 10.1192/bjp.173.1.11.
10. Cuijpers P, Smit F (2002) Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord* 72: 227–236. doi: 10.1016/S0165-0327(01)00413-X.
11. Hardeveld F, Spijker J, de Graaf R, Nolen WA, Beekman ATF (2010) Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatr Scand* 122: 184–191. doi: 10.1111/j.1600-0447.2009.01519.x.
12. Spijker J, de Graaf R, Bijl RV, Beekman ATF, Ormel J, Nolen WA (2002) Duration of

- major depressive episodes in the general population: results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Br J Psychiatry* 181: 208–213. doi: 10.1192/bjp.181.3.208.
13. Biesheuvel-Leliefeld KEM, Kok GD, Bockting CLH, Cuijpers P, Hollon SD, van Marwijk HWJ, Smit F (2015) Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. *J Affect Disord* 174: 400–410. doi: 10.1016/j.jad.2014.12.016.
 14. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER (1993) The economic burden of depression in 1990. *J Clin Psychiatry* 54: 405–418.
 15. Wang PS, Simon G, Kessler RC (2003) The economic burden of depression and the cost-effectiveness of treatment. *Int J Methods Psychiatr Res* 12: 22–33. doi: 10.1002/mpr.139.
 16. Demyttenaere K, Bruffaerts R, Posada-Villa J et al. (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *J Am Med Assoc* 291: 2581–2590. doi: 10.1001/jama.291.21.2581.
 17. Cearns M, Hahn T, Baune BT (2019) Recommendations and future directions for supervised machine learning in psychiatry. *Transl Psychiatry* 9: 271. doi: 10.1038/s41398-019-0607-2.
 18. Iniesta R, Stahl D, McGuffin P (2016) Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med* 46: 2455–2465. doi: 10.1017/S0033291716001367.
 19. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. *N Engl J Med* 380: 1347–1358. doi: 10.1056/NEJMra1814259.
 20. Ferreira MF, Castanheira L, Sebastião AM, Telles-Correia D (2018) Depression assessment in clinical trials and pre-clinical tests: a critical review. *Curr Top Med Chem* 18: 1677–1703. doi: 10.2174/1568026618666181115095920.
 21. Sjöberg L, Karlsson B, Atti AR, Skoog I, Fratiglioni L, Wang HX (2017) Prevalence of depression: comparisons of different depression definitions in population-based samples of older adults. *J Affect Disord* 221: 123–131. doi: 10.1016/j.jad.2017.06.011.
 22. American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (fifth edition). American Psychiatric Association Publishing, Washington, D.C.
 23. Sullivan PF, Neale MC, Kendler KS (2000) Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157: 1552–1562. doi: 10.1176/appi.ajp.157.10.1552.

24. Gao S, Calhoun VD, Sui J (2018) Machine learning in major depression: from classification to treatment outcome prediction. *CNS Neurosci Ther* 24: 1037–1052. doi: 10.1111/cns.13048.
25. Hahn T, Nierenberg AA, Whitfield-Gabrieli S (2017) Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 22: 37–43. doi: 10.1038/mp.2016.201.
26. Grace AA (2016) Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nat Rev Neurosci* 17: 524–532. doi: 10.1038/nrn.2016.57.
27. Hamon M, Blier P (2013) Monoamine neurocircuitry in depression and strategies for new treatments. *Prog Neuro-Psychopharmacology Biol Psychiatry* 45: 54–63. doi: 10.1016/j.pnpbp.2013.04.009.
28. Ferrari F, Villa RF (2017) The neurobiology of depression: an integrated overview from biological theories to clinical evidence. *Mol Neurobiol* 54: 4847–4865. doi: 10.1007/s12035-016-0032-y.
29. Ménard C, Hodes GE, Russo SJ (2016) Pathogenesis of depression: insights from human and rodent studies. *Neuroscience* 321: 138–162. doi: 10.1016/j.neuroscience.2015.05.053.
30. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M (2006) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 163: 1905–1917. doi: 10.1176/ajp.2006.163.11.1905.
31. Cearns M, Opel N, Clark S, Kaehler C, Thalamuthu A, Heindel W, Winter T, Teismann H, Minnerup H, Dannlowski U, Berger K, Baune BT (2019) Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl Psychiatry* 9: 285. doi: 10.1038/s41398-019-0615-2.
32. Buckman JEJ, Underwood A, Clarke K, Saunders R, Hollon SD, Fearon P, Pilling S (2018) Risk factors for relapse and recurrence of depression in adults and how they operate: a four-phase systematic review and meta-synthesis. *Clin Psychol Rev* 64: 13–38. doi: 10.1016/j.cpr.2018.07.005.
33. Hughes S, Cohen D (2009) A systematic review of long-term studies of drug treated and non-drug treated depression. *J Affect Disord* 118: 9–18. doi: 10.1016/j.jad.2009.01.027.

34. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J (2010) Antidepressant drug effects and depression severity: a patient-level meta-analysis. *J Am Med Assoc* 303: 47–53. doi: 10.1001/jama.2009.1943.
35. Farah WH, Alsawas M, Mainou M, Alahdab F, Farah MH, Ahmed AT, Mohamed EA, Almasri J, Gionfriddo MR, Castaneda-Guarderas A, Mohammed K, Wang Z, Asi N, Sawchuk CN, Williams MD, Prokop LJ, Murad MH, LeBlanc A (2016) Non-pharmacological treatment of depression: a systematic review and evidence map. *Evid Based Med* 21: 214–221. doi: 10.1136/ebmed-2016-110522.
36. Königbauer J, Letsch J, Doebler P, Ebert D, Baumeister H (2017) Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. *J Affect Disord* 223: 28–40. doi: 10.1016/j.jad.2017.07.021.
37. Kircher T, Wöhr M, Nenadic I, Schwarting R, Schrott G, Alferink J, Culmsee C, Garn H, Hahn T, Müller-Myhsok B, Dempfle A, Hahmann M, Jansen A, Pfefferle P, Renz H, Rietschel M, Witt SH, Nöthen M, Krug A, Dannlowski U (2019) Neurobiology of the major psychoses: a translational perspective on brain structure and function - the FOR2107 consortium. *Eur Arch Psychiatry Clin Neurosci* 269: 949–962. doi: 10.1007/s00406-018-0943-x.
38. Srivastava K, Prakash J, Shashikumar R, Ryali V, Bhat P, Khan S (2010) Neuropsychophysiological correlates of depression. *Ind Psychiatry J* 19: 82–89. doi: 10.4103/0972-6748.90336.
39. Goodwin GM (1997) Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. *J Psychopharmacol* 11: 115–122. doi: 10.1177/026988119701100204.
40. McDermott LM, Ebmeier KP (2009) A meta-analysis of depression severity and cognitive function. *J Affect Disord* 119: 1–8. doi: 10.1016/j.jad.2009.04.022.
41. Austin MP, Mitchell P, Goodwin GM (2001) Cognitive deficits in depression: possible implications for functional neuropathology. *Br J Psychiatry* 178: 200–206. doi: 10.1192/bjp.178.3.200.
42. Austin MP, Mitchell P, Wilhelm K, Parker G, Hickie I, Brodaty H, Chan J, Eyers K, Milic M, Hadzi-Pavlovic D (1999) Cognitive function in depression: a distinct pattern of frontal impairment in melancholia? *Psychol Med* 29: 73–85. doi: 10.1017/S0033291798007788.
43. Sternberg DE, Jarvik ME (1976) Memory functions in depression: improvement with antidepressant medication. *Arch Gen Psychiatry* 33: 219–224. doi: 10.1001/archpsyc.1976.01770020055009.

44. Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA (2012) A meta-analysis of cognitive deficits in first-episode major depressive disorder. *J Affect Disord* 140: 113–124. doi: 10.1016/j.jad.2011.10.023.
45. McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallagher LA, Kudlow P, Alsuwaidan M, Baskaran A (2013) Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. *Depress Anxiety* 30: 515–527. doi: 10.1002/da.22063.
46. Biringier E, Lundervold A, Stordal K, Mykletun A, Egeland J, Bottlender R, Lund A (2005) Executive function improvement upon remission of recurrent unipolar depression. *Eur Arch Psychiatry Clin Neurosci* 255: 373–380. doi: 10.1007/s00406-005-0577-7.
47. Beblo T, Baumann B, Bogerts B, Wallesch CW, Herrmann M (1999) Neuropsychological correlates of major depression: a short-term follow-up. *Cogn Neuropsychiatry* 4: 333–341. doi: 10.1080/135468099395864.
48. Herrmann LL, Goodwin GM, Ebmeier KP (2007) The cognitive neuropsychology of depression in the elderly. *Psychol Med* 37: 1693–1702. doi: 10.1017/S0033291707001134.
49. Basso MR, Lowery N, Ghormley C, Combs D, Purdie R, Neel J, Davis M, Bornstein R (2007) Comorbid anxiety corresponds with neuropsychological dysfunction in unipolar depression. *Cogn Neuropsychiatry* 12: 437–456. doi: 10.1080/13546800701446517.
50. Lin K, Xu G, Lu W, Ouyang H, Dang Y, Lorenzo-Seva U, Guo Y, Bessonov D, Akiskal HS, So KF, Lee TMC (2014) Neuropsychological performance in melancholic, atypical and undifferentiated major depression during depressed and remitted states: a prospective longitudinal study. *J Affect Disord* 168: 184–191. doi: 10.1016/j.jad.2014.06.032.
51. Schmaal L, Veltman DJ, van Erp TGM et al. (2016) Subcortical brain alterations in major depressive disorder: findings from the ENIGMA major depressive disorder working group. *Mol Psychiatry* 21: 806–812. doi: 10.1038/mp.2015.69.
52. Schmaal L, Hibar DP, Sämann PG et al. (2017) Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA major depressive disorder working group. *Mol Psychiatry* 22: 900–909. doi: 10.1038/mp.2016.60.
53. Zhang FF, Peng W, Sweeney JA, Jia ZY, Gong QY (2018) Brain structure alterations in depression: psychoradiological evidence. *CNS Neurosci Ther* 24: 994–1003. doi: 10.1111/cns.12835.
54. Arnone D, McIntosh AM, Ebmeier KP, Munafò MR, Anderson IM (2012) Magnetic

- resonance imaging studies in unipolar depression: systematic review and meta-regression analyses. *Eur Neuropsychopharmacol* 22: 1–16. doi: 10.1016/j.euroneuro.2011.05.003.
55. Koolschijn PCMP, van Haren NEM, Lensvelt-Mulders GJLM, Hulshoff Pol HE, Kahn RS (2009) Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Hum Brain Mapp* 30: 3719–3735. doi: 10.1002/hbm.20801.
56. Lorenzetti V, Allen NB, Fornito A, Yücel M (2009) Structural brain abnormalities in major depressive disorder: a selective review of recent MRI studies. *J Affect Disord* 117: 1–17. doi: 10.1016/j.jad.2008.11.021.
57. Mitchell TM (1997) *Machine learning*. McGraw-Hill, New York
58. Liu Y, Chen PHC, Krause J, Peng L (2019) How to read articles that use machine learning: users' guides to the medical literature. *J Am Med Assoc* 322: 1806–1816. doi: 10.1001/jama.2019.16489.
59. Gabrieli JDE, Ghosh SS, Whitfield-Gabrieli S (2015) Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85: 11–26. doi: 10.1016/j.neuron.2014.10.047.
60. Lueken U, Hahn T (2016) Functional neuroimaging of psychotherapeutic processes in anxiety and depression: from mechanisms to predictions. *Curr Opin Psychiatry* 29: 25–31. doi: 10.1097/YCO.0000000000000218.
61. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B (2017) Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage* 145: 166–179. doi: 10.1016/j.neuroimage.2016.10.038.
62. Whelan R, Garavan H (2014) When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol Psychiatry* 75: 746–748. doi: 10.1016/j.biopsych.2013.05.014.
63. Bishop CM (2006) *Pattern recognition and machine learning*. Springer, New York
64. Leenings R, Winter NR, Plagwitz L, Holstein V, Ernsting J, Steenweg J, Gebker J, Sarink K, Emden D, Grotegerd D, Opel N, Risse B, Jiang X, Dannlowski U, Hahn T (2020) PHOTON - a Python API for rapid machine learning model development (version 1). In: arXiv, Cornell University (Hrsg.), <http://arxiv.org/abs/2002.05426v1>, heruntergeladen am 12.05.2021
65. Bzdok D, Altman N, Krzywinski M (2018) Points of significance: statistics versus machine learning. *Nat Methods* 15: 233–234. doi: 10.1038/nmeth.4642.
66. Lee Y, Ragguett R-M, Mansur RB et al. (2018) Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and

- systematic review. *J Affect Disord* 241: 519–532. doi: 10.1016/j.jad.2018.08.073.
67. Patel MJ, Khalaf A, Aizenstein HJ (2016) Studying depression using imaging and machine learning methods. *NeuroImage Clin* 10: 115–123. doi: 10.1016/j.nicl.2015.11.003.
68. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G (2017) Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* 74: 370–378. doi: 10.1001/jamapsychiatry.2017.0025.
69. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, Cannon TD, Krystal JH, Corlett PR (2016) Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 3: 243–250. doi: 10.1016/S2215-0366(15)00471-X.
70. Ozomaro U, Wahlestedt C, Nemeroff CB (2013) Personalized medicine in psychiatry: problems and promises. *BMC Med* 11: 132. doi: 10.1186/1741-7015-11-132.
71. Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Burger C, Munker S, Muhlmann L, Wahl P, Heindel W, Arolt V, Alferink J, Zwanzger P, Zavorotnyy M, Kugel H, Dannlowski U (2016) Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* 73: 557–564. doi: 10.1001/jamapsychiatry.2016.0316.
72. Pigoni A, Delvecchio G, Madonna D, Bressi C, Soares J, Brambilla P (2019) Can machine learning help us in dealing with treatment resistant depression? a review. *J Affect Disord* 259: 21–26. doi: 10.1016/j.jad.2019.08.009.
73. McGorry P, Keshavan M, Goldstone S, Amminger P, Allott K, Berk M, Lavoie S, Pantelis C, Yung A, Wood S, Hickie I (2014) Biomarkers and clinical staging in psychiatry. *World Psychiatry* 13: 211–223. doi: 10.1002/wps.20144.
74. Lee SH, Ripke S, Neale BM et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45: 984–994. doi: 10.1038/ng.2711.
75. Del Boca FK, Noll JA (2000) Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction* 95: 347–360. doi: 10.1080/09652140020004278.
76. Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K (2017) Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain Behav* 7. doi: 10.1002/brb3.633.
77. Ma Y, Ji J, Huang Y, Gao H, Li Z, Dong W, Zhou S, Zhu Y, Dang W, Zhou T, Yu H, Yu

- B, Long Y, Liu L, Sachs G, Yu X (2019) Implementing machine learning in bipolar diagnosis in China. *Transl Psychiatry* 9: 305. doi: 10.1038/s41398-019-0638-8.
78. Hahn T, Marquand AF, Ehlis AC, Dresler T, Kittel-Schneider S, Jarczok TA, Lesch KP, Jakob PM, Mourao-Miranda J, Brammer MJ, Fallgatter AJ (2011) Integrating neurobiological markers of depression. *Arch Gen Psychiatry* 68: 361–368. doi: 10.1001/archgenpsychiatry.2010.178.
79. Mumtaz W, Ali SSA, Yasin MAM, Malik AS (2018) A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Med Biol Eng Comput* 56: 233–246. doi: 10.1007/s11517-017-1685-z.
80. Kim EY, Lee MY, Kim SH, Ha K, Kim KP, Ahn YM (2017) Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm. *Prog Neuro-Psychopharmacology Biol Psychiatry* 76: 65–71. doi: 10.1016/j.pnpbp.2017.02.014.
81. Wardenaar KJ, van Loo HM, Cai T et al. (2014) The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychol Med* 44: 3289–3302. doi: 10.1017/S0033291714000993.
82. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, Ebert DD, Hwang I, Li J, de Jonge P, Nierenberg AA, Petukhova MV, Rosellini AJ, Sampson NA, Schoevers RA, Wilcox MA, Zaslavsky AM (2016) Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry* 21: 1366–1371. doi: 10.1038/mp.2015.198.
83. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S et al. (2018) Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* 75: 1156–1172. doi: 10.1001/jamapsychiatry.2018.2165.
84. Janssen RJ, Mourão-Miranda J, Schnack HG (2018) Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biol Psychiatry Cogn Neurosci Neuroimaging* 3: 798–808. doi: 10.1016/j.bpsc.2018.04.004.
85. Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, Winter NR, Leenings R, Eickhoff SB, Kircher T, Krug A, Nenadic I, Arolt V, Clark S, Baune BT, Jiang X, Dannlowski U, Hahn T (2021) Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology*. doi: 10.1038/s41386-021-01020-7.
86. Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017) Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage* 145: 137–165. doi:

- 10.1016/j.neuroimage.2016.02.079.
87. Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. *Neural Comput* 8: 1341–1390. doi: 10.1162/neco.1996.8.7.1341.
 88. Wittchen HU, Wunderlich U, Gruschwitz S, Zaudig M (1997) SKID I, Strukturiertes Klinisches Interview für DSM-IV, Achse I: Psychische Störungen, Interviewheft und Beurteilungsheft; eine deutschsprachige, erweiterte Bearbeitung der amerikanischen Originalversion des SKID I. Hogrefe, Göttingen
 89. Sass H, Wittchen HU, Zaudig M, Houben I (2003) Diagnostisches und Statistisches Manual Psychischer Störungen – Textrevision – DSM-IV-TR. Hogrefe, Göttingen
 90. Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23: 56–62. doi: 10.1136/jnnp.23.1.56.
 91. Hautzinger M, Keller F, Kühner C (2006) BDI-II Beck Depressions-Inventar, Manual und Fragebogenblock. Harcourt Test Services, Frankfurt
 92. Beck AT, Steer RA, Brown GK (1996) Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio
 93. Bullinger M (2000) Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36-Health Survey. *Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz* 43: 190–197. doi: 10.1007/s001030050034.
 94. Tarlov AR, Ware Jr. JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M (1989) The Medical Outcomes Study: an application of methods for monitoring the results of medical care. *J Am Med Assoc* 262: 925–930. doi: 10.1001/jama.262.7.925.
 95. Williams JBW, Terman M (2003) Structured interview guide for the Hamilton Depression Rating Scale with atypical depression supplement (SIGH-ADS). New York State Psychiatric Institute, New York
 96. Möller H-J (2001) Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. *Eur Arch Psychiatry Clin Neurosci* 251: 13–20. doi: 10.1007/bf03035121.
 97. Jonkisz E, Moosbrugger H, Brandt H (2012) Planung und Entwicklung von Tests und Fragebogen. In: Moosbrugger H, Kelava A (Hrsg.) *Testtheorie und Fragebogenkonstruktion*, Springer, Berlin, S. 27–74
 98. Helmstaedter CA, Lendt M, Lux S (2001) *Verbaler Lern- und Merkfähigkeitstest*. Beltz Test, Göttingen
 99. Reitan R (1992) *Trail Making Test: manual for administration and scoring*. Reitan Neuropsychology Laboratory, Tucson
 100. Härting C, Wechsler D (2000) *Wechsler-Gedächtnistest - revidierte Fassung WMS-R*,

- Manual, deutsche Adaptation der revidierten Fassung der Wechsler Memory scale.
Huber, Bern
101. von Aster M, Neubauer A, Horn R (2006) Wechsler Intelligenztest für Erwachsene (WIE): deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler. Harcourt Test Services, Frankfurt
 102. Tewes U (1991) Hamburg-Wechsler-Intelligenztest für Erwachsene - Revision 1991: HAWIE-R. Huber, Bern
 103. Brickenkamp R (2002) Test d2: Aufmerksamkeits-Belastungs-Test. Hogrefe, Göttingen
 104. Aschenbrenner S, Tucha O, Lange KW (2001) RWT: Regensburger Wortflüssigkeits-Test. Hogrefe, Göttingen
 105. Lehrl S (2005) Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B. Spitta, Balingen
 106. Volz-Sidiropoulou E, Poll E, Forkmann T, Gauggel S (2010) Erweiterte Altersnormen zum Verbalen Lern- und Merkfähigkeitstest (VLMT). *Klin Diagnostik und Eval* 3: 226–243.
 107. Bowie CR, Harvey PD (2006) Administration and interpretation of the Trail Making Test. *Nat Protoc* 1: 2277–2281. doi: 10.1038/nprot.2006.390.
 108. Crowe S (1998) The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. *J Clin Psychol* 54: 585–591. doi: 10.1002/(SICI)1097-4679(199808)54:5<585::AID-JCLP4>3.0.CO;2-K.
 109. Flint C, Förster K, Koser SA et al. (2020) Biological sex classification with structural MRI data shows increased misclassification in transgender women. *Neuropsychopharmacology* 45: 1758–1765. doi: 10.1038/s41386-020-0666-3.
 110. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31: 968–980. doi: 10.1016/j.neuroimage.2006.01.021.
 111. Fischl B (2012) FreeSurfer. *Neuroimage* 62: 774–781. doi: 10.1016/j.neuroimage.2012.01.021.
 112. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33: 341–355. doi: 10.1016/S0896-6273(02)00569-X.
 113. Stein JL, Medland SE, Vasquez AA et al. (2012) Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 44: 552–561.

- doi: 10.1038/ng.2250.
114. Vogelbacher C, Möbius TWD, Sommer J, Schuster V, Dannlowski U, Kircher T, Dempfle A, Jansen A, Bopp MHA (2018) The Marburg-Münster Affective Disorders Cohort Study (MACS): a quality assurance protocol for MR neuroimaging data. *Neuroimage* 172: 450–460. doi: 10.1016/j.neuroimage.2018.01.079.
 115. Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 46: 399–424. doi: 10.1080/00273171.2011.568786.
 116. Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab* 10: 486–489. doi: 10.5812/ijem.3505.
 117. Zimmerman DW (1987) Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances. *J Exp Educ* 55: 171–174. doi: 10.1080/00220973.1987.10806451.
 118. D’agostino RB, Belanger A, D’agostino Jr. RB (1990) A suggestion for using powerful and informative tests of normality. *Am Stat* 44: 316–321. doi: 10.1080/00031305.1990.10475751.
 119. Virtanen P, Gommers R, Oliphant TE et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17: 261–272. doi: 10.1038/s41592-019-0686-2.
 120. McKinney W (2010) Data structures for statistical computing in Python. In: van der Walt S, Millman J (Hrsg.) *Proceedings of the 9th Python in Science Conference*, Austin, S. 56–61
 121. Hunter JD (2007) Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9: 90–95. doi: 10.1109/MCSE.2007.55.
 122. Waskom M, Botvinnik O, Gelbart M et al. (2020) Seaborn: statistical data visualization (v0.11.0). In: Zenodo, CERN (Hrsg.), <https://zenodo.org/record/4019146>, heruntergeladen am 12.05.2021
 123. Leenings R, Winter NR, Sarink K, Ernsting J, Jiang X, Dannlowski U, Hahn T (2020) The PHOTON wizard - towards educational machine learning code generators. In: arXiv, Cornell University (Hrsg.), <http://arxiv.org/abs/2002.05432>, heruntergeladen am 12.05.2021
 124. Harris CR, Millman KJ, van der Walt SJ et al. (2020) Array programming with NumPy. *Nature* 585: 357–362. doi: 10.1038/s41586-020-2649-2.
 125. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher

- M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12: 2825–2830.
126. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Mellish CS (Hrsg.) *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Francisco, S. 1137–1143
127. Poldrack RA, Huckins G, Varoquaux G (2020) Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* 77: 534–540. doi: 10.1001/jamapsychiatry.2019.3671.
128. Cawley GC, Talbot NLC (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 11: 2079–2107.
129. Vanwinckelen G, Blockeel H (2012) On estimating model accuracy with repeated cross-validation. In: De Baets B, Manderick B, Rademaker M, Waegeman W (Hrsg.) *BeneLearn 2012: proceedings of the 21st Belgian-Dutch conference on machine learning*, Benelearn 2012 Organization Committee, Ghent, S. 39–44
130. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: StratifiedKFold. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html, heruntergeladen am 12.05.2021
131. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: LabelEncoder. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>, heruntergeladen am 12.05.2021
132. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: SimpleImputer. <https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html>, heruntergeladen am 12.05.2021
133. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: StandardScaler. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>, heruntergeladen am 12.05.2021
134. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: SVM. <https://scikit-learn.org/stable/modules/svm.html#svm>, heruntergeladen am 12.05.2021
135. Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola AJ, Bartlett P, Schölkopf B, Schuurmans D (Hrsg.) *Advances in large margin classifiers*, MIT Press, Cambridge, S. 61–74
136. Chang C-C, Lin C-JL (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2. doi: 10.1145/1961189.1961199.
137. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: SVC.

- <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>,
heruntergeladen am 12.05.2021
138. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: SVR. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>,
heruntergeladen am 12.05.2021
139. Sordo M, Zeng Q (2005) On sample size and classification accuracy: a performance comparison. In: Oliveira JL, Maojo V, Martín-Sánchez F, Pereira AS (Hrsg.) Proceedings of the ISBMDA: International Symposium on Biological and Medical Data Analysis, Springer, Berlin, S. 193–201
140. Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. doi: 10.1023/A:1010933404324.
141. Ishwaran H, Lu M (2019) Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* 38: 558–582. doi: 10.1002/sim.7803.
142. Breiman L (1996) Bagging predictors. *Mach Learn* 24: 123–140. doi: 10.1007/bf00058655.
143. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20: 832–844. doi: 10.1109/34.709601.
144. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. *Mach Learn* 63: 3–42. doi: 10.1007/s10994-006-6226-1.
145. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: RandomForestClassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>, heruntergeladen am 12.05.2021
146. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: RandomForestRegressor. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>, heruntergeladen am 12.05.2021
147. Hofmann T, Schölkopf B, Smola AJ (2008) Kernel methods in machine learning. *Ann Stat* 36: 1171–1220. doi: 10.1214/0090536070000000677.
148. Liaw A, Wiener M (2002) Classification and regression by randomforest. *R News* 2: 18–22.
149. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: mean_absolute_error. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html, heruntergeladen am 12.05.2021
150. Kelleher JD, Mac Namee B, D’Arcy A (2015) Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT Press,

Cambridge

151. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: `balanced_accuracy_score`. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html, heruntergeladen am 12.05.2021
152. Brodersen K, Ong CS, Stephan K, Buhmann J (2010) The balanced accuracy and its posterior distribution. In: Boyer KL, Cetin M, Lee S-W (Hrsg.) Proceedings of the 20th International Conference on Pattern Recognition, IEEE, Los Alamitos, S. 3121–3124
153. Urbanowicz RJ, Moore JH (2015) ExSTraCS 2.0: description and evaluation of a scalable learning classifier system. *Evol Intell* 8: 89–116. doi: 10.1007/s12065-015-0128-8.
154. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: `DummyClassifier`. <https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html>, heruntergeladen am 12.05.2021
155. Pedregosa F, Varoquaux G, Gramfort A et al. (2021) Scikit-learn documentation: `DummyRegressor`. <https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html>, heruntergeladen am 12.05.2021
156. Golland P, Fischl B (2003) Permutation tests for classification: towards statistical significance in image-based studies. In: Taylor C, Noble JA (Hrsg.) Proceedings of the 18th International Conference of Information Processing in Medical Imaging, Springer, Berlin, S. 330–341
157. Ojala M, Garriga GC (2010) Permutation tests for studying classifier performance. *J Mach Learn Res* 11: 1833–1863.
158. Dinga R, Marquand AF, Veltman DJ, Beekman ATF, Schoevers RA, van Hemert AM, Penninx BWJH, Schmaal L (2018) Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry* 8: 241. doi: 10.1038/s41398-018-0289-1.
159. Moritz S, Stöckert K, Hauschildt M, Lill H, Jelinek L, Beblo T, Diedrich S, Arlt S (2017) Are we exaggerating neuropsychological impairment in depression? reopening a closed chapter. *Expert Rev Neurother* 17: 839–846. doi: 10.1080/14737175.2017.1347040.
160. Reitan RM, Wolfson D (1995) Influence of age and education on neuropsychological test results. *Clin Neuropsychol* 9: 151–158. doi: 10.1080/13854049508401597.
161. Le Carret N, Lafont S, Letenneur L, Dartigues JF, Mayo W, Fabrigoule C (2003) The effect of education on cognitive performances and its implication for the constitution of the cognitive reserve. *Dev Neuropsychol* 23: 317–337. doi:

- 10.1207/S15326942DN2303_1.
162. Ho AJ, Raji CA, Becker JT, Lopez OL, Kuller LH, Hua X, Dinov ID, Stein JL, Rosano C, Toga AW, Thompson PM (2011) The effects of physical activity, education, and body mass index on the aging brain. *Hum Brain Mapp* 32: 1371–1382. doi: 10.1002/hbm.21113.
163. Liu Y, Julkunen V, Paajanen T, Westman E, Wahlund LO, Aitken A, Sobow T, Mecocci P, Tsolaki M, Vellas B, Muehlboeck S, Spenger C, Lovestone S, Simmons A, Soininen H (2012) Education increases reserve against Alzheimer's disease: evidence from structural MRI analysis. *Neuroradiology* 54: 929–938. doi: 10.1007/s00234-012-1005-0.
164. Fervaha G, Zakzanis KK, Foussias G, Graff-Guerrero A, Agid O, Remington G (2014) Motivational deficits and cognitive test performance in schizophrenia. *JAMA Psychiatry* 71: 1058–1065. doi: 10.1001/jamapsychiatry.2014.1105.
165. Elliott R, Sahakian BJ, McKay AP, Herrod JJ, Robbins TW, Paykel ES (1996) Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychol Med* 26: 975–989. doi: 10.1017/s0033291700035303.
166. Krosnick JA (1999) Survey research. *Annu Rev Psychol* 50: 537–567. doi: 10.1146/annurev.psych.50.1.537.
167. Grotegerd D, Stuhrmann A, Kugel H, Schmidt S, Redlich R, Zwanzger P, Rauch AV, Heindel W, Zwieterlood P, Arolt V, Suslow T, Dannlowski U (2014) Amygdala excitability to subliminally presented emotional faces distinguishes unipolar and bipolar depression: an fMRI and pattern classification study. *Hum Brain Mapp* 35: 2995–3007. doi: 10.1002/hbm.22380.
168. Bartsch LJ, Butterworth P, Byles JE, Mitchell P, Shaw J, Anstey KJ (2011) Examining the SF-36 in an older population: analysis of data and presentation of Australian adult reference scores from the Dynamic Analyses to Optimise Ageing (DYNOPTA) project. *Qual Life Res* 20: 1227–1236. doi: 10.1007/s11136-011-9864-0.
169. Moosbrugger H, Kelava A (2012) Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: Moosbrugger H, Kelava A (Hrsg.) *Testtheorie und Fragebogenkonstruktion*, Springer, Berlin, S. 7–26
170. Satzger W, Fessmann H, Engel R (2002) Liefern HAWIE-R, WST und MWT-B vergleichbare IQ-Werte? *Zeitschrift für Differ und Diagnostische Psychol* 23: 159–170. doi: 10.1024//0170-1789.23.2.159.
171. Jack CR, Bentley MD, Twomey CK, Zinsmeister AR (1990) MR imaging-based volume

- measurements of the hippocampal formation and anterior temporal lobe: validation studies. *Radiology* 176: 205–209. doi: 10.1148/radiology.176.1.2353093.
172. Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasti P, Locascio JJ, Bakkour A, van der Kouwe A, Triantafyllou C, Corkin S, Dickerson BC (2009) Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. *Neuroimage* 44: 1324–1333. doi: 10.1016/j.neuroimage.2008.10.037.
173. Schmitt F, Grosu D, Mohr C, Purdy D, Salem K, Scott KT, Stoeckel B (2004) 3 Tesla-MRT: der Erfolg höherer Feldstärken. *Radiologe* 44: 31–47. doi: 10.1007/s00117-003-1000-x.
174. Livshits I, Hussein S, Kennedy C, Weinstock-Guttman B, Hojnacki D, Zivadinov R (2012) Comparison of a 1.5T standard vs. 3T optimized protocols in multiple sclerosis patients. *Minerva Med* 103: 97–102.

8 Anhang

8.1 Ausschlusskriterien (somatische Erkrankungen)

In Tabelle 8-1 sind die somatischen Erkrankungen dargestellt, bei deren Vorliegen Probanden aus der Studie ausgeschlossen wurden.

Tabelle 8-1: Übersicht über die Ausschlusskriterien (somatische Erkrankungen)

Kein Einschluss:	Einschluss auf individueller Basis entschieden:
Colitis ulcerosa	Allergien generell
Enzephalopathie	Asthma
Epilepsie	Borreliose
erblich bedingter Haltetremor	chron. Bronchitis
fazioskapulohumerale Muskeldystrophie	chron. Leukozytose
Hepatitis C	Diabetes mellitus Typ 1
Karzinome	Eisendepoterkrankung
Lupus	Hashimoto
Morbus Bechterew (HLA B27 pos.)	Hepatitis B
Morbus Crohn	lymphozytäre Thyreoiditis
Multiple Sklerose	Migräne mit Aura
Narkolepsie	Neurodermitis
Neuroborreliose	paroxysmale Hemikranie
Neuronale Muskelatrophie	Rheuma
Sarkoidose	Restless Legs Syndrom (RLS)
Sjörger-Syndrom	Schlafapnoe
Sklerodermie	Tinnitus
undifferenzierte Kollagenose	Tuberkulose

8.2 Variablenübersicht (Modalität: neuropsychologische Testung)

In der Tabelle 8-2 und der Tabelle 8-3 sind die Prädiktoren der Analysen mit den Daten der neuropsychologischen Testbatterie dargestellt.

Tabelle 8-2: Variablenübersicht für die Modalität: neuropsychologische Testung (Teil 1)

Test	Variablenname	Bedeutung
VLMT	VLMT_RichtigA1	Anzahl der richtigen Wörter (1. Durchgang)
	VLMT_FPA1	Anzahl der Falsch Positiven (1. Durchgang)
	VLMT_PA1	Anzahl der Perseverationen (1. Durchgang)
	VLMT_RichtigA2	Anzahl der richtigen Wörter (2. Durchgang)
	VLMT_FPA2	Anzahl der Falsch Positiven (2. Durchgang)
	VLMT_PA2	Anzahl der Perseverationen (2. Durchgang)
	VLMT_RichtigA3	Anzahl der richtigen Wörter (3. Durchgang)
	VLMT_FPA3	Anzahl der Falsch Positiven (3. Durchgang)
	VLMT_PA3	Anzahl der Perseverationen (3. Durchgang)
	VLMT_RichtigA4	Anzahl der richtigen Wörter (4. Durchgang)
	VLMT_FPA4	Anzahl der Falsch Positiven (4. Durchgang)
	VLMT_PA4	Anzahl der Perseverationen (4. Durchgang)
	VLMT_RichtigA5	Anzahl der richtigen Wörter (5. Durchgang)
	VLMT_FPA5	Anzahl der Falsch Positiven (5. Durchgang)
	VLMT_PA5	Anzahl der Perseverationen (5. Durchgang)
	VLMT_RichtigB1	Anzahl der richtigen Wörter (Durchgang B)
	VLMT_RichtigA6	Anzahl der richtigen Wörter (6. Durchgang)
	VLMT_RichtigInA6	Anzahl Interferenzen (6. Durchgang)
	VLMT_RichtigA7	Anzahl der richtigen Wörter (7. Durchgang)
	VLMT_RichtigInA7	Anzahl Interferenzen (7. Durchgang)
	VLMT_TrefferA	Anzahl der Treffer (Wiedererkennung)
	VLMT_FehlerB	Anzahl der Fehler (Wiedererk., Liste B)
	VLMT_FehlerNeuesWort	Anzahl der Fehler (Wiedererk., neues Wort)
	VLMT_Sum_Richtige	Summe der Richtigen (Durchgang 1 bis 5)
	VLMT_VerlustInteferenz	VLMT_RichtigA5 - VLMT_RichtigA6
	VLMT_VerlustZeitlicheVerzoeg.	VLMT_RichtigA6 - VLMT_RichtigA7
	VLMT_RecognitionCorrected	VLMT_TrefferA - (VLMT_FehlerB + VLMT_FehlerNeuesWort)

Tabelle 8-3: Variablenübersicht für die Modalität: neuropsychologische Testung (Teil 2)

Test	Variablenname	Bedeutung
VLMT	VLMT_Sum_FalschPositive	Summe der Falsch Positiven (Durchgang 1 bis 5)
	VLMT_Sum_Perseverationen	Summe der Perseverationen (Durchgang 1 bis 5)
	VLMT_Sum_Interferenz	Summe der Interferenzen (Durchgang 6 und 7)
TMT	TMT_PfadfinderA	Zeit in Sekunden (Teil A)
	TMT_PfadfinderB	Zeit in Sekunden (Teil B)
	TMT_Differenz	Differenz der Zeiten von Teil B und A
Blockspanne	Blockspanne_Vorwaerts	Anzahl richtiger Folgen (vorwärts)
	Blockspanne_Rueckwaerts	Anzahl richtiger Folgen (rückwärts)
	Blockspanne_Gesamt	Anzahl aller richtigen Folgen
BZT	BZT_Sum	Summer der richtigen Reihen
ZST	ZST_Sum	Gesamtsumme richtiger Symbole
d2	D2_GZ	Gesamtzahl der bearbeiteten Zeichen
	D2_F1	Auslassungsfehler
	D2_F2	Verwechslungsfehler
	D2_KL	Konzentrationsleistung (bearbeitete Zielobjekte - D2_F1 - D2_F2)
RWT	RWT_Tiere_Sum	Summe der Wörter (Kategorie Tiere)
	RWT_Tiere_Pers	Perseverationen (Kategorie Tiere)
	RWT_Tiere_Fehler	Fehler (Kategorie Tiere)
	RWT_P_Sum	Summe der Wörter (Buchstabe P)
	RWT_P_Pers	Perseverationen (Buchstabe P)
	RWT_P_Fehler	Fehler (Buchstabe P)
	RWT_Altern_Sum	Summer der Wörter (altern. Kategorien)
	RWT_Altern_Pers	Perseverationen (altern. Kategorien)
	RWT_Altern_Fehler	Fehler (altern. Kategorien)
	RWT_Tiere_SumCorrected	korrigierte Summe (Kategorie Tiere)
	RWT_P_SumCorrected	korrigierte Summe (Buchstabe P)
	RWT_Altern_SumCorrected	korrigierte Summe (altern. Kategorien)
MWT-B	IQ	Übersetzung der Anzahl der richtigen Items in einen IQ-Wert (anhand einer Tabelle)

8.3 Variablenübersicht (Modalität: strukturelle MRT-Daten)

In der Tabelle 8-4, der Tabelle 8-5 und der Tabelle 8-6 sind sämtliche Prädiktoren der Analysen mit den strukturellen MRT-Daten gelistet. Wie in Kapitel 2.1.5 beschrieben, wurden die entsprechenden Variablen mithilfe der Freesurfer Software (<https://surfer.nmr.mgh.harvard.edu/>) (111) generiert. Die genauen Spezifikationen der Variablen können der entsprechenden Dokumentation entnommen werden.

Tabelle 8-4: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 1)

Variablenname	Variablenname
L_bankssts_thickavg	R_bankssts_thickavg
L_caudalanteriorcingulate_thickavg	R_caudalanteriorcingulate_thickavg
L_caudalmiddlefrontal_thickavg	R_caudalmiddlefrontal_thickavg
L_cuneus_thickavg	R_cuneus_thickavg
L_entorhinal_thickavg	R_entorhinal_thickavg
L_fusiform_thickavg	R_fusiform_thickavg
L_inferiorparietal_thickavg	R_inferiorparietal_thickavg
L_inferiortemporal_thickavg	R_inferiortemporal_thickavg
L_isthmuscingulate_thickavg	R_isthmuscingulate_thickavg
L_lateraloccipital_thickavg	R_lateraloccipital_thickavg
L_lateralorbitofrontal_thickavg	R_lateralorbitofrontal_thickavg
L_lingual_thickavg	R_lingual_thickavg
L_medialorbitofrontal_thickavg	R_medialorbitofrontal_thickavg
L_middletemporal_thickavg	R_middletemporal_thickavg
L parahippocampal_thickavg	R parahippocampal_thickavg
L_paracentral_thickavg	R_paracentral_thickavg
L_parsopercularis_thickavg	R_parsopercularis_thickavg
L_parsorbitalis_thickavg	R_parsorbitalis_thickavg
L_parstriangularis_thickavg	R_parstriangularis_thickavg
L_pericalcarine_thickavg	R_pericalcarine_thickavg
L_postcentral_thickavg	R_postcentral_thickavg
L_posteriorcingulate_thickavg	R_posteriorcingulate_thickavg
L_precentral_thickavg	R_precentral_thickavg
L_precuneus_thickavg	R_precuneus_thickavg

Tabelle 8-5: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 2)

Variablenname	Variablenname
L_rostralanteriorcingulate_thickavg	R_rostralanteriorcingulate_thickavg
L_rostralmiddlefrontal_thickavg	R_rostralmiddlefrontal_thickavg
L_superiorfrontal_thickavg	R_superiorfrontal_thickavg
L_superiorparietal_thickavg	R_superiorparietal_thickavg
L_superiortemporal_thickavg	R_superiortemporal_thickavg
L_supramarginal_thickavg	R_supramarginal_thickavg
L_frontalpole_thickavg	R_frontalpole_thickavg
L_temporalpole_thickavg	R_temporalpole_thickavg
L_transversetemporal_thickavg	R_transversetemporal_thickavg
L_insula_thickavg	R_insula_thickavg
LThickness	RThickness
LSurfArea	RSurfArea
L_bankssts_surfavg	R_bankssts_surfavg
L_caudalanteriorcingulate_surfavg	R_caudalanteriorcingulate_surfavg
L_caudalmiddlefrontal_surfavg	R_caudalmiddlefrontal_surfavg
L_cuneus_surfavg	R_cuneus_surfavg
L_entorhinal_surfavg	R_entorhinal_surfavg
L_fusiform_surfavg	R_fusiform_surfavg
L_inferiorparietal_surfavg	R_inferiorparietal_surfavg
L_inferiortemporal_surfavg	R_inferiortemporal_surfavg
L_isthmuscingulate_surfavg	R_isthmuscingulate_surfavg
L_lateraloccipital_surfavg	R_lateraloccipital_surfavg
L_lateralorbitofrontal_surfavg	R_lateralorbitofrontal_surfavg
L_lingual_surfavg	R_lingual_surfavg
L_medialorbitofrontal_surfavg	R_medialorbitofrontal_surfavg
L_middletemporal_surfavg	R_middletemporal_surfavg
L parahippocampal_surfavg	R parahippocampal_surfavg
L_paracentral_surfavg	R_paracentral_surfavg
L_parsopercularis_surfavg	R_parsopercularis_surfavg
L_parsorbitalis_surfavg	R_parsorbitalis_surfavg
L_parstriangularis_surfavg	R_parstriangularis_surfavg
L_pericalcarine_surfavg	R_pericalcarine_surfavg
L_postcentral_surfavg	R_postcentral_surfavg
L_posteriorcingulate_surfavg	R_posteriorcingulate_surfavg
L_precentral_surfavg	R_precentral_surfavg

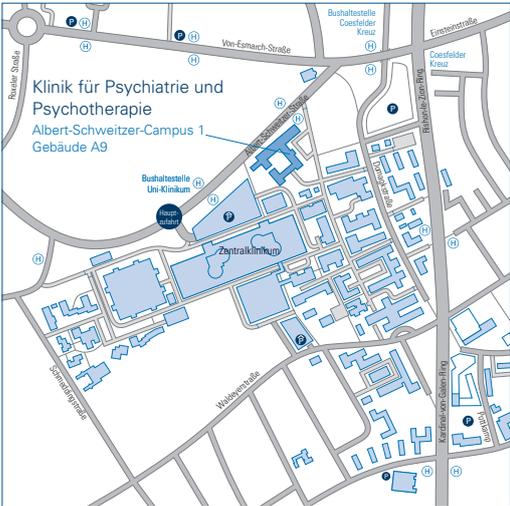
Tabelle 8-6: Variablenübersicht für die Modalität: strukturelle MRT-Daten (Teil 3)

Variablenname	Variablenname
L_precuneus_surfavg	R_precuneus_surfavg
L_rostralanteriorcingulate_surfavg	R_rostralanteriorcingulate_surfavg
L_rostralmiddlefrontal_surfavg	R_rostralmiddlefrontal_surfavg
L_superiorfrontal_surfavg	R_superiorfrontal_surfavg
L_superiorparietal_surfavg	R_superiorparietal_surfavg
L_superiortemporal_surfavg	R_superiortemporal_surfavg
L_supramarginal_surfavg	R_supramarginal_surfavg
L_frontalpole_surfavg	R_frontalpole_surfavg
L_temporalpole_surfavg	R_temporalpole_surfavg
L_transversetemporal_surfavg	R_transversetemporal_surfavg
L_insula_surfavg	R_insula_surfavg
ICV	
LLatVent	RLatVent
Lthal	Rthal
Lcaud	Rcaud
Lput	Rput
Lpal	Rpal
Lhippo	Rhippo
Lamyg	Ramyg
Laccumb	Raccumb
BrainSegVol	
BrainSegVolNotVent	
lhCortexVol	rhCortexVol
CortexVol	
lh-WhiteMatterVol	rh-WhiteMatterVol
SubCortGrayVol	
TotalGrayVol	

8.4 Probandenaufwurf am Studienstandort Münster

In den folgenden Abbildungen (s. Abbildung 8-1 und Abbildung 8-2) ist der Probandenaufwurf dargestellt, wie er am Studienstandort Münster ausgelegt wurde. Zum Zeitpunkt der Erstellung dieser Arbeit betrug die aktuelle Vergütung 70 Euro.

Anfahrt



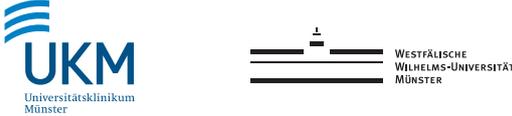
Klinik für Psychiatrie und Psychotherapie
Albert-Schweitzer-Campus 1, Gebäude A9
48149 Münster

Klinik für Psychiatrie und Psychotherapie
Forschungsbereich Translationale Psychiatrie
Prof. Dr. Dr. Udo Dannlowski
Albert-Schweitzer-Campus 1, Gebäude A9
48149 Münster

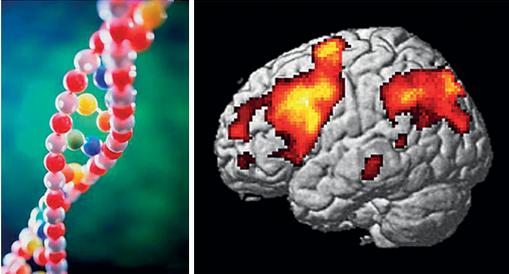
Wenn Sie sich für eine Teilnahme an unserer Studie interessieren, dann melden Sie sich unter:
mrt.studie@uni-muenster.de, T 0251 83-57215

Weitere Informationen erhalten Sie auf unserer Homepage: www.FOR2107.de

Impressum
Herausgeber: Universitätsklinikum Münster, Unternehmenskommunikation
T 0251 83-55866, unternehmenskommunikation@ukmuenster.de
Quelle Titelfotos: FOR2107



Klinik für Psychiatrie und Psychotherapie

Probandenaufwurf

Wir suchen Gesunde und Patienten mit Depression, bipolarer Störung, Schizophrenie oder schizoaffektiver Störung.





2015-0491-G

Abbildung 8-1: Probandenaufwurf am Studienstandort Münster (Teil 1)

Hintergrund dieses Projektes

Neurobiologie psychischer Störungen

Das Auftreten von affektiven und psychotischen Störungen wird durch das Erbgut und Umweltfaktoren bestimmt, die das Erkrankungsrisiko erhöhen können (z.B. erhöhter Stress in der Kindheit).

Ob nun die Erkrankung ausbricht oder nicht, erklären wir uns durch ein Wechselspiel von verschiedenen Genen und Umweltfaktoren. Anhand unserer groß angelegten Untersuchung wollen wir neue Erkenntnisse bezüglich dieses Wechselspiels und des Auftretens von weitverbreiteten psychischen Störungen gewinnen.

Wen suchen wir?

- Probanden ohne psychiatrische Vorerkrankung
- Patienten, die aktuell oder in der Vergangenheit an
 - einer Depression
 - einer bipolaren Störung
 - Schizophrenie
 - einer schizoaffektiven Störung erkrankt sind oder waren
- Probanden, die zwischen 18 und 65 Jahre alt sind

Sie können leider nicht teilnehmen wenn Sie...

- neurologische Vorerkrankungen (z.B. Schlaganfall, Epilepsie) haben
- Metall im Körper haben
- aktuell oder in der Vergangenheit alkohol- oder drogenabhängig sind oder waren

Was wünschen wir uns?

Im Rahmen der Studienteilnahme würden Sie folgende Schritte durchlaufen:

1. Interview und Fragebögen
2. Untersuchung von Gehirnfunktion und -struktur mittels Magnetresonanztomographie (MRT)
3. Neuropsychologische Testung (z.B. Gedächtnis und Sprache)
4. Blutentnahme und die Abgabe von weiteren Proben (z.B. Speichel, Haare)

Die gesamten Untersuchungen sind gesundheitlich völlig unbedenklich.

Zeitaufwand

Für die Teilnahme an der Studie sollten Sie etwa 4 Stunden einplanen (plus 1 Stunde für das Ausfüllen von Fragebögen zu Hause). Darüber hinaus besteht die Möglichkeit, an einer zweiten Erhebung nach 2 Jahren teilzunehmen, welche erwünscht, aber nicht verpflichtend ist.

Aufwandsentschädigung

Nach erfolgreicher Teilnahme erhalten Sie 50 € sowie – falls gewünscht – eine Aufnahme Ihres eigenen Gehirns.

Abbildung 8-2: Probandenaufruf am Studienstandort Münster (Teil 2)

8.5 Probandenaufwurf am Studienstandort Marburg

Der entsprechende Probandenaufwurf für den Studienstandort Marburg kann den folgenden Abbildungen (s. Abbildung 8-3 und Abbildung 8-4) entnommen werden.



Universitätsklinik für Psychiatrie und
Psychotherapie
Rudolf-Bultmann-Straße 8
35039 Marburg

Weitere Informationen erhalten Sie auf unserer Homepage:
<http://www.FOR2107.de>



Wir freuen uns von Ihnen zu hören!



Wir suchen

Patienten mit Schizophrenie,
bipolarer oder
schizoaffektiver Erkrankung



©Tiefen: Fotolia.com - contrastwerkstatt ©PA

Abbildung 8-3: Probandenaufwurf am Studienstandort Marburg (Teil 1)

Studie an der Universitätsklinik Marburg zum Thema „Neurobiologie psychischer Erkrankungen“

Für unsere Studie suchen wir **Patienten mit Schizophrenie, bipolarer oder schizoaffektiver Erkrankung**.

Für die Teilnahme müssen Sie zwischen 18 und 65 Jahre alt sein.

Hintergrund dieses Projektes

Das Auftreten von affektiven und psychotischen Erkrankungen wird durch das Erbgut und Umweltfaktoren bestimmt, die das Erkrankungsrisiko erhöhen können (z.B. erhöhter Stress in der Kindheit). Ob nun die Erkrankung ausbricht oder nicht, erklären wir uns durch ein Wechselspiel von verschiedenen Genen und Umweltfaktoren. Anhand unserer groß angelegten Untersuchung wollen wir neue Erkenntnisse bezüglich dieses Wechselspiels und des Auftretens von weitverbreiteten psychischen Störungen gewinnen.

Was wünschen wir uns von Ihnen?

Im Rahmen der Studienteilnahme würden Sie folgende Schritte durchlaufen:

- Interview und Fragebögen
- Untersuchung von Gehirnfunktion und -struktur mittels Magnetresonanztomographie (MRT)
- Neuropsychologische Testung (z.B. Gedächtnis und Sprache)
- Blutentnahme und die Abgabe von weiteren Proben (z.B. Speichel, Haare)

Die gesamten Untersuchungen sind völlig unbedenklich (keine radioaktive Strahlung, keine Medikamente, kein Kontrastmittel, etc.).

Zeitaufwand

Für die Teilnahme an der Studie sollten Sie etwa 3 Stunden einplanen (plus 1 Stunde für das Ausfüllen von Fragebögen, gerne auch zu Hause). Eine identische, zweite Erhebung erfolgt nach 2 Jahren, eine Teilnahme an dieser Wiederholungsuntersuchung ist ebenfalls gewünscht, aber natürlich nicht verpflichtend.

Aufwandsentschädigung

Nachdem Sie erfolgreich an der Studie teilgenommen haben, erhalten Sie **50 €** und – falls gewünscht – ein Bild von Ihrem Gehirn.

Ausschlusskriterien

Sie können nicht teilnehmen, wenn Sie eine aktuelle oder frühere Alkohol- oder Drogenabhängigkeit aufweisen.

Außerdem sind Deutschkenntnisse auf Muttersprachenniveau für die Testungen erforderlich.

Kontakt

Wenn Sie sich für die Teilnahme an unserer Studie interessieren, melden Sie sich bitte für weitere Informationen und zur Terminabsprache unter:

E-Mail: info@for2107.de
Telefon: 06421 – 58 67048

Abbildung 8-4: Probandenaufruf am Studienstandort Marburg (Teil 2)

8.6 Screening-Fragebogen

Im Folgenden ist der Fragebogen dargestellt, anhand dessen die Möglichkeit zum Studieneinschluss geprüft wurde.

Name, Vorname _____ Interviewer _____
 Geburtsdatum _____ Datum _____
 Körpergewicht _____
 Körpergröße _____
 Telefon _____
 Mail _____

- 1) Sind oder waren Sie jemals in psychotherapeutischer oder psychiatrischer Behandlung oder Beratung?
 Ja Nein

Falls ja, aufgrund welcher Beschwerden? (*Unipolar, soziale Phobie → TIP Screening*)

Beurteile: Falls nein, Proband unter 40 Jahren?

- Ja (→ Risiko-Screening, TIP Screening) Nein

- 2) Welcher Nationalität gehören Sie, Ihre Eltern und Ihre Großeltern an (*Ausschluss bei nicht-kaukasischer Abstammung*)?

- 3) Haben Sie bereits an einer MRT-Studie am Uniklinikum Münster (*v.a. NAE*) teilgenommen?

- Ja Nein

- 4) Sind Sie bereits einmal am Herzen oder am Kopf operiert worden? Hatten Sie sonstige Operationen, bei denen Rückstände im Körper zurückgeblieben sind (*z.B. Gefäßclips, Stents, Prothesen, Schrauben*)?

- Ja Nein

- 5) Befinden sich metallische Objekte in Ihrem Körper (*z.B. Metallstifte, Piercings, Spirale, Metallsplitter*)?

- Ja Nein

- 6) Tragen Sie eine Tätowierung oder „Permanent Make-Up“?

- Ja Nein

- 7) Sind Sie in einem Metall verarbeitenden Beruf tätig oder tätig gewesen oder besteht die Möglichkeit, dass Sie schon einmal von Metallsplittern im Auge getroffen worden sind?

- Ja Nein

- 8) Hatten Sie schon einmal ein Schädelhirntrauma/eine Gehirnerschütterung?

- Ja Nein

- 9) Haben Sie schon einmal für mehr als 10 Minuten das Bewusstsein verloren?

- Ja Nein

- 10) Leiden oder litten Sie jemals unter **Erkrankungen des Gehirns oder Nervensystems** (*z.B. Migräne, Schlaganfall, Hirnhautentzündung, Multiple Sklerose, Alzheimer, Lähmungen, Demenz, Hirntumore*)?

- Ja Nein

- 11) Leiden oder litten Sie jemals unter einer **Autoimmunerkrankung** (*z.B. HIV, Colitis ulcerosa, Hashimoto-Thyreoiditis, Lupus, Morbus Crohn, Narkolepsie, Sarkoidose, Sklerodermie, Sjörgen-Syndrom*)?

- Ja Nein

12) Leiden oder litten Sie jemals unter einer **Erkrankung des Herzens**, einer **Krebserkrankung**, **Tuberkulose**, **Hepatitis B+C** oder einem (instabilen) **Diabetes** (z.B. *Koronare Herzkrankheit, Herzinfarkte, Herzrhythmusstörungen*)?

Ja Nein

13) Haben oder hatten Sie jemals epileptische Anfälle?

Ja Nein

14) Leiden Sie unter Hörschwierigkeiten (z.B. *Tinnitus*)?

Ja Nein

15) Leiden Sie unter Farbblindheit/Rot-Grün-Blindheit?

Ja Nein

16) Tragen Sie eine Sehhilfe (*Kontaktlinsen, Brille*)?

Ja Nein

Falls ja, welche Sehstärke haben Sie? (hat aber auch wegwerfkontaktlinsen)

Links:

Rechts:

17) Leiden Sie unter einer Hornhautverkrümmung? (*Wenn ja, wie stark ist ihre Sehfähigkeit ohne Sehhilfe eingeschränkt?*)

Ja Nein

18) Besteht bei Ihnen die Möglichkeit, eine Haarprobe zu entnehmen (*Ausschluss bei Glatze, Dreadlocks oder Haarlänge unter 3 cm*)?

Ja Nein

19) Leiden Sie unter Platzangst?

Ja Nein

20) Besteht die Möglichkeit, dass Sie schwanger sind?

Ja Nein

21) Nehmen Sie regelmäßig Medikamente ein (*Ausschluss bei Benzodiazepinen*)?

Ja Nein

Falls ja, welche Medikamente?

22) Fühlten Sie sich jemals von einem ärztlich verschriebenen Medikament abhängig oder nahmen Sie mehr davon ein, als Ihnen verschrieben wurde?

Ja Nein

23) Haben Sie jemals Drogen genommen?

Ja Nein

24) Gab es einmal eine Zeit in Ihrem Leben, in der Sie 5 oder mehr Gläser Alkohol (Bier, Wein oder Likör) auf einmal getrunken haben?

Ja Nein

Falls ja, wie würden Sie Ihre Trinkgewohnheiten beschreiben? (Frequenz/Monat, Menge)

Falls ja, Wann in Ihrem Leben haben Sie am meisten getrunken? Während dieser Zeit, wie oft und wie viel haben Sie da getrunken?

Beurteile: Gibt es Hinweise auf eine aktuelle oder frühere Alkoholabhängigkeit?

Ja (→ E7-E13) Nein

25) Wie sind Sie auf unsere Studie aufmerksam geworden?

26) Bekommen Sie in der nächsten Zeit EKT?

Ja Nein

- 27) Hinweis, dass die Auszahlung der Vergütung bis zu 3 Monate in Anspruch nehmen kann.
 erledigt
- 28) Hinweis, dass wir verpflichtet sind, Zufallsbefunde im MRT an den Probanden zu melden. Von Neuropsych, SKID & Fragebögen können jedoch keine Ergebnisse zurückgemeldet werden, da diese in anonymisierter Form verarbeitet werden.
 erledigt
- 29) Hinweis, dass MRT-Termine sehr teuer sind und deshalb möglichst nicht kurzfristig abgesagt werden sollten (Ausnahmen sind natürlich, wenn der Proband plötzlich erkrankt, einen Unfall hat usw.)
 erledigt

Screening auf Substanzabhängigkeit

- E7) Ist es schon öfter vorgekommen, dass Sie viel mehr SUBSTANZ konsumiert haben, als Sie es sich eigentlich vorgenommen hatten? (...oder über einen längeren Zeitraum, als vorgenommen)
 Ja Nein
- E8) Haben Sie jemals versucht, den Konsum von SUBSTANZ zu reduzieren oder ganz einzustellen? (anhaltender Wunsch oder erfolglose Versuche zu reduzieren / einzustellen)
 Ja Nein
- E9) Haben Sie schon einmal viel Zeit damit verbracht SUBSTANZ zu konsumieren oder zu beschaffen? Brauchten Sie lange, um wieder zu Ihrem normalen Zustand zurückzufinden?
 Ja Nein
- E10) Gab es Zeiten, in denen Sie SUBSTANZ so oft genommen haben, dass Sie deswegen nicht mehr arbeiten oder Ihre Zeit mit Hobbies, Ihrer Familie oder Freunden verbringen konnten?
 Ja Nein
- E11) Verursachte SUBSTANZ psychische Probleme, wurden Sie z.B. depressiv? Oder verursachte SUBSTANZ körperliche Probleme oder verschlimmerte sich durch SBSTANZ ein körperliches Problem? (=körperliche/psychische Folgeschäden)
 Ja Nein
- E12) Haben Sie bemerkt, dass Sie deutlich mehr von SUBSTANZ nehmen mussten, um die gleiche Wirkung zu erzielen als zu der Zeit in der Sie anfangen? Ka es auch vor, dass Sie durch die gleiche Dosis wie früher eine geringere Wirkung erzielten? (= Toleranzentwicklung)
 Ja Nein
- E13) Hatten Sie jemals körperliche oder psychische Entzugserscheinungen, d.h. fühlten Sie sich schlecht, wenn Sie weniger nahmen als gewöhnlich oder ganz damit aufhörten?
 Ja Nein
- Beurteile:** Mindestens 3 Kriterien erfüllt? (= Abhängigkeit)
 Ja Nein

8.7 Studienaufklärung und Einverständniserklärung (Allgemein)

Die genannten Dokumente sind beispielhaft für den Standort Münster dargestellt.



Projektleiter FOR 2107:
Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski
 Klinik für Psychiatrie und Psychotherapie Universitätsklinikum Münster
 Albert-Schweitzer-Campus 1, Gebäude A9, 48149 Münster
 Tel: 0251-83-56601

Probanden-Aufklärung für ein Forschungsvorhaben mit neuropsychologischen und MRT-Untersuchungen

**zur Vorbereitung der mündlichen Aufklärung durch den Arzt oder Psychologen für die Studie:
 „Neurobiologie affektiver Störungen:
 Eine translationale Perspektive auf Hirnstruktur und -funktion“**

Sehr geehrte Probandin, sehr geehrter Proband,

wir möchten Sie bitten, an o.g. Studie teilzunehmen. In diesem Informationsblatt geben wir Ihnen im Folgenden Hintergrundinformationen zur Studie, beschreiben die geplanten Untersuchungen, klären über mögliche Risiken auf, beschreiben den Umgang mit sog. „Zufallsbefunden“, konkretisieren den Umgang mit den erhobenen Daten und nennen abschließend Ausschlusskriterien für die Teilnahme an der Studie. Über die Sammlung von Biomaterialien in einer Biodatenbank werden sie gesondert aufgeklärt und erhalten eine gesonderte Einwilligungserklärung.

1. Inhalt und Zweck der Studie:

Der Hintergrund der Untersuchung sind die zunehmenden Befunde von speziellen Varianten im Erbgut (Kandidatengene), sowie bestimmten Umwelteinflüssen, welche die Auftretenswahrscheinlichkeit von psychischen Störungen modulieren können. Schon seit längerer Zeit weiß man, dass das Auftreten von Depressionen oder Bipolarer Störung in bedeutsamer aber nicht vollständiger Weise durch das Erbgut

mitbestimmt wird. Zusätzlich sind einige Umweltfaktoren wie etwa erhöhter Stress in der Kindheit, bekannt, die das Risiko zu erkranken erhöhen können. Die Schwierigkeit dieser Ergebnisse liegt darin, dass einerseits nicht alle Patienten mit diesen Erkrankungen jene Genvarianten haben oder Umwelteinflüsse erlebten und andererseits die Mehrzahl der Träger dieser Varianten oder Umwelteinflüsse psychisch gesund sind. Wir erklären dies durch ein hochkomplexes Wechselspiel von ganz verschiedenen Kandidatengenen und Umweltfaktoren.

Diese Untersuchung findet statt im Rahmen einer von der Deutschen Forschungsgemeinschaft geförderten Studie zur Untersuchung von Effekten genetischer Risikovarianten und verschiedener Umweltfaktoren auf Hirnfunktionen sowie die Hirnstruktur. Zusätzlich soll untersucht werden, wie sich die o.g. Faktoren auf eine Reihe von sog. Biomarkern auswirken. Zu diesen Biomarkern gehören z.B. verschiedene Hormone, Proteine, Immunfaktoren im Blut und Gehirnparameter, aber auch die Frage, inwieweit sich das Vorliegen oder die Abwesenheit von bestimmten Umweltfaktoren auf die Genexpression (also wie häufig ein Gen „abgelesen“ wird) und andere Faktoren auswirkt. Im Rahmen dieser Studie findet neben mehreren Verhaltenstests auch eine Untersuchung im Kernspintomographen (MRT) statt. Wir möchten damit untersuchen, welchen Einfluss Genvarianten und Umweltfaktoren, die sich als bedeutsam für psychische Erkrankungen erwiesen haben, auf Erleben und Verhalten haben und wie sie zur Entstehung von psychischen Erkrankungen beitragen können. Ziel der Studie ist es, ein besseres Verständnis über die Entstehung und den Verlauf von affektiven Störungen (Depression und Bipolare Störung) zu erhalten, das sich auf Erleben und Verhalten, Hirnmarker sowie weitere Biomarker stützt. Dabei ist es wichtig, auch zu untersuchen, wie sich diese Parameter im zeitlichen Verlauf verändern. Aus diesem Grund erklären Sie sich bereit, an einer weiteren Untersuchung mit denselben Untersuchungsmethoden in 2 Jahren erneut teilzunehmen, für die wir Sie kontaktieren werden. Die Untersuchungen nach 2 Jahren sind identisch mit denen für die aktuelle Teilnahme. Die Einwilligung zu dieser Studie zu diesem Zeitpunkt setzt jedoch nicht automatisch voraus, dass Sie auch zu einem späteren Zeitpunkt erneut teilnehmen müssen. Für die Teilnahme an der Studie zahlen wir Ihnen eine Aufwandsentschädigung von 70 Euro.

2. Betroffener Personenkreis

Für das Studienvorhaben werden insbesondere Patienten gesucht, die entweder aktuell oder in der Vergangenheit an einer psychischen Störung (d.h. Depression, Bipolare Störung, Schizophrenie) erkrankt sind bzw. waren. Außerdem werden gesunde Kontrollprobanden (mit und ohne genetischem bzw. umweltbedingtem Risiko für eine psychische Erkrankung) als Vergleichsgruppe eingeschlossen. Im Rahmen der Studie werden auch anonymisierte Informationen zu Verwandten (z.B. psychische Erkrankungen im Familienkreis) erfragt.

3. Zu erhebende Daten

- a) *Wiederholte Untersuchungen zu neuropsychologischen Variablen, Persönlichkeitsfaktoren sowie relevanten Umwelteinflüssen*
- b) *Klinisches Interview*
- c) *Wiederholte Untersuchungen im Kernspintomographen*

a) Neuropsychologische Untersuchung, Persönlichkeit und Umwelteinflüsse

Es findet eine neuropsychologische Testung statt, bei der u.a. Gedächtnisfunktionen, Sprachfertigkeiten und Aufmerksamkeit untersucht werden. Diese Untersuchungen werden insgesamt ca. 50 Minuten dauern. Zusätzlich erhalten Sie einige Fragebögen zum Ausfüllen, die Ihr generelles Erleben und Verhalten erfassen. Z.B. wird erfragt, ob Sie generell eher ängstlich, fröhlich, offen im Kontakt mit Mitmenschen etc. sind. Diese Untersuchung wird ebenfalls ca. 30 Minuten dauern. Weiterhin erhalten Sie Fragebögen, mit denen wichtige Umwelteinflüsse wie etwa traumatische Erlebnisse in Ihrer Kindheit und sonstige wichtige Lebensereignisse erfragt werden. Die Bearbeitung dieser Fragebögen dauert ebenfalls ca. 30 Minuten. Die Untersuchung der zeitlich veränderbaren Variablen wird in zwei Jahren wiederholt.

b) Klinisches Interview

Im Rahmen der Studie sollen Sie ein Interview führen, wobei wir mit Ihnen anhand einer Reihe von Fragen Ihre aktuelle Krankheitsgeschichte sowie Ihre derzeitigen Symptome abklären werden. Dazu gehört neben den bisherigen stationären Aufenthalten und Behandlungen auch, wie sich Ihre derzeitige Symptomatik darstellt und in welchen Bereichen (z.B. Stimmung, Antrieb, Schlaf etc.) Sie derzeit

besondere Probleme haben. Dieses Interview wird ca. 1 Stunde lang dauern. Das Interview wird in zwei Jahren wiederholt.

c) Funktionelle Untersuchung im Kernspintomographen

Das menschliche Gehirn besteht aus einer Vielzahl von Molekülen, die jeweils spezifische magnetische Eigenschaften aufweisen (sog. Kernspinresonanz). Bei Anwendung von starken Magnetfeldern können diese gemessen und im Anschluss daran die Konzentration von Molekülen im Gehirn festgestellt werden. Auf diese Weise lassen sich Aussagen über die Struktur und die Durchblutung des Gehirns ermitteln, genauer die kontinuierlichen Veränderungen des Sauerstoffgehalts im Blut innerhalb weniger Sekunden. Diese Sauerstoffgehaltsveränderungen ergeben spezifische Muster, je nachdem, welche Sinnesreize oder Vorstellungen die Versuchsperson während der Messung innerlich verarbeitet. Auf diese Weise können unterschiedliche Funktionen des Gehirns sichtbar gemacht werden. Man spricht hier von funktioneller Magnetresonanztomographie (fMRT). Um bei den Messungen verlässliche Ergebnisse zu erhalten, ist Ihre Kooperationsbereitschaft erforderlich. Sie liegen während der Untersuchung auf einem Tisch, der Sie mit dem Oberkörper in die Öffnung des MR-Tomographen hineinführt. Zusätzlich wird ein Magnetspulenrahmen, ähnlich einem Helm, um Ihren Kopf gelegt. Während der Messung entstehen unterschiedlich laute Geräusche des MR-Tomographen, die von elektrischen Umschaltvorgängen der Magnetfelder herrühren. Um die Geräusche zu dämpfen, tragen Sie einen Kopfhörer. Sie haben während der Untersuchung jederzeit die Möglichkeit, mit den Untersuchern über eine Wechselsprechanlage in Kontakt zu treten. Zusätzlich bekommen Sie einen gesonderten Alarmknopf für Notfälle, so dass sie auf Ihren Wunsch hin jederzeit aus dem MR-Tomographen herausgefahren werden können.

Mögliche Gefahren: Die MRT-Technologie ist für den Körper nach heutigem Erkenntnisstand unschädlich. Sie basiert auf mehr als 25 Jahren Erfahrung und wird täglich in allen größeren Kliniken eingesetzt. Bekannte Risiken ergeben sich ausschließlich durch metallische Gegenständen oder Stoffe mit magnetischen Eigenschaften, die sich am oder im Körper befinden. Diese können sich erhitzen und zu Verbrennungen führen. Lose Metallteile können durch das Magnetfeld beschleunigt werden und dann zu Verletzungen führen. Daher sind Personen von

der Teilnahme an der Studie ausgeschlossen, die elektrische Geräte (z.B. Herzschrittmacher, Medikamentenpumpen) oder Metallteile (z.B. Schrauben nach Knochenbruch) im oder am Körper haben.

Risiken unabhängig von den Genannten sind bislang nicht bekannt. Abgesehen von möglichen Unbequemlichkeiten, die vom stillen Liegen in der engen MRT-Röhre resultieren, und der Lautstärke des Tomographen, sollten keine Beschwerden während der Untersuchung auftreten.

Wir möchten Sie allerdings darauf hinweisen, dass über mögliche langfristige Risiken bei wiederholten MRT-Messungen bisher keine wissenschaftlich abgesicherten Ergebnisse vorliegen.

Die Untersuchung im Kernspintomographen gliedert sich in mehrere Teile, die einzelnen Aufgaben werden Ihnen vorher ausführlich erklärt:

1. In den ersten 17 Minuten werden strukturelle Messungen durchgeführt. Dabei ist keine Aufgabe zu bearbeiten – bitte bleiben Sie einfach ruhig liegen. Wenn Sie möchten, können Sie in dieser Zeit Bilder auf dem Monitor betrachten oder Radio hören. Sie können Ihre Augen geöffnet oder geschlossen halten.
2. Wir zeigen Ihnen nacheinander eine Reihe von jeweils drei Gesichtern. Diese können einen neutralen, ängstlichen, fröhlichen oder traurigen Ausdruck haben. Ihre Aufgabe besteht darin, durch einen Tastendruck eines von zwei Gesichtern einem dritten zuzuordnen, das mit diesem identisch ist. Diese Aufgabe dauert ca. 6 Minuten.
3. In der darauf folgenden Bewertungsaufgabe werden Ihnen für kurze Zeit Gesichter gezeigt. Zwischen den – oberflächlich betrachtet – neutralen Gesichtern bestehen sehr feine Ausdrucksunterschiede. Achten Sie genau auf den Gesichtsausdruck. Ihre Aufgabe ist es, so schnell wie möglich zu entscheiden, ob und wie stark ein Gesicht ein positives oder negatives Gefühl ausdrückt. Diese Aufgabe dauert ca. 6 Minuten.
4. Die dritte Aufgabe umfasst 30 Gesichter sowie 30 verpixelte Gesichter. Ihre Aufgabe besteht darin, sich die Gesichter zu merken und das Geschlecht der dargestellten Person durch Tastendruck zu bestimmen. Diese Aufgabe dauert ca. 5 Minuten.

5. Anschließend wird Ihr Gehirn in Ruhe untersucht. Dabei müssen Sie keine Aufgabe bearbeiten, sondern mit geschlossenen Augen ruhig liegen. Diese Messung dauert ca. 5 Minuten.

Insgesamt wird diese Untersuchung im Kernspintomographen ca. 40 Minuten dauern.

4. Umgang mit Zufallsbefunden

Bei der Studie handelt es sich um eine Forschungsstudie. Eine neuroradiologische Befundung der MR-Bilder im Sinne einer klinisch orientierten Diagnostik findet daher nicht statt. Dennoch kann es vorkommen, dass in den MR-Bildern Signalauffälligkeiten entdeckt werden, die eine mögliche klinische Relevanz haben („Zufallsbefund“). Falls sich bei der Untersuchung Anhaltspunkte für einen Zufallsbefund ergeben, die eine fachärztliche neuro-radiologische Diagnostik empfehlenswert erscheinen lassen, würde Sie der Versuchsleiter persönlich darüber informieren und Ihnen eine fachärztlich neuroradiologische Diagnostik empfehlen. Weiterhin dienen alle durchgeführten Tests und Fragebögen nicht der Diagnostik von psychischen oder somatischen Störungen oder anderen medizinischen Abklärungen. Trotzdem kann es sein, dass die Auswertung einzelner Verfahren Ergebnisse liefert, die für Ihre Gesundheit von erheblicher Bedeutung sind. In diesem Fall wird ein Studienarzt mit Ihnen das weitere Vorgehen besprechen. Falls Sie über einen Zufallsbefund nicht informiert werden wollen, stellt dies ein Ausschlusskriterium für die Teilnahme an der Studie dar.

5. Lagerung und Pseudonymisierung von Daten

Ihre personenbezogenen Daten werden maschinell gespeichert und weiterverarbeitet. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen ohne Namensnennung und setzt vor Teilnahme an der Studie Ihre freiwillige Einwilligung voraus. Die Zuordnung Ihrer Daten zu Ihrer Person ist nur am Studienzentrum in Münster durch den Studienleiter möglich. Ihre Daten sind zu jeder Zeit pseudonymisiert, d.h. alle Arbeiten an Ihren Daten erfolgen ausschließlich über einen Code. Nur der Studienleiter kann diesen Ihrer Person zuordnen. Pseudonymisierte Studiendaten werden in einer Datenbank des Koordinierungszentrums für Klinische Studien an der Philipps-Universität Marburg gespeichert.

6. Zusammenarbeit mit anderen Forschungsgruppen

Wie oben ausgeführt, dient unsere Studie letztlich der Erforschung der Ursachen und Verlauf sowie der Verbesserung von Diagnostik und Therapie seelischer Erkrankungen. Um diese Forschung effizient zu gestalten, ist eine enge Zusammenarbeit mit wissenschaftlichen Arbeitsgruppen in Marburg, anderen nationalen und internationalen Forschergruppen von essentieller Bedeutung. Dies bedeutet, dass hierfür Information über klinische und MRT-Daten einer möglichst großen Anzahl international forschender Arbeitsgruppen zur Verfügung gestellt werden müssen. Dies geschieht immer gemäß den gesetzlichen Datenschutzrichtlinien und unter Wahrung der Pseudonymisierung. Darüber hinaus möchten wir Daten in pseudonymisierter Form in internationale Datenbanken einbringen. Der Zugang zu diesen Datenbanken ist reglementiert, d.h. die Daten werden nur ausgewählten Wissenschaftlern auf Antrag zeitlich limitiert zur Verfügung gestellt. Die Datenweitergabe wird durch spezielle Gremien streng kontrolliert.

7. Konkrete Dauer der Speicherung

Personenbezogene Daten werden über einen Zeitraum von 10 Jahren aufbewahrt um eine erneute Kontaktaufnahme zu Studienzwecken gewährleisten zu können. Nach Ablauf dieser Frist werden die erhobenen Daten ausschließlich in anonymisierter Form weiterverwendet.

8. Ausschlusskriterien

Die Anwendung von Magnetfeldern bei der MRT-Untersuchung schließt die Teilnahme von Personen aus, die elektrische Geräte (z.B. Herzschrittmacher, Medikamentenpumpen usw.) oder Metallteile (z.B. Schrauben nach Knochenbruch, „Spirale“) im oder am Körper haben. Frauen, die schwanger sind, werden nicht als Probandinnen zugelassen. Weitere Ausschlusskriterien für die Untersuchung sind akute oder chronische neurologische, internistische oder sonstige Erkrankungen. Probanden, die nicht über mögliche Zufallsbefunde informiert werden wollen, können ebenfalls nicht an der Studie teilnehmen.

9. Rechtsgrundlagen

Die Rechtsgrundlage zur Verarbeitung der genannten personenbezogenen Daten bildet die Einwilligung gemäß Art. 6 (1) Buchstabe a EU-DSGVO im zweiten Teil dieses Dokumentes.

10. Widerruf seitens des Betroffenen

Die Teilnahme an dieser Studie ist freiwillig, Sie können zu jeder Zeit und ohne Angabe von Gründen Ihre datenschutzrechtliche Einwilligung zurückziehen, ohne dass Ihnen daraus Nachteile entstehen. Durch den Widerruf der Einwilligung wird die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung nicht berührt. Richten Sie den Widerruf an den Verantwortlichen. Nach Eingang des Widerrufs ist Ihre Teilnahme an der Studie beendet und alle personenbezogenen Daten werden gelöscht. Daten, die im Rahmen von Forschung bereits in Analysten, Statistiken, etc. eingeflossen sind, können i.d.R. nicht rückwirkend herausgenommen werden.

11. Weitere Rechte der Betroffenen

Gemäß Art. 13 II b der Datenschutzgrundverordnung haben Sie das Recht auf

- Auskunft (Art 15 DSGVO und §34 BDSG)
- Widerspruch (Art. 21 DSGVO und §36 BDSG)
- Datenübertragbarkeit (Art 20 DSGVO)
- Löschung (Art 17 DSGVO und §35 BDSG)
- Einschränkung der Verarbeitung (Art 18 DSGVO)
- Berichtigung (Art 16 DSGVO)

Möchten Sie eins dieser Rechte in Anspruch nehmen, wenden Sie sich bitte an den Datenschutzbeauftragten des UKM. Weiterhin haben Sie das Recht, Beschwerde bei der Aufsichtsbehörde einzulegen:

Landesbeauftragte für Datenschutz und Informationsfreiheit
Nordrhein-Westfalen
Postfach 20 04 44
40102 Düsseldorf
Tel.:0211/38424-0

12. Kontaktdaten des Datenschutzbeauftragten

Universitätsklinikum Münster
Datenschutzbeauftragter UKM
Albert-Schweitzer-Campus 1
48149 Münster
Telefon: 0049 (0)251 - 83- 49694
E-Mail: datenschutz@ukmuenster.de

**Wenn weitere Fragen bestehen, werden diese gerne vom jeweiligen
Untersuchungsleiter beantwortet.**



Projektleiter FOR 2107:
Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski
Klinik für Psychiatrie und Psychotherapie Universitätsklinikum Münster
Albert-Schweitzer-Campus 1, Gebäude A9, 48149 Münster
Tel: 0251-83-56601

**Einwilligungserklärung für Probanden zur Teilnahme an dem
Forschungsvorhaben „Neurobiologie affektiver Störungen: Eine
translationale Perspektive auf Hirnstruktur und –funktion“**

Bei Ihrer Bereitschaft zur Teilnahme bitten wir Sie, die Einwilligungserklärung vor der Untersuchung vollständig auszufüllen und zu unterschreiben.

Ich bestätige hiermit, dass ich durch den Untersucher, Herrn/Frau..... über Wesen, Bedeutung, Risiken und Tragweite der beabsichtigten Untersuchung aufgeklärt wurde und für meine Entscheidung genügend Bedenkzeit hatte. Ich habe die Probanden-Aufklärung gelesen. Ich fühle mich ausreichend informiert und habe verstanden, worum es geht. Der Untersucher hat mir ausreichend Gelegenheit gegeben, Fragen zu stellen, die alle für mich ausreichend beantwortet wurden. Ich hatte genügend Zeit, mich zu entscheiden.

Ich wurde darauf hingewiesen, dass es sich bei der Studie um eine Forschungsstudie handelt. Eine individuelle Befundung der erhobenen Daten im Sinne einer klinisch orientierten Diagnostik findet daher nicht statt. Dennoch kann es vorkommen, dass in den MR-Bildern Signalauffälligkeiten oder in weiteren Daten Auffälligkeiten entdeckt werden, die eine mögliche klinische Relevanz haben („Zufallsbefund“). Mir ist bekannt, dass der Versuchsleiter mich informieren würde, falls sich bei der Untersuchung Anhaltspunkte für einen Zufallsbefund ergeben, die eine fachärztliche Diagnostik empfehlenswert erscheinen lassen.

Ich habe verstanden, dass bei wissenschaftlichen Studien persönliche Daten und medizinische Befunde erhoben werden. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen und setzt vor Teilnahme an der Studie meine freiwillige Einwilligung voraus. Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten auf Fragebögen und elektronischen Datenträgern aufgezeichnet und ohne Namensnennung zum Zwecke wissenschaftlicher Auswertung analysiert werden. Ich willige ein, dass meine Daten und Materialien in pseudonymisierter Form an nationale und internationale Datenbanken und Kooperationspartner zu wissenschaftlichen Zwecken übergeben werden können. Ich erkläre mich damit einverstanden, dass ich zur möglichen Teilnahme an der gleichen Forschungsstudie zu einem späteren Zeitpunkt kontaktiert werde. Mir wurde versichert, dass eine kommerzielle Verwendung der Daten nicht stattfinden wird.

Ich habe eine Kopie der Probanden-Aufklärung und dieser unterschriebenen Einwilligungserklärung erhalten. Meine Einwilligung, an diesem Forschungsvorhaben als Proband teilzunehmen, erfolgt freiwillig. Ich wurde darauf hingewiesen, dass ich meine Einwilligung jederzeit ohne Angabe von Gründen und ohne Nachteile widerrufen kann.

Mir ist bewusst, dass ich keinen persönlichen Vorteil aus diesem Forschungsprojekt haben werde und dass mir die Ergebnisse nicht mitgeteilt werden. Nur dann, wenn durch die Auswertung einzelner Tests oder Fragebögen wider Erwarten medizinische Erkenntnisse gewonnen werden, die für die Erhaltung oder Wiederherstellung meiner eigenen Gesundheit von erheblicher Bedeutung sind, wird mein Studienarzt darüber informiert, damit er mit mir das weitere Vorgehen besprechen kann.

Ich willige hiermit ein, als Probandin/Proband an dem Forschungsvorhaben „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und –funktion“ teilzunehmen.

VON DER PROBANDIN / VOM PROBANDEN AUSZUFÜLLEN:

Name:

Geburtsdatum:

Datum:

Uhrzeit:

Ort:

Unterschrift:

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken des Forschungsvorhabens aufgeklärt.

Datum:

Uhrzeit:

Ort:

Untersucher:

Ich willige hiermit ein, als Probandin/Proband an dem Forschungsvorhaben „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und –funktion“ teilzunehmen.

VON DER PROBANDIN / VOM PROBANDEN AUSZUFÜLLEN:

Name:

Geburtsdatum:

Datum:

Uhrzeit:

Ort:

Unterschrift:

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken des Forschungsvorhabens aufgeklärt.

Datum:

Uhrzeit:

Ort:

Untersucher:

8.8 Studienaufklärung und Einverständniserklärung (Bioproben)

Auch die Aufklärung und Einverständniserklärung für die Bioprobenentnahme und -verarbeitung ist im Folgenden beispielhaft für den Studienstandort Münster dargestellt.



Projektleiter FOR 2107:
Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannowski
 Klinik für Psychiatrie und Psychotherapie Universitätsklinikum Münster
 Albert-Schweitzer-Campus 1, Gebäude A9, 48149 Münster
 Tel: 0251-83-56601

Probanden-Aufklärung

bzgl. der Sammlung von Biomaterialien in einer Biobank

zur Vorbereitung der mündlichen Aufklärung durch den Arzt oder Psychologen für die Studie:
 „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und –funktion“

Sehr geehrte Probandin, sehr geehrter Proband,

wir möchten Sie bitten, an o.g. Studie teilzunehmen. In einer gesonderten Aufklärung mit der Überschrift „Probanden-Aufklärung für ein Forschungsvorhaben mit neuropsychologischen und MRT-Untersuchungen“ wurden Sie bereits ausführlich über die Hintergründe der Studie „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und –funktion“ informiert und sind nun als Proband* (*Im Rahmen dieses Textes schließt die männliche Bezeichnung stets die weibliche Bezeichnung mit ein) an der Studie beteiligt. Bei der Biobank handelt es sich um eine Sammlung von menschlichen Biomaterialien wie z.B. Blut, verknüpft mit ausgewählten medizinischen Daten.

Die Untersuchung von menschlichen Biomaterialien und die Analyse der daraus gewonnenen oder zu gewinnenden Daten sind zu einem wichtigen Instrument medizinischer Forschung geworden. Deshalb fragen wir unsere Probanden und daher auch Sie, ob Sie bereit sind, uns bestimmte Körpermaterialien für die Forschung zur Verfügung zu stellen. Ihre Teilnahme ist völlig freiwillig. Soweit Sie

sich nicht beteiligen möchten oder Ihre Zustimmung später widerrufen möchten, erwachsen Ihnen daraus keine Nachteile.

Im Folgenden informieren wir Sie über die Ziele der Biobank, die Verfahrensweisen und die Maßnahmen zum Schutz Ihrer personenbezogenen Daten, damit Sie sich auf dieser Grundlage Ihre eigene Meinung bilden und eine Entscheidung treffen können.

Sollte Ihnen etwas unklar sein, fragen Sie bitte Ihren behandelnden Arzt bzw. Ihren Studienarzt, bevor Sie Ihre Zustimmung erteilen. Sie können sich wegen Rückfragen auch zu einem späteren Zeitpunkt an die Studienleitung Prof. Dr. Dannlowski in der Klinik für Psychiatrie und Psychotherapie der Universitätsklinik Münster, wenden.

1. Ziele der Biobank

Die Untersuchung von menschlichen Biomaterialien und die Analyse der daraus gewonnenen oder zu gewinnenden Daten sind zu einem wichtigen Instrument medizinischer Erkenntnisgewinnung geworden. Die Biobank dient der Förderung medizinischer Forschung. In der Biobank werden Biomaterialien und ausgewählte Daten langfristig aufbewahrt und für die Wissenschaft zur Verfügung gestellt, um die Vorbeugung, Erkennung und Behandlung von affektiven Störungen, d.h. Depressionen und bipolaren Störungen, zu verbessern. In dieser Biobank werden Körpermaterialien der Studienprobanden sowie Probanddaten gesammelt und so aufbereitet und gelagert, dass Sie für Laboranalysen eingesetzt werden können, mit denen wir Wechselwirkungen zwischen genetisch festgelegten Merkmalen und Einflüssen aus der Umwelt untersuchen und besser verstehen können. Diese Untersuchungen können uns helfen, die Ursachen von Depressionen und manisch-depressiven Störungen zu bestimmen. Diese Biobank ist Teil einer Studie zur Entstehungsgeschichte von Psychosen, die in der Göttinger Universitätsmedizin durchgeführt wird (Klinische Forschergruppe 241, Prof. Dr. Schulze). Vom Vergleich der Laboranalysen beider Studien erhoffen wir uns neue Erkenntnisse zu gemeinsamen und speziellen Risikofaktoren und Krankheitswegen beider Erkrankungstypen. Damit kann die Biobank dazu beitragen, neue Wege der Verhinderung wie auch der Behandlung dieser psychischen Erkrankungen zu finden.

2. Um welche Art von Biomaterialien und Daten handelt es sich?

Bei dem von Ihnen erbetenen Biomaterial handelt es sich um Körperflüssigkeiten (Blut, Urin, Speichel), eine Haar- und eine Stuhlprobe sowie Wattebauschabstrich der Mundhöhle, die Ihnen anlässlich Ihrer derzeitigen Studienteilnahme zusätzlich entnommen werden sollen.

- Die *Blutentnahme* (ca. 70 ml), um die wir Sie bitten, wird über eine kleine Kanüle aus Metall erfolgen, wie sie auch im Krankenhausalltag oder beim Hausarzt durchgeführt wird.
- Des Weiteren bitten wir Sie um einen *Mundhöhlenabstrich*, der mit Hilfe eines sterilen Wattetupfers an der Wangeninnenseite abgenommen wird.
- Außerdem bitten wir Sie um *Spontanurin* (Mittelstrahl, ca. 20 ml) und *Stuhl* (ca. 5 g). Zur Gewinnung der Proben erhalten Sie vom Studienpersonal entsprechende sterile Sammelgefäße, die eine einfache Gewinnung der Materialien ermöglichen.
- Ebenfalls bitten wir Sie um eine *Speichelprobe*. Die Speichelprobe gewinnen wir mittels einer Salivette. Dabei handelt es sich um einen Kaugummi ähnlichen sterilen Kunststoffpfropfen, den Sie ca. 2 min kauen und anschließend in ein kleines Plastikbehältnis spucken.
- Ebenso erbitten wir eine *Haarprobe*. Hierzu schneidet Ihnen das Studienpersonal bei 20-50 Haaren etwa 2 cm lange Stückchen aus dem hinteren Haupthaar ab.

Die Daten, die von Ihnen im Rahmen dieser Biobank gesammelt werden, umfassen Daten, die sich auf die von Ihnen gewonnen Proben beziehen und die ausschließlich pseudonymisiert in der Biodatenbank vorliegen. Somit ist es den Mitarbeitern der Biobank nicht möglich genetische Marker wie Risikogene und krankheitsrelevante Daten wie Entzündungsmarker zusammenzuführen und unabhängig zu nutzen. Da die Biobank die Proben ausschließlich pseudonymisiert erhält und verarbeitet, liegen der Biobank auch keine identifizierenden Informationen zu Ihrer Person vor.

Im Rahmen der Studie sind weitere Folgeuntersuchungen geplant, zu denen Sie in entsprechenden Zeitintervallen eingeladen werden. Auch an diesen zukünftigen Folgeuntersuchungen möchten wir Sie bitten, uns die oben genannten Körpermaterialien zur Verfügung zu stellen.

3. Wie werden die Biomaterialien und Daten verwendet?

Die von Ihnen zur Verfügung gestellten Biomaterialien und Daten werden ausschließlich für die Erforschung psychiatrischer Erkrankungen verwendet. Die genauen Fragestellungen können jedoch zum derzeitigen Zeitpunkt noch nicht alle konkret benannt werden. An Ihren Biomaterialien werden möglicherweise auch genetische Untersuchungen, also Untersuchungen der Erbsubstanz, durchgeführt, und zwar unter Umständen auch eine Untersuchung Ihres gesamten Genoms. Die Biomaterialien und Daten sollen für 10 Jahre aufbewahrt und für die medizinische Forschung bereitgestellt werden. Aus logistischen Gründen ist es der Biobank nicht möglich, individuelle Eingrenzungen (z.B. Ausschluss bestimmter Forschung, Ausschluss der Weitergabe der Materialien an Dritte) vorzunehmen. Wenn Sie mit der beschriebenen Art und Dauer der Nutzung nicht in vollem Umfang einverstanden sind, werden Ihre Biomaterialien und Daten nicht für die Biobank verwendet.

4. Welche Risiken sind mit Ihrer Spende verbunden?

a. Gesundheitliche Risiken:

Wir möchten Ihnen 70 ml Blut abnehmen (das entspricht etwa 10 Esslöffeln). Ihre Bereitschaft, Blut für die Biobank zu spenden, ist für Sie nur mit den geringen Risiken einer normalen Blutentnahme verbunden. Die Blutentnahme wird über eine kleine Kanüle aus Metall erfolgen, wie sie auch im Krankenhausalltag oder beim Hausarzt durchgeführt wird. Es handelt sich hierbei um einen in der Medizin täglich praktizierten Routineeingriff, der mit sehr geringen Risiken verbunden ist. Dieses besteht im Einzelnen aus einem Blutungsrisiko und einem Infektionsrisiko. Da die Kanüle nur wenige Sekunden in der Vene liegt, ist das Infektionsrisiko außerordentlich gering. Auch das Blutungsrisiko ist bei der Punktion einer Vene extrem gering, so wie bei einer Blutentnahme. Wegen des geringen Blutdrucks einer Vene besteht hier lediglich das Risiko ungefährlicher Einblutungen (Bluterguss).

Für die Entnahme der weiteren Körpermaterialien ergeben sich für keine Risiken.

b. Weitere Risiken:

Die im Rahmen der Biobank aus Ihren Proben erhaltenen Daten werden an die Studienleitung zur Auswertung weitergeleitet. Wie bei jeder Erhebung, Speicherung und Übermittlung von Daten aus Biomaterialien im Rahmen von Forschungsprojekten können trotz mehrfacher Pseudonymisierung Vertraulichkeitsrisiken entstehen (z.B. die Möglichkeit, Sie zu identifizieren), insbesondere im Hinblick auf die Information zu Ihrer Erbsubstanz. Diese Risiken lassen sich nicht

völlig ausschließen und steigen, je mehr Daten miteinander verknüpft werden können, insbesondere auch dann, wenn Sie selbst (z.B. zur Ahnenforschung) genetische Daten im Internet veröffentlichen. Das Studienteam versichert Ihnen, alles nach dem Stand der Technik Mögliche zum Schutz Ihrer Privatsphäre zu tun und Proben und Daten nur an Projekte weiterzugeben, die ein geeignetes Datenschutzkonzept vorweisen können (siehe Punkt 8: „Wer hat Zugang zu Ihren Biomaterialien und Daten?“).

5. Welcher Nutzen ergibt sich für Sie persönlich?

Persönlich können Sie für Ihre Gesundheit keinen unmittelbaren Vorteil oder Nutzen aus der Spende Ihrer Proben und Daten erwarten. Die Ergebnisse sind ausschließlich zu Forschungszwecken bestimmt.

6. Welcher Nutzen ergibt sich für unsere Gesellschaft?

Alle derzeit durchgeführten wie auch künftige medizinisch-wissenschaftliche Forschungsvorhaben zielen auf eine Verbesserung unseres Verständnisses für die Krankheitsentstehung und die Diagnosestellung und auf dieser Basis auf die Neuentwicklung von verbesserten Behandlungsansätzen.

7. Welche Voraussetzungen gelten für die Nutzung Ihrer Biomaterialien und Daten?

- a. Voraussetzung für die Gewinnung und Nutzung Ihrer Biomaterialien samt den zugehörigen personenbezogenen Daten zu Forschungszwecken ist Ihre schriftliche Einwilligung. Ihre Einwilligung ist freiwillig und kann jederzeit widerrufen werden (siehe auch Punkt 12 „Was beinhaltet Ihr Widerrufsrecht?“).
- b. Ihre Biomaterialien und Daten werden in der Biobank unter standardisierten Qualitäts- und Sicherheitsbedingungen aufbewahrt. Sie werden nach dem Stand der Technik vor unbefugtem Zugriff gesichert.
- c. Voraussetzung für die Verwendung der Biomaterialien und Daten für ein konkretes medizinisches Forschungsprojekt ist, dass das Forschungsvorhaben durch eine Ethikkommission bewertet wurde. Die Ethikkommission hat die ethischen und rechtlichen Aspekte des Forschungsprojekts geprüft und für gut befunden.
- d. Mit der Überlassung der Biomaterialien an die Biobank bleiben Sie als Spender Eigentümer, ermächtigen aber die Klinik für Psychiatrie und Psychotherapie der

Universitätsklinik Münster, Ihre Daten zu nutzen. Dabei haben Sie das Recht, fehlerhaft gespeicherte oder bearbeitete Daten jederzeit zu berichtigen.

8. Wer hat Zugang zu Ihren Biomaterialien und Daten und wie werden sie geschützt?

a. Alle unmittelbar Ihre Person identifizierenden Daten (Name, Geburtsdatum, Anschrift etc.) werden unverzüglich nach ihrer Gewinnung durch einen Code ersetzt (pseudonymisiert, verschlüsselt). Danach wird der Datensatz nochmals neu kodiert und gespeichert. Diese doppelte Kodierung schließt eine Identifizierung Ihrer Person durch Unbefugte nach heutiger Kenntnis weitestgehend aus. Erst in dieser Form werden die Biomaterialien und Daten für Forschungszwecke zur Verfügung gestellt.

b. Die Sie unmittelbar identifizierenden Daten werden völlig getrennt von den Biomaterialien und medizinischen Daten gespeichert. Ein Zugriff auf sie erfolgt nur, um entweder weitere oder fehlende medizinische Daten aus Ihren Krankenunterlagen zu ergänzen oder erneut mit Ihnen in Kontakt zu treten (siehe Punkt 10).

c. Die doppelt verschlüsselten Biomaterialien und medizinischen Daten können auf Antrag für medizinische Forschungszwecke nach zuvor festgelegten Kriterien an andere Universitäten und Forschungsinstitute, ggf. auch im Ausland weitergegeben werden. Dabei werden die Daten unter Umständen auch mit medizinischen Datensätzen in anderen Datenbanken verknüpft, sofern die gesetzlichen Voraussetzungen hierfür erfüllt sind.

d. Biomaterialien und Daten, die an Dritte weitergegeben wurden, dürfen nur für den beantragten Forschungszweck verwendet und vom Empfänger nicht zu anderen Zwecken weitergegeben werden. Nicht verbrauchtes Material wird an die Biobank zurückgegeben oder vernichtet.

e. Wissenschaftliche Veröffentlichungen von Ergebnissen erfolgen ausschließlich anonymisiert, also in einer Form, die keine Rückschlüsse auf Ihre Person zulässt.

9. Erlangen Sie einen finanziellen Vorteil aus der Nutzung Ihrer Biomaterialien und Daten?

Für die Überlassung Ihrer Biomaterialien und Daten erhalten Sie kein Entgelt. Für die Bereitstellung der Proben und Daten kann die Biobank von den Nutzern eine Aufwandsentschädigung erheben.

10. Erfolgt eine erneute Kontaktaufnahme mit Ihnen?

Ziel der Studie ist es, ein besseres Verständnis über die Entstehung und den Verlauf von affektiven Störungen (Depression und Bipolare Störung) zu erhalten, das sich auf Erleben und Verhalten, Hirnmarker sowie weitere Biomarker stützt. Dabei ist es wichtig, auch zu untersuchen, wie sich diese Parameter im zeitlichen Verlauf verändern. Aus diesem Grund erklären Sie sich bereit, an einer weiteren Untersuchung mit denselben Untersuchungsmethoden in 2 Jahren erneut teilzunehmen, für die wir Sie kontaktieren werden.

Eine erneute Kontaktaufnahme und Rückmeldung kann außerdem erfolgen, wenn eine Information für Ihre Gesundheit aus unserer Sicht von erheblicher Bedeutung ist. Dies ist dann der Fall, wenn der Ausbruch einer (ggf. lebensbedrohlichen) Krankheit vermieden oder eine bisher möglicherweise nicht erkannte Krankheit behandelt werden kann. Unter Umständen müssen Sie die Information dann allerdings bei anderen Stellen (z.B. vor Abschluss einer Versicherung) offenbaren.

11. Rechtsgrundlagen

Die Rechtsgrundlage zur Verarbeitung der genannten personenbezogenen Daten bildet die Einwilligung gemäß Art. 6 (1) Buchstabe a EU-DSGVO im zweiten Teil dieses Dokumentes.

12. Was beinhaltet Ihr Widerrufsrecht?

Sie können Ihre Einwilligung zur Verwendung Ihrer Biomaterialien und Daten jederzeit ohne Angabe von Gründen und ohne nachteilige Folgen für Sie widerrufen. Richten Sie den Widerruf an den Verantwortlichen. Im Falle eines Widerrufs können Sie entscheiden, ob Ihre Biomaterialien vernichtet und die dazu gehörenden Daten gelöscht werden sollen oder ob sie in anonymisierter Form für weitere Forschungsvorhaben verwendet werden dürfen. Sobald der Bezug der Biomaterialien und der übrigen Daten zu Ihrer Person gelöscht wurde, ist eine Vernichtung jedoch nicht mehr möglich. Zudem können Daten aus bereits durchgeführten Analysen nicht mehr entfernt werden.

13. Weitere Rechte der Betroffenen

Gemäß Art. 13 II b der Datenschutzgrundverordnung haben Sie das Recht auf

Auskunft (Art 15 DSGVO und §34 BDSG)

Widerspruch (Art. 21 DSGVO und §36 BDSG)

Datenübertragbarkeit (Art 20 DSGVO)
Löschung (Art 17 DSGVO und §35 BDSG)
Einschränkung der Verarbeitung (Art 18 DSGVO)
Berichtigung (Art 16 DSGVO)

Möchten Sie eins dieser Rechte in Anspruch nehmen, wenden Sie sich bitte an den Datenschutzbeauftragten des UKM. Weiterhin haben Sie das Recht, Beschwerde bei der Aufsichtsbehörde einzulegen:

Landesbeauftragte für Datenschutz und Informationsfreiheit
Nordrhein-Westfalen
Postfach 20 04 44
40102 Düsseldorf
Tel.:0211/38424-0

14. Kontaktdaten des Datenschutzbeauftragten

Universitätsklinikum Münster
Datenschutzbeauftragter UKM
Albert-Schweitzer-Campus 1
48149 Münster
Telefon: 0049 (0)251 - 83- 49694
E-Mail: datenschutz@ukmuenster.de

Wenn weitere Fragen bestehen, werden diese gerne vom jeweiligen Untersuchungsleiter beantwortet.



Projektleiter FOR 2107:
Univ.-Prof. Dr. med. Dr. phil. Dipl.-Psych. Udo Dannlowski
 Klinik für Psychiatrie und Psychotherapie Universitätsklinikum Münster
 Albert-Schweitzer-Campus 1, Gebäude A9, 48149 Münster
 Tel: 0251-83-56601

Einwilligungserklärung für Probanden

bzgl. der Sammlung von Biomaterialien in einer Biobank

**im Forschungsvorhaben „Neurobiologie affektiver Störungen: Eine translationale
 Perspektive auf Hirnstruktur und –funktion“**

**Bei Ihrer Bereitschaft zur Teilnahme bitten wir Sie, diese
 Einwilligungserklärung vor der Untersuchung vollständig auszufüllen und zu
 unterschreiben.**

Ich bestätige hiermit, dass ich durch den Untersucher, Herrn/Frau.....
 über Wesen, Bedeutung, Risiken und Tragweite der beabsichtigten Untersuchung
 aufgeklärt wurde und für meine Entscheidung genügend Bedenkzeit hatte. Ich habe
 die Probanden-Aufklärung gelesen. Ich fühle mich ausreichend informiert und habe
 verstanden, worum es geht. Der Untersucher hat mir ausreichend Gelegenheit
 gegeben, Fragen zu stellen, die alle für mich ausreichend beantwortet wurden. Ich
 hatte genügend Zeit, mich zu entscheiden.

Ich erkläre mich bereit, an einer weiteren Untersuchung mit denselben
 Untersuchungsmethoden in zwei Jahren erneut teilzunehmen, für die ich von Ihnen
 kontaktiert werde.

Ich wurde darauf hingewiesen, dass es sich bei der Studie um eine Forschungsstudie
 handelt. Eine individuelle Befundung der Biomaterialien im Sinne einer klinisch
 orientierten Diagnostik findet daher nicht statt. Ich willige jedoch ein, dass eine
 erneute Kontaktaufnahme und Rückmeldung erfolgen kann, wenn eine Information
 für meine Gesundheit aus unserer Sicht von erheblicher Bedeutung ist. Dies ist dann
 der Fall, wenn der Ausbruch einer (ggf. lebensbedrohlichen) Krankheit vermieden

oder eine bisher möglicherweise nicht erkannte Krankheit behandelt werden kann. Mir ist bewusst, dass die Situation eintreten kann, dass ich diese Information dann bei anderen Stellen (z.B. vor Abschluss einer Versicherung) offenbaren muss.

Meine Einwilligung, an diesem Forschungsvorhaben als Proband teilzunehmen, erfolgt freiwillig. Ich wurde darauf hingewiesen, dass ich meine Einwilligung jederzeit ohne Angabe von Gründen und ohne Nachteile widerrufen kann.

Datenschutzerklärung:

Ich erkläre mich damit einverstanden, dass die Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Münster wie in der Information beschrieben

- personenbezogene Daten von mir erhebt und speichert,
- ggf. weitere Angaben über meine Gesundheit aus meinen Krankenunterlagen entnimmt,
- und die Daten gemeinsam mit meinen Biomaterialien pseudonymisiert (das heißt verschlüsselt) an die Biobank für medizinische Forschungsvorhaben zur Verfügung stellt.

Die Biomaterialien und Daten dürfen für 10 Jahre für medizinische Forschungsvorhaben verwendet werden.

Sie dürfen pseudonymisiert an Universitäten oder Forschungsinstitute, ggf. auch im Ausland, zu Zwecken medizinischer Forschung weitergegeben werden.

Ich bin darüber aufgeklärt worden, dass ich meine Einwilligung gegenüber der Biobank ohne Angabe von Gründen jederzeit widerrufen kann. Beim Widerruf werden auf mein Verlangen die verbliebenen Biomaterialien und die erhobenen Daten vernichtet bzw. gelöscht oder anonymisiert. Daten aus bereits durchgeführten Analysen können nicht mehr entfernt werden.

Eine Kopie der Probanden-Aufklärung und Einwilligungserklärung habe ich erhalten.
Das Original verbleibt in der Klinik für Psychiatrie und Psychotherapie des
Universitätsklinikums Münster.

VOM PROBANDEN AUSZUFÜLLEN:

Name:

Geburtsdatum:

Datum:

Uhrzeit:

Ort:

Unterschrift:.....

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken
des Forschungsvorhabens aufgeklärt.

Datum:

Uhrzeit:

Ort:

Untersucher:.....

Eine Kopie der Probanden-Aufklärung und Einwilligungserklärung habe ich erhalten.
Das Original verbleibt in der Klinik für Psychiatrie und Psychotherapie des
Universitätsklinikums Münster.

VOM PROBANDEN AUSZUFÜLLEN:

Name:

Geburtsdatum:

Datum:

Uhrzeit:

Ort:

Unterschrift:.....

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken
des Forschungsvorhabens aufgeklärt.

Datum:

Uhrzeit:

Ort:

Untersucher:.....

8.9 Studienaufklärung und Einverständniserklärung (MRT)

Im Folgenden ist die Aufklärung und die Einverständniserklärung für die MRT Untersuchung beispielhaft für den Studienstandort Münster dargestellt.



Merkblatt für Teilnehmer an einer MR-Untersuchung zu wissenschaftlichen Zwecken

Sehr geehrte Probandin, sehr geehrter Proband!

Wir möchten Sie bitten, an einer Untersuchung teilzunehmen, bei der das Gehirn mittels MR-Bildgebung dargestellt wird. Während der Untersuchung werden Ihnen gegebenenfalls weitere Aufgaben gestellt. Die Messung im MR-Tomographen dauert etwa 30 bis 45 Minuten.

Wenn Sie Ihr Einverständnis erklären, werden Sie in einem Magnetresonanztomographen (auch Kernspintomograph genannt) des Instituts für Klinische Radiologie des Universitätsklinikums Münster bei einer Magnetfeldstärke von 3 Tesla untersucht. Vor Beginn einer Messung werden Sie vom Untersuchungsleiter ausführlich über die geplante Untersuchung informiert.

Für die MR-Messung müssen Sie sich auf eine Liege legen, der Kopf wird in eine Spule gelegt. Mit der Liege werden Sie langsam in die Röhre des Magnetresonanztomographen geschoben. Dort befinden Sie sich während der gesamten MR-Messung. Sie erfolgt in einem starken Magnetfeld; während der Messung werden zusätzliche Hochfrequenzfelder und Magnetfeldgradienten eingeschaltet. Dies bemerken Sie an klopfenden Geräuschen. Während der Untersuchung sollten Sie möglichst ruhig liegen, besonders wenn Sie im Rahmen einer funktionellen Magnetresonanztomographie Aufgaben gestellt bekommen. Um dies zu erleichtern, wird Ihr Kopf mit Polstern fixiert. Während der Messung können Sie sich über eine Klingel (Druckball) und eine Sprechanlage gegebenenfalls mit uns verständigen. Sie können jederzeit aus dem MR-Magneten hinausgefahren werden, wenn Sie dies wünschen.

Vor der Untersuchung müssen Sie alle metallenen Gegenstände ablegen, die Sie sicher in einem Schließfach aufbewahren können.

Ionisierende Strahlen (Röntgenstrahlen oder radioaktive Strahlen) werden im MR-Gerät nicht verwendet. Schädliche Wirkungen der bei der MR-Tomographie verwendeten Magnetfelder auf Körpergewebe sind bei richtiger Durchführung ausgeschlossen. Die Wirkung dieser Felder auf andere Materialien - vor allem aus Metall - können aber erhebliche Gefahren hervorrufen. Daher müssen Sie vor dem Betreten des Messraumes alle Metallgegenstände ablegen und den Mitarbeitern, die die Messung durchführen, vollständig Auskunft geben über mögliche Einschränkungen der Untersuchungsfähigkeit (z.B. falls Implantate vorhanden sind). Sie werden nach möglichen Untersuchungshindernissen gefragt werden. Insbesondere im Untersuchungsraum müssen Sie den Anweisungen der Mitarbeiter genau folgen.

Die bei dieser Messung durchgeführte Bildgebung stellt keine umfassende diagnostische Untersuchung dar. Sollten sich Anhaltspunkte für krankhafte Veränderungen des Gehirns ergeben,

werden wir Sie informieren. Sie haben die Möglichkeit, diese Befunde mit einem Arzt zu besprechen und sich bezüglich weiterer Untersuchungen beraten zu lassen. Sie haben jederzeit das Recht, ohne Angabe von Gründen die Teilnahme abzulehnen oder abzuberechnen. Die ärztliche Schweigepflicht wird gewahrt. Die Bestimmungen des Datenschutzes werden eingehalten.

Bei Fragen oder Unklarheiten fragen Sie bitte nach !

Folgende Punkte müssen besonders beachtet werden:

- Elektronisch gesteuerte Geräte wie Herzschrittmacher, Neurostimulatoren, Insulinpumpen oder Hörhilfen können in ihrer Funktion gestört werden, daher dürfen Sie den Untersuchungsraum nicht mit derartigen Geräten betreten.
- Auf ferromagnetische Gegenstände (z.B. Münzen, Schlüssel, Messer, Haarspangen) wird eine starke Anziehungskraft ausgeübt. Dadurch werden die Gegenstände mit großer Geschwindigkeit in den Magneten gezogen und können Personen im oder am Magneten erheblich verletzen. Diese Gegenstände dürfen nicht in den Untersuchungsraum gebracht werden.
- Metallische Implantate können ebenfalls ferromagnetisch sein, durch magnetische Kräfte können sie ihre Position im Körper verändern und dadurch innere Verletzungen hervorrufen. Auch in nichtmagnetischen Implantaten wie Clips, Stents, künstliche Herzklappen können - insbesondere bei der in diesem Gerät verwendeten hohen Feldstärke - unter Umständen Erwärmungen durch elektrische Induktion auftreten. Über alle bekannten oder möglichen Implantate muss der Arzt oder die Ärztin vor der Untersuchung informiert werden.
- Bei der Messung kommt es zur Abstrahlung von hochfrequenter elektromagnetischer Strahlung, wie sie z.B. bei Radiosendern oder Funktelefonen auftritt. Dies kann zu einer geringfügigen Erwärmung des Körpers führen. Diese Erwärmung ist aber durch technische Sicherheitsmaßnahmen auf ungefährliche Beträge begrenzt.
- In lang gestreckten metallischen Leitern (Kabel, Drähte, Elektroden), aber auch in Karbonfasern können durch die Hochfrequenz sehr hohe Spannungen induziert werden. Daher dürfen keine derartigen Leiter in das MR-Gerät gebracht werden.
- Kreditkarten oder andere kodierte Karten mit Magnetstreifen oder Chips werden im Magnetfeld gelöscht und sollen daher ebenfalls nicht in den Untersuchungsraum gebracht werden.
- Das Schalten der Gradienten kann Geräusche mit Lautstärken bis zu 100 dB erzeugen. Deshalb müssen bei allen Messungen entweder schallabsorbierende Kopfhörer oder Lärmschutz-Ohrstopfen getragen werden, die von uns zur Verfügung gestellt werden. Bei Einhaltung dieser Vorsichtsmaßnahme kann eine Gehörschädigung ausgeschlossen werden.

Aus diesen Gründen dürfen Sie den MR-Untersuchungsraum nicht betreten, wenn Sie

- Träger eines Herzschrittmachers oder anderer elektronisch gesteuerter Hilfsmittel sind
- Ferromagnetische Implantate im Körper haben (auch Metallsplitter durch Unfälle)
- metallische Implantate im Kopfbereich haben (dies gilt unter Umständen auch bei Zahnersatz)
- Implantierte oder extrakorporale Elektroden oder andere Drähte tragen (Sonden, Stimulatorelektroden, EKG-Elektroden und Anschlusskabel bei Langzeit-EKG, auch wenn sie nicht angeschlossen sind)



Institut für Klinische Radiologie • Universitätsklinikum Münster
 Direktor: Univ.-Prof. Dr. med. W. Heindel • 48129 Münster



Fragebogen für Teilnehmer/innen an Messungen mit der Magnetresonanztomographie (MR-Tomographie)

Name

Vorname.....

Geburtsdatum.....

Körpergewicht.....

Straße.....

Körpergröße.....

Wohnort.....

Telefon.....

Beantworten Sie bitte folgende Fragen (Zutreffendes ankreuzen):

Sind Sie Träger eines Herzschrittmachers oder anderer elektronisch gesteuerter Geräte wie Hörhilfen, Neurostimulatoren oder Insulinpumpen?

Ja Nein

Sind Sie bereits einmal am Herzen oder am Kopf operiert worden?

Ja Nein

Befinden sich metallische Objekte in Ihrem Körper (beispielsweise Gefäßclips, Gelenkprothesen, Metallstifte, Zahnschrauben, Piercings, mechanische Verhütungsmittel, Metallsplitter)? Tragen Sie eine Tätowierung oder "Permanent Make-up"?

Ja Nein

Wenn ja, bitte angeben:

Sind Sie in einem Metall verarbeitenden Beruf tätig oder tätig gewesen, oder besteht die Möglichkeit, dass Sie einmal von Metallsplittern im Auge getroffen worden sind?

Sind Sie schon einmal durch Metallteile am Auge verletzt worden (auch geringfügige Verletzungen)?

Ja Nein

Leiden Sie unter Platzangst?

Ja Nein

Besteht die Möglichkeit, dass Sie schwanger sind?

Ja Nein

Sie müssen alle metallenen und magnetischen Gegenstände vor dem Betreten des Untersuchungsraumes ablegen. Sie werden in einem Fach für Sie aufbewahrt.

8.10 Skripte der Machine Learning-Pipelines

Im Folgenden ist das für die PHOTONAI Machine Learning-Pipeline genutzte Skript beispielhaft für eine Klassifikations- sowie eine Regressionsanalyse dargestellt. Die entsprechenden Vor- und Nachbereitungsschritte der Daten sind nicht aufgeführt.

Beispielskript für eine Klassifikationsanalyse:

```

1  # Define features and target variable individually for each analysis
2  X = np.asarray(sample.loc[:, features])
3  y = np.asarray(sample.loc[:, label])
4
5  # Define the individual pipeline (referred to as "hyperpipe")
6  hyperpipe = Hyperpipe(classification_example,
7                        optimizer="grid_search",
8                        metrics=['accuracy', 'balanced_accuracy', 'auc', 'sensitivity', 'specificity',
9                                'f1_score', 'recall', 'precision'],
10                       best_config_metric="balanced_accuracy",
11                       outer_cv=StratifiedKFold(n_splits=10, shuffle=True),
12                       inner_cv=StratifiedKFold(n_splits=5, shuffle=True),
13                       verbosity=0,
14                       output_settings=OutputSettings(
15                           project_folder="...",
16                           mongodb_connect_url="...",
17                           overwrite_results=False,
18                           save_output=True))
19
20 # Add preprocessing elements to the pipeline
21 preprocessing_pipe = Preprocessing()
22 hyperpipe += preprocessing_pipe
23 preprocessing_pipe += PipelineElement("LabelEncoder")
24
25 # Add transformer elements to the pipeline
26 hyperpipe += PipelineElement("SimpleImputer", hyperparameters={},
27                               test_disabled=False, missing_values=np.nan, strategy='mean', fill_value=0)
28 hyperpipe += PipelineElement("StandardScaler", hyperparameters={},
29                               test_disabled=False, with_mean=True, with_std=True)
30
31 # Add estimator to the pipeline
32 estimator_switch = Switch('EstimatorSwitch')
33 estimator_switch += PipelineElement("SVC",
34                                     hyperparameters={'C': [0.001, 0.1, 1, 3, 10], 'kernel': ['linear', 'rbf']},
35                                     gamma='scale', max_iter=1000000)
36 estimator_switch += PipelineElement("RandomForestClassifier",
37                                     hyperparameters={'n_estimators': [100],
38                                                       'max_depth': [2, 4, 10, None],
39                                                       'min_samples_leaf': [1, 10, 0.1]},
40                                     criterion='gini')
41 hyperpipe += estimator_switch
42
43 # Fit the pipeline specified above to the data
44 results = hyperpipe.fit(X, y)

```

Beispielskript für eine Regressionsanalyse:

```

1  # Define features and target variable individually for each analysis
2  X = np.asarray(sample.loc[:, features])
3  y = np.asarray(sample.loc[:, label])
4
5  # Define the individual pipeline (referred to as "hyperpipe")
6  hyperpipe = Hyperpipe(regression_example,
7                        optimizer="grid_search",
8                        metrics=['mean_squared_error', 'mean_absolute_error',
9                                'explained_variance', 'pearson_correlation', 'r2'],
10                       best_config_metric="mean_absolute_error
11                       outer_cv=StratifiedKfoldRegression(n_splits=10, shuffle=True),
12                       inner_cv=StratifiedKfoldRegression(n_splits=5, shuffle=True),
13                       verbosity=0,
14                       output_settings=OutputSettings(
15                           project_folder="...",
16                           mongodb_connect_url="...",
17                           overwrite_results=False,
18                           save_output=True))
19
20 # Add transformer elements to the pipeline
21 hyperpipe += PipelineElement("SimpleImputer", hyperparameters={},
22                              test_disabled=False, missing_values=np.nan, strategy='mean', fill_value=0)
23 hyperpipe += PipelineElement("StandardScaler", hyperparameters={},
24                              test_disabled=False, with_mean=True, with_std=True)
25
26 # Add estimator to the pipeline
27 estimator_switch = Switch('EstimatorSwitch')
28 estimator_switch += PipelineElement("SVR", hyperparameters={'C': [0.001, 0.1, 1, 3, 10],
29                                                            'kernel': ['linear', 'rbf']},
30                                   max_iter=10000000, gamma='scale', epsilon=0.1)
31 estimator_switch += PipelineElement("RandomForestRegressor",
32                                   hyperparameters={'min_samples_leaf': [1, 10, 0.1],
33                                                    'max_depth': [2, 4, 10, None],
34                                                    'n_estimators': [100]},
35                                   criterion='mse')
36 hyperpipe += estimator_switch
37
38 # Fit the pipeline specified above to the data
39 results = hyperpipe.fit(X, y)

```

8.11 Ergebnistabellen zur Modalität: neuropsychologische Testung

Wie in Kapitel 3 beschrieben, sind in dieser Arbeit der Übersichtlichkeit halber nicht alle Ergebnistabellen im Hauptteil dargestellt. Die Kriterien für die Entscheidung, wann eine Tabelle im Anhang bzw. im Hauptteil dargestellt ist, finden sich ebenfalls in Kapitel 3. Im Folgenden finden sich entsprechend die Ergebnistabellen der Analysen, in denen die Prädiktion der Zielvariablen nicht über alle Subgruppen hinweg weitestgehend konsistent besser als anhand des Dummys gelang. In diesem Kapitel werden nur die Ergebnisse der Analysen anhand der Modalität der neuropsychologischen Testung dargestellt. Die Verweise auf die entsprechenden Tabellen finden sich im Hauptteil dieser Arbeit.

Tabelle 8-7: Klassifikationsanalyse: rezidivierend vs. einzelne Episode, Modalität: neuropsych. Testung

	rez. ^a		nicht rez. ^a	Dummy Test		Test		Perf. ^b
	n	n	n	BAC M (SD)		BAC M (SD)		BAC _{test} - BAC _{dum.}
komplette Stichprobe	736	474	262	0.50	(0.00)	0.50	(0.04)	0.00
Geschlecht								
männlich	253	160	93	0.50	(0.00)	0.49	(0.08)	-0.01
weiblich	483	314	169	0.50	(0.00)	0.50	(0.07)	0.00
Alter								
> 55	77	55	22	0.50	(0.00)	0.52	(0.12)	0.02
≥ 30 und ≤ 55	352	236	116	0.50	(0.00)	0.51	(0.07)	0.01
< 30	307	183	124	0.50	(0.00)	0.52	(0.09)	0.02
höchster Bildungsgrad								
Lehre, Abi, Meister, Uni ^c	504	326	178	0.50	(0.00)	0.52	(0.03)	0.02
Schulabschluss ^d	230	146	84	0.50	(0.00)	0.47	(0.07)	-0.03
zeitl. Krankheitsverlauf								
rezidivierend	474	474	0	0.00	(0.00)	0.00	(0.00)	0.00
einzelne Episode	262	0	262	0.00	(0.00)	0.00	(0.00)	0.00
Hospitalisierungsgrad^e								
mind. einmal	492	347	145	0.50	(0.00)	0.51	(0.03)	0.01
noch nie	244	127	117	0.50	(0.00)	0.51	(0.08)	0.01
Hospitalisierungsstatus^e								
hospitalisiert	218	150	68	0.50	(0.00)	0.46	(0.08)	-0.04
nicht hospitalisiert	518	324	194	0.50	(0.00)	0.52	(0.05)	0.02
Remissionsstatus								
akut	312	216	96	0.50	(0.00)	0.51	(0.07)	0.01
partielle Remission	190	134	56	0.50	(0.00)	0.51	(0.09)	0.01
volle Remission	234	124	110	0.50	(0.00)	0.56	(0.04)	0.06
Antidepressiva								
mind. eins	275	207	68	0.50	(0.00)	0.49	(0.03)	-0.01
keins	186	100	86	0.50	(0.00)	0.53	(0.09)	0.03
Psych. Komorbidität								
mind. eine	208	151	57	0.50	(0.00)	0.49	(0.03)	-0.01
keine	327	199	128	0.50	(0.00)	0.59	(0.06)	0.09
Studienzentrum								
Marburg	323	217	106	0.50	(0.00)	0.50	(0.06)	0.00
Münster	413	257	156	0.50	(0.00)	0.51	(0.06)	0.01

Metriken: gemittelte Werte der 10 äußeren Faltungen, BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden, die eine rezidivierende MDD (engl.: major depressive disorder, MDD) oder eine einzelne Episode hatten, ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^e(teil-)stationäre Behandlungen

Tabelle 8-8: Klassifikationsanalyse: hospitalisiert vs. nicht hosp., Modalität: neuropsych. Testung

	hosp. ^a		nicht hosp. ^a	Dummy Test		Test		Perf. ^b
	n	n	n	BAC M (SD)		BAC M (SD)		BAC _{test} - BAC _{dum.}
komplette Stichprobe	736	218	518	0.50	(0.00)	0.52	(0.03)	0.02
Geschlecht								
männlich	253	96	157	0.50	(0.00)	0.47	(0.09)	-0.03
weiblich	483	122	361	0.50	(0.00)	0.49	(0.02)	-0.01
Alter								
> 55	77	16	61	0.50	(0.00)	0.55	(0.21)	0.05
≥ 30 und ≤ 55	352	106	246	0.50	(0.00)	0.47	(0.06)	-0.03
< 30	307	96	211	0.50	(0.00)	0.50	(0.06)	0.00
höchster Bildungsgrad								
Lehre, Abi, Meister, Uni ^c	504	134	370	0.50	(0.00)	0.50	(0.02)	0.00
Schulabschluss ^d	230	84	146	0.50	(0.00)	0.57	(0.12)	0.07
zeitl. Krankheitsverlauf								
rezidivierend	474	150	324	0.50	(0.00)	0.52	(0.05)	0.02
einzelne Episode	262	68	194	0.50	(0.00)	0.52	(0.05)	0.02
Hospitalisierungsgrad^e								
mind. einmal	492	218	274	0.50	(0.00)	0.55	(0.04)	0.05
noch nie	244	0	244	0.00	(0.00)	0.00	(0.00)	0.00
Hospitalisierungsstatus^e								
hospitalisiert	218	218	0	0.00	(0.00)	0.00	(0.00)	0.00
nicht hospitalisiert	518	0	518	0.00	(0.00)	0.00	(0.00)	0.00
Remissionsstatus								
akut	312	159	153	0.50	(0.00)	0.52	(0.07)	0.02
partielle Remission	190	54	136	0.50	(0.00)	0.54	(0.11)	0.04
volle Remission	234	5	229	0.00	(0.00)	0.00	(0.00)	0.00
Antidepressiva								
mind. eins	275	129	146	0.50	(0.00)	0.56	(0.07)	0.06
keins	186	27	159	0.50	(0.00)	0.50	(0.07)	0.00
Psych. Komorbidität								
mind. eine	208	64	144	0.50	(0.00)	0.51	(0.07)	0.01
keine	327	88	239	0.50	(0.00)	0.50	(0.04)	0.00
Studienzentrum								
Marburg	323	87	236	0.50	(0.00)	0.50	(0.07)	0.00
Münster	413	131	282	0.50	(0.00)	0.51	(0.03)	0.01

Metriken: gemittelte Werte der 10 äußeren Faltungen, BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden, die sich in (teil-)stationärer Behandlung befanden oder nicht, ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^e(teil-)stationäre Behandlungen

Tabelle 8-9: Regressionsanalyse: HAMD-Wert, Modalität: neuropsychologische Testung

	n	HAMD		Dummy Test		Test		Perf. ^a		Test	
		M (SD)		MAE M (SD)		MAE M (SD)		$\frac{MAE_{test}}{MAE_{dum.}}$	r	M (SD)	
komplette Stichprobe	736	9.14 (7.09)		5.88 (0.09)		5.82 (0.12)		0.99	0.08	(0.08)	
Geschlecht											
männlich	253	8.91 (6.83)		5.75 (0.27)		5.88 (0.54)		1.02	-0.03	(0.23)	
weiblich	483	9.26 (7.23)		5.96 (0.18)		5.92 (0.33)		0.99	0.12	(0.18)	
Alter											
> 55	77	8.21 (6.65)		5.51 (0.74)		5.64 (0.73)		1.02	0.03	(0.38)	
≥ 30 und ≤ 55	352	9.31 (7.32)		6.06 (0.17)		6.01 (0.22)		0.99	0.10	(0.21)	
< 30	307	9.18 (6.94)		5.76 (0.15)		5.79 (0.26)		1.01	0.12	(0.15)	
höchster Bildungsgrad											
Lehre, Abi, Meister, Uni ^b	504	8.54 (6.91)		5.67 (0.16)		5.64 (0.25)		0.99	0.13	(0.13)	
Schulabschluss ^c	230	10.43 (7.34)		6.16 (0.11)		6.15 (0.13)		1.00	-0.02	(0.18)	
zeitl. Krankheitsverlauf											
rezidivierend	474	9.89 (7.03)		5.89 (0.14)		5.99 (0.34)		1.02	0.07	(0.16)	
einzelne Episode	262	7.77 (7.00)		5.68 (0.28)		5.64 (0.32)		0.99	0.08	(0.22)	
Hospitalisierungsgrad^d											
mind. einmal	492	10.34 (7.26)		6.08 (0.11)		6.07 (0.18)		1.00	0.02	(0.12)	
noch nie	244	6.73 (6.07)		4.95 (0.23)		4.89 (0.27)		0.99	0.06	(0.14)	
Hospitalisierungsstatus^d											
hospitalisiert	218	12.68 (6.93)		5.73 (0.23)		5.84 (0.35)		1.02	0.08	(0.20)	
nicht hospitalisiert	518	7.65 (6.62)		5.43 (0.12)		5.33 (0.14)		0.98	0.12	(0.13)	
Remissionsstatus											
akut	312	14.05 (6.50)		5.35 (0.13)		5.43 (0.26)		1.01	0.02	(0.21)	
partielle Remission	190	7.86 (5.46)		4.47 (0.08)		4.45 (0.08)		0.99	-0.19	(0.19)	
volle Remission	234	3.64 (3.76)		3.03 (0.25)		2.87 (0.33)		0.95	0.11	(0.24)	
Antidepressiva											
mind. eins	275	10.87 (7.07)		5.91 (0.19)		5.97 (0.20)		1.01	0.04	(0.18)	
keins	186	6.49 (6.59)		5.24 (0.34)		5.00 (0.41)		0.95	0.01	(0.22)	
Psych. Komorbidität											
mind. eine	208	10.94 (6.95)		5.78 (0.28)		5.76 (0.38)		1.00	0.09	(0.33)	
keine	327	7.30 (6.24)		5.16 (0.18)		5.04 (0.20)		0.98	0.05	(0.19)	
Studienzentrum											
Marburg	323	8.28 (6.77)		5.67 (0.22)		5.58 (0.35)		0.98	0.15	(0.14)	
Münster	413	9.82 (7.27)		6.02 (0.19)		5.83 (0.37)		0.97	0.20	(0.19)	

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, HAMD = Hamilton Depressions-Skala, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlung

Tabelle 8-10: Regressionsanalyse: BDI-Wert, Modalität: neuropsychologische Testung

	n	BDI		Dummy Test		Test		Perf. ^a		Test	
		M (SD)		MAE M (SD)		MAE M (SD)		$\frac{MAE_{test}}{MAE_{dum.}}$		r M (SD)	
komplette Stichprobe	736	17.14 (10.91)		9.12 (0.14)		9.11 (0.28)		1.00		0.09 (0.14)	
Geschlecht											
männlich	253	16.68 (10.27)		8.70 (0.36)		8.85 (0.39)		1.02		-0.03 (0.22)	
weiblich	483	17.38 (11.23)		9.33 (0.23)		9.25 (0.27)		0.99		0.15 (0.09)	
Alter											
> 55	77	15.81 (10.71)		8.69 (0.55)		8.21 (1.99)		0.95		0.22 (0.41)	
≥ 30 und ≤ 55	352	17.11 (11.00)		9.27 (0.26)		9.29 (0.42)		1.00		0.06 (0.13)	
< 30	307	17.51 (10.86)		8.99 (0.27)		9.13 (0.34)		1.02		-0.02 (0.19)	
höchster Bildungsgrad											
Lehre, Abi, Meister, Uni ^b	504	16.38 (10.91)		9.13 (0.20)		9.03 (0.43)		0.99		0.13 (0.11)	
Schulabschluss ^c	230	18.72 (10.76)		8.89 (0.20)		8.88 (0.22)		1.00		0.06 (0.14)	
zeitl. Krankheitsverlauf											
rezidivierend	474	18.15 (10.67)		8.85 (0.21)		8.88 (0.44)		1.00		0.11 (0.14)	
einzelne Episode	262	15.31 (11.12)		9.27 (0.34)		9.06 (0.39)		0.98		0.13 (0.14)	
Hospitalisierungsgrad^d											
mind. einmal	492	18.94 (10.94)		9.07 (0.16)		9.14 (0.24)		1.01		0.09 (0.15)	
noch nie	244	13.51 (9.92)		8.14 (0.41)		8.14 (0.47)		1.00		-0.06 (0.22)	
Hospitalisierungsstatus^d											
hospitalisiert	218	22.27 (10.64)		8.76 (0.32)		8.73 (0.54)		1.00		0.14 (0.19)	
nicht hospitalisiert	518	14.98 (10.29)		8.57 (0.13)		8.51 (0.21)		0.99		0.11 (0.20)	
Remissionsstatus											
akut	312	23.93 (9.80)		7.87 (0.24)		8.00 (0.32)		1.02		0.00 (0.14)	
partielle Remission	190	17.04 (9.24)		7.68 (0.16)		7.79 (0.29)		1.01		0.00 (0.20)	
volle Remission	234	8.17 (5.95)		4.80 (0.27)		4.71 (0.41)		0.98		0.18 (0.20)	
Antidepressiva											
mind. eins	275	20.61 (11.02)		9.10 (0.27)		9.00 (0.51)		0.99		0.16 (0.12)	
keins	186	13.59 (9.90)		8.13 (0.46)		7.88 (0.51)		0.97		-0.13 (0.13)	
Psych. Komorbidität											
mind. eine	208	19.62 (10.47)		8.67 (0.43)		8.91 (0.66)		1.03		-0.08 (0.18)	
keine	327	15.01 (10.27)		8.55 (0.34)		8.55 (0.32)		1.00		0.09 (0.15)	
Studienzentrum											
Marburg	323	18.17 (10.75)		9.15 (0.21)		9.17 (0.41)		1.00		0.04 (0.27)	
Münster	413	16.33 (10.97)		9.02 (0.32)		8.94 (0.52)		0.99		0.15 (0.08)	

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, BDI = Beck-Depressions-Inventar, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlungen

Tabelle 8-11: Regressionsanalyse: GAF-Wert, Modalität: neuropsychologische Testung

	n	GAF		Dummy Test		Test		Perf. ^a		Test	
		M (SD)	M (SD)	MAE M (SD)	MAE M (SD)	$\frac{MAE_{test}}{MAE_{dum.}}$	r M (SD)				
komplette Stichprobe	530	65.33 (13.56)	11.10 (0.07)	10.90 (0.40)	0.98	0.20 (0.12)					
Geschlecht											
männlich	181	63.57 (13.25)	10.72 (0.31)	10.88 (0.94)	1.02	0.11 (0.24)					
weiblich	349	66.24 (13.64)	11.21 (0.25)	10.91 (0.44)	0.97	0.22 (0.11)					
Alter											
> 55	46	65.65 (12.17)	9.31 (2.38)	9.61 (2.57)	1.03	0.37 (0.51)					
≥ 30 und ≤ 55	260	64.51 (14.24)	11.86 (0.25)	11.88 (0.41)	1.00	0.17 (0.13)					
< 30	224	66.21 (13.00)	10.45 (0.54)	9.86 (0.77)	0.94	0.30 (0.12)					
höchster Bildungsgrad											
Lehre, Abi, Meister, Uni ^b	363	66.49 (13.29)	10.88 (0.31)	10.88 (0.79)	1.00	0.17 (0.16)					
Schulabschluss ^c	165	62.81 (13.89)	11.34 (0.87)	11.54 (1.16)	1.02	0.16 (0.23)					
zeitl. Krankheitsverlauf											
rezidivierend	345	63.23 (12.77)	10.30 (0.35)	10.33 (0.54)	1.00	0.12 (0.14)					
einzelne Episode	185	69.25 (14.12)	11.64 (0.46)	11.72 (0.98)	1.01	0.18 (0.23)					
Hospitalisierungsgrad^d											
mind. einmal	348	61.90 (12.72)	10.01 (0.27)	9.88 (0.39)	0.99	0.16 (0.13)					
noch nie	182	71.90 (12.68)	10.54 (0.40)	10.43 (0.41)	0.99	0.06 (0.25)					
Hospitalisierungsstatus^d											
hospitalisiert	151	56.29 (9.11)	7.30 (0.43)	6.93 (0.97)	0.95	0.34 (0.25)					
nicht hospitalisiert	379	68.93 (13.36)	10.96 (0.14)	10.93 (0.40)	1.00	0.18 (0.16)					
Remissionsstatus											
akut	220	56.61 (8.93)	7.06 (0.25)	6.70 (0.50)	0.95	0.33 (0.21)					
partielle Remission	138	63.41 (9.81)	7.68 (0.38)	7.84 (0.50)	1.02	0.00 (0.34)					
volle Remission	172	78.03 (11.23)	9.25 (0.49)	9.44 (0.38)	1.02	0.10 (0.24)					
Antidepressiva											
mind. eins	149	61.19 (12.00)	9.32 (0.66)	9.07 (1.04)	0.97	0.34 (0.24)					
keins	106	70.25 (15.43)	12.85 (0.79)	12.19 (2.14)	0.95	0.35 (0.28)					
Psych. Komorbidität											
mind. eine	207	61.77 (12.47)	9.94 (0.46)	9.66 (0.50)	0.97	0.15 (0.29)					
keine	323	67.61 (13.75)	11.22 (0.32)	10.81 (0.52)	0.96	0.29 (0.14)					
Studienzentrum											
Marburg	240	64.89 (13.83)	11.43 (0.12)	10.61 (0.92)	0.93	0.36 (0.20)					
Münster	290	65.70 (13.33)	10.82 (0.19)	10.79 (0.20)	1.00	0.07 (0.12)					

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, GAF = Global Assessment of Functioning, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlungen

8.12 Ergebnistabellen zur Modalität: strukturelle MRT-Daten

Wie in Kapitel 3 beschrieben, sind in dieser Arbeit der Übersichtlichkeit halber nicht alle Ergebnistabellen im Hauptteil dargestellt. Die Kriterien für die Entscheidung, wann eine Tabelle im Anhang bzw. im Hauptteil dargestellt ist, finden sich ebenfalls in Kapitel 3. Im Folgenden finden sich entsprechend die Ergebnistabellen der Analysen, in denen die Prädiktion der Zielvariablen nicht über alle Subgruppen hinweg weitestgehend konsistent besser als anhand des Dummys gelang. In diesem Kapitel werden nur die Ergebnisse der Analysen anhand der Modalität der strukturellen MRT-Daten dargestellt. Die Verweise auf die entsprechenden Tabellen finden sich im Hauptteil dieser Arbeit.

Tabelle 8-12: Klassifikationsanalyse: rezidivierend vs. einzelne Episode, Modalität: strukt. MRT-Daten

	rez. ^a		nicht rez. ^a	Dummy Test		Test	Perf. ^b
	n	n	n	BAC M (SD)		BAC M (SD)	BAC _{test} – BAC _{dum.}
Komplette Stichprobe	432	288	144	0.50	(0.00)	0.55 (0.07)	0.05
Geschlecht							
männlich	153	102	51	0.50	(0.00)	0.62 (0.09)	0.12
weiblich	279	186	93	0.50	(0.00)	0.54 (0.07)	0.04
Alter							
> 55	51	38	13	0.50	(0.00)	0.71 (0.22)	0.21
≥ 30 und ≤ 55	201	140	61	0.50	(0.00)	0.51 (0.09)	0.01
< 30	180	110	70	0.50	(0.00)	0.49 (0.11)	-0.01
höchster Bildungsgrad							
Lehre, Abi, Meister, Uni ^c	291	196	95	0.50	(0.00)	0.55 (0.07)	0.05
Schulabschluss ^d	140	91	49	0.50	(0.00)	0.57 (0.09)	0.07
zeitl. Krankheitsverlauf							
rezidivierend	288	288	0	0.00	(0.00)	0.00 (0.00)	0.00
einzelne Episode	144	0	144	0.00	(0.00)	0.00 (0.00)	0.00
Hospitalisierungsgrad^e							
mind. einmal	302	215	87	0.50	(0.00)	0.56 (0.07)	0.06
noch nie	130	73	57	0.50	(0.00)	0.50 (0.10)	0.00
Hospitalisierungsstatus^e							
hospitalisiert	149	101	48	0.50	(0.00)	0.54 (0.10)	0.04
nicht hospitalisiert	283	187	96	0.50	(0.00)	0.56 (0.06)	0.06
Remissionsstatus							
akut	198	140	58	0.50	(0.00)	0.47 (0.11)	-0.03
partielle Remission	109	80	29	0.50	(0.00)	0.48 (0.03)	-0.02
volle Remission	125	68	57	0.50	(0.00)	0.58 (0.12)	0.08
Antidepressiva							
mind. eins	261	196	65	0.50	(0.00)	0.50 (0.06)	0.00
keins	171	92	79	0.50	(0.00)	0.49 (0.07)	-0.01
Psych. Komorbidität							
mind. eine	85	67	18	0.50	(0.00)	0.48 (0.03)	-0.02
keine	158	103	55	0.50	(0.00)	0.54 (0.09)	0.04
Studienzentrum							
Marburg	271	187	84	0.50	(0.00)	0.50 (0.06)	0.00
Münster	161	101	60	0.50	(0.00)	0.57 (0.10)	0.07

Metriken: gemittelte Werte der 10 äußeren Faltungen, BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden, die eine rezidivierende MDD (engl.: major depressive disorder, MDD) oder eine einzelne Episode hatten, ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^e(teil-)stationäre Behandlungen

Tabelle 8-13: Klassifikationsanalyse: hospitalisiert vs. nicht hospitalisiert, Modalität: strukt. MRT-Daten

	hosp. ^a		nicht hosp. ^a	Dummy Test		Test	Perf. ^b
	n	n	n	BAC M (SD)		BAC M (SD)	BAC _{test} – BAC _{dum.}
Komplette Stichprobe	432	149	283	0.50 (0.00)	0.54 (0.04)	0.04	
Geschlecht							
männlich	153	73	80	0.50 (0.00)	0.58 (0.13)	0.08	
weiblich	279	76	203	0.50 (0.00)	0.53 (0.11)	0.03	
Alter							
> 55	51	11	40	0.50 (0.00)	0.59 (0.21)	0.09	
≥ 30 und ≤ 55	201	68	133	0.50 (0.00)	0.52 (0.09)	0.02	
< 30	180	70	110	0.50 (0.00)	0.53 (0.06)	0.03	
höchster Bildungsgrad							
Lehre, Abi, Meister, Uni ^c	291	90	201	0.50 (0.00)	0.49 (0.08)	-0.01	
Schulabschluss ^d	140	59	81	0.50 (0.00)	0.55 (0.11)	0.05	
zeitl. Krankheitsverlauf							
rezidivierend	288	101	187	0.50 (0.00)	0.53 (0.05)	0.03	
einzelne Episode	144	48	96	0.50 (0.00)	0.52 (0.10)	0.02	
Hospitalisierungsgrad^e							
mind. einmal	302	149	153	0.50 (0.00)	0.58 (0.08)	0.08	
noch nie	130	0	130	0.00 (0.00)	0.00 (0.00)	0.00	
Hospitalisierungsstatus^e							
hospitalisiert	149	149	0	0.00 (0.00)	0.00 (0.00)	0.00	
nicht hospitalisiert	283	0	283	0.00 (0.00)	0.00 (0.00)	0.00	
Remissionsstatus							
akut	198	116	82	0.50 (0.00)	0.60 (0.11)	0.10	
partielle Remission	109	30	79	0.50 (0.00)	0.49 (0.09)	-0.01	
volle Remission	125	3	122	0.00 (0.00)	0.00 (0.00)	0.00	
Antidepressiva							
mind. eins	261	123	138	0.50 (0.00)	0.56 (0.08)	0.06	
keins	171	26	145	0.50 (0.00)	0.48 (0.04)	-0.02	
Psych. Komorbidität							
mind. eine	85	35	50	0.50 (0.00)	0.54 (0.10)	0.04	
keine	158	52	106	0.50 (0.00)	0.56 (0.12)	0.06	
Studienzentrum							
Marburg	271	77	194	0.50 (0.00)	0.47 (0.06)	-0.03	
Münster	161	72	89	0.50 (0.00)	0.55 (0.10)	0.05	

Metriken: gemittelte Werte der 10 äußeren Faltungen, BAC: korrigierte Klassifikationsgenauigkeit (engl.: balanced accuracy, BAC), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), ^aAnzahl der Probanden, die sich in (teil-)stationärer Behandlung befanden oder nicht, ^bPerformance: Differenz zwischen BAC (Test) und BAC (Dummy Test), ^cBachelor bzw. Masterabschluss, ^dVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^e(teil-)stationäre Behandlungen

Tabelle 8-14: Regressionsanalyse: HAMD-Wert, Modalität: strukturelle MRT-Daten

		HAMD	Dummy Test	Test	Perf.^a	Test
	n	M (SD)	MAE M (SD)	MAE M (SD)	$\frac{MAE_{test}}{MAE_{dum.}}$	R M (SD)
komplette Stichprobe	432	9.10 (7.27)	6.07 (0.13)	6.03 (0.13)	0.99	-0.05 (0.22)
Geschlecht						
männlich	153	9.51 (7.26)	6.13 (0.43)	6.18 (0.49)	1.01	-0.22 (0.15)
weiblich	279	8.88 (7.29)	6.02 (0.11)	6.02 (0.20)	1.00	-0.03 (0.15)
Alter						
> 55	51	8.67 (7.13)	5.92 (0.99)	6.09 (1.12)	1.03	-0.09 (0.46)
≥ 30 und ≤ 55	201	9.33 (7.47)	6.28 (0.35)	6.17 (0.37)	0.98	-0.13 (0.20)
< 30	180	8.97 (7.12)	5.88 (0.15)	5.90 (0.23)	1.00	0.02 (0.15)
höchster Bildungsgrad						
Lehre, Abi, Meister, Uni ^b	291	8.41 (7.26)	5.99 (0.19)	5.91 (0.19)	0.99	0.02 (0.18)
Schulabschluss ^c	140	10.54 (7.15)	6.07 (0.19)	6.19 (0.25)	1.02	-0.05 (0.27)
zeitl. Krankheitsverlauf						
rezidivierend	288	9.78 (7.13)	6.05 (0.13)	6.09 (0.19)	1.01	-0.09 (0.15)
einzelne Episode	144	7.74 (7.38)	5.91 (0.40)	5.81 (0.48)	0.98	-0.09 (0.28)
Hospitalisierungsgrad^d						
mind. einmal	302	10.67 (7.37)	6.18 (0.24)	6.19 (0.28)	1.00	0.04 (0.17)
noch nie	130	5.46 (5.57)	4.51 (0.30)	4.20 (0.40)	0.93	0.34 (0.25)
Hospitalisierungsstatus^d						
hospitalisiert	149	13.21 (6.79)	5.66 (0.29)	5.56 (0.55)	0.98	0.26 (0.24)
nicht hospitalisiert	283	6.94 (6.56)	5.35 (0.26)	5.18 (0.35)	0.97	0.04 (0.17)
Remissionsstatus						
akut	198	13.84 (6.59)	5.41 (0.23)	5.37 (0.36)	0.99	0.13 (0.21)
partielle Remission	109	7.44 (5.69)	4.68 (0.36)	4.53 (0.38)	0.97	0.18 (0.22)
volle Remission	125	3.04 (3.37)	2.70 (0.25)	2.64 (0.43)	0.98	0.14 (0.27)
Antidepressiva						
mind. eins	261	10.87 (7.14)	5.96 (0.24)	5.93 (0.47)	1.00	0.12 (0.22)
keins	171	6.40 (6.64)	5.24 (0.28)	5.01 (0.33)	0.96	0.07 (0.17)
Psych. Komorbidität						
mind. eine	85	10.74 (6.99)	5.89 (0.51)	6.18 (0.73)	1.05	-0.07 (0.26)
keine	158	6.74 (5.94)	4.89 (0.29)	4.74 (0.34)	0.97	-0.08 (0.20)
Studienzentrum						
Marburg	271	8.22 (6.71)	5.58 (0.21)	5.53 (0.23)	0.99	-0.12 (0.19)
Münster	161	10.58 (7.95)	6.74 (0.36)	6.77 (0.47)	1.01	-0.11 (0.20)

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, HAMD = Hamilton Depressions-Skala, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlung

Tabelle 8-15: Regressionsanalyse: BDI-Wert, Modalität: strukturelle MRT-Daten

	n	BDI		Dummy Test		Test		Perf. ^a		Test	
		M (SD)		MAE M (SD)		MAE M (SD)		$\frac{MAE_{test}}{MAE_{dum.}}$		R M (SD)	
komplette Stichprobe	432	17.72 (11.09)		9.34 (0.20)		9.53 (0.41)		1.02		0.01 (0.13)	
Geschlecht											
männlich	153	18.02 (10.43)		8.74 (0.59)		8.82 (0.70)		1.01		0.07 (0.24)	
weiblich	279	17.55 (11.45)		9.66 (0.21)		9.69 (0.28)		1.00		-0.11 (0.18)	
Alter											
> 55	51	15.78 (11.18)		9.02 (1.50)		9.14 (2.32)		1.01		-0.04 (0.40)	
≥ 30 und ≤ 55	201	18.17 (11.06)		9.37 (0.30)		9.38 (0.29)		1.00		-0.19 (0.27)	
< 30	180	17.76 (11.11)		9.20 (0.27)		9.29 (0.32)		1.01		-0.17 (0.25)	
höchster Bildungsgrad											
Lehre, Abi, Meister, Uni ^b	291	16.93 (11.47)		9.71 (0.31)		9.60 (0.32)		0.99		-0.05 (0.13)	
Schulabschluss ^c	140	19.31 (10.15)		8.37 (0.26)		8.41 (0.22)		1.01		-0.06 (0.21)	
zeitl. Krankheitsverlauf											
rezidivierend	288	18.66 (10.89)		9.14 (0.25)		9.15 (0.25)		1.00		-0.03 (0.11)	
einzelne Episode	144	15.82 (11.29)		9.37 (0.75)		9.28 (1.02)		0.99		-0.09 (0.21)	
Hospitalisierungsgrad^d											
mind. einmal	302	20.13 (10.88)		9.08 (0.27)		9.05 (0.33)		1.00		0.09 (0.10)	
noch nie	130	12.10 (9.46)		7.63 (0.42)		7.42 (0.47)		0.97		0.15 (0.23)	
Hospitalisierungsstatus^d											
hospitalisiert	149	22.97 (10.34)		8.51 (0.31)		8.69 (0.68)		1.02		0.06 (0.25)	
nicht hospitalisiert	283	14.95 (10.47)		8.77 (0.30)		8.62 (0.37)		0.98		0.06 (0.17)	
Remissionsstatus											
akut	198	24.01 (9.92)		7.95 (0.37)		7.98 (0.64)		1.00		0.13 (0.23)	
partielle Remission	109	17.34 (9.53)		7.91 (0.45)		7.91 (0.58)		1.00		-0.20 (0.21)	
volle Remission	125	8.08 (5.93)		4.75 (0.51)		4.83 (0.49)		1.02		0.12 (0.23)	
Antidepressiva											
mind. eins	261	20.60 (10.96)		9.03 (0.31)		9.01 (0.57)		1.00		0.05 (0.22)	
keins	171	13.31 (9.79)		7.97 (0.38)		7.72 (0.47)		0.97		-0.08 (0.22)	
Psych. Komorbidität											
mind. eine	85	21.45 (10.26)		8.67 (0.69)		8.71 (0.62)		1.00		-0.04 (0.30)	
keine	158	15.27 (10.28)		8.49 (0.50)		8.31 (0.57)		0.98		-0.11 (0.27)	
Studienzentrum											
Marburg	271	18.14 (10.72)		9.10 (0.23)		9.14 (0.23)		1.00		-0.15 (0.12)	
Münster	161	16.99 (11.69)		9.73 (0.59)		9.86 (0.75)		1.01		-0.01 (0.23)	

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, BDI = Beck-Depressions-Inventar, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlung

Tabelle 8-16: Regressionsanalyse: GAF-Wert, Modalität: strukturelle MRT-Daten

	GAF		Dummy Test		Test		Perf. ^a		Test	
	n	M (SD)	MAE M (SD)	MAE M (SD)	MAE M (SD)	$\frac{MAE_{test}}{MAE_{dum.}}$	R	M (SD)		
komplette Stichprobe	238	65.31 (14.23)	11.69 (0.43)	11.61 (0.67)	0.99	0.00 (0.17)				
Geschlecht										
männlich	90	62.07 (13.73)	10.92 (0.75)	10.93 (1.10)	1.00	0.06 (0.30)				
weiblich	148	67.28 (14.22)	11.63 (0.64)	11.89 (0.96)	1.02	0.06 (0.29)				
Alter										
> 55	24	65.79 (11.96)	9.04 (3.66)	8.87 (3.41)	0.98	0.41 (0.76)				
≥ 30 und ≤ 55	112	63.23 (14.32)	11.59 (0.89)	11.53 (1.00)	1.00	0.02 (0.29)				
< 30	102	67.47 (14.41)	11.93 (0.82)	12.06 (1.21)	1.01	0.06 (0.31)				
höchster Bildungsgrad										
Lehre, Abi, Meister, Uni ^b	156	68.54 (13.90)	11.41 (0.65)	11.63 (0.88)	1.02	-0.02 (0.23)				
Schulabschluss ^c	81	59.09 (12.89)	10.40 (0.68)	10.45 (0.77)	1.00	0.01 (0.31)				
zeitl. Krankheitsverlauf										
rezidivierend	165	63.47 (13.08)	10.49 (0.75)	10.35 (0.83)	0.99	-0.14 (0.18)				
einzelne Episode	73	69.47 (15.86)	13.31 (1.00)	13.80 (1.70)	1.04	0.05 (0.43)				
Hospitalisierungsgrad^d										
mind. einmal	168	61.43 (13.07)	10.14 (0.71)	10.04 (0.75)	0.99	0.07 (0.30)				
noch nie	70	74.60 (12.58)	10.27 (0.77)	8.63 (1.57)	0.84	0.53 (0.26)				
Hospitalisierungsstatus^d										
hospitalisiert	86	56.12 (9.91)	7.97 (1.04)	7.89 (1.39)	0.99	0.03 (0.31)				
nicht hospitalisiert	152	70.51 (13.68)	11.31 (0.45)	11.47 (0.97)	1.01	0.10 (0.19)				
Remissionsstatus										
akut	109	56.28 (9.43)	7.50 (0.67)	7.68 (0.83)	1.02	-0.08 (0.35)				
partielle Remission	64	64.61 (10.48)	8.48 (1.28)	8.59 (1.16)	1.01	-0.34 (0.29)				
volle Remission	65	81.12 (9.94)	8.14 (0.94)	8.16 (1.11)	1.00	0.29 (0.32)				
Antidepressiva										
mind. eins	142	61.46 (12.10)	9.35 (0.84)	9.53 (1.28)	1.02	-0.05 (0.26)				
keins	96	71.00 (15.27)	12.61 (0.93)	12.88 (1.19)	1.02	0.06 (0.27)				
Psych. Komorbidität										
mind. eine	84	61.25 (13.78)	11.05 (1.01)	11.36 (1.08)	1.03	0.10 (0.27)				
keine	154	67.52 (14.03)	11.30 (0.53)	11.42 (0.56)	1.01	-0.08 (0.14)				
Studienzentrum										
Marburg	191	64.54 (13.58)	11.16 (0.36)	11.13 (0.42)	1.00	0.09 (0.22)				
Münster	47	68.40 (16.43)	13.73 (2.62)	15.33 (3.79)	1.12	-0.25 (0.47)				

Metriken: gemittelte Werte der 10 äußeren Faltungen, MAE: Mittlerer absoluter Fehler (engl.: mean absolute error, MAE), M: Mittelwert (engl.: mean, M), SD: Standardabweichung (engl.: standard deviation, SD), r: Pearson-Korrelation, Dummy-Strategie: Prädiktion des Mittelwertes der Zielvariable, GAF = Global Assessment of Functioning, ^aPerformance: MAE (Test) geteilt durch MAE (Dummy Test), ^bBachelor bzw. Masterabschluss, ^cVolks-, Haupt-, Realschulabschluss bzw. Fachhochschulreife, ^d(teil-)stationäre Behandlung

8.13 Votum der Ethik-Kommission am Studienstandort Marburg

Im Folgenden ist das Ethik-Votum für die FOR 2107-Studie am Standort Marburg dargestellt.



Philipps-Universität - 35032 Marburg

Herrn
Prof. Dr. Tilo Kircher
Universitätsklinikum Gießen und Marburg
GmbH
Standort Marburg
Klinik für Psychiatrie und Psychotherapie
Rudolf-Bultmann-Str. 8
D-35033 Marburg

Fachbereich Medizin
Dekanat/Ethikkommission

Prof. Dr. med. Gerd Richter (Vors.)

Tel.: 06421 586 6487
Fax: 06421 586 6585
Sek.: D. Raiss/M. Backes
E-Mail: ethikkom@staff.uni-marburg.de
Anschrift: Baldingerstrasse/Postfach 2360
35032 Marburg
Web: www.med.uni-marburg.de/ethikkomm
Az.: Studie 07/2014

Marburg, den 07.02.2014

Studie: DFG-Forscherguppe 2107 „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und –funktion“.

- WP1 – Analyse der Neurobiologie im Verlauf affektiver Störungen – die Marburg Affective Disorder Cohort Study (MACS), Prof. Udo Dannlowski, Prof. Carsten Konrad, Prof. Tilo Kircher
- WP3 – microRNAs als Regulatoren der Neuroplastizität bei affektiven Störungen, Prof. Gerhard Schratt
- WP4 – Zellulärer Stress und mitochondriale Funktionsstörungen bei affektiven Erkrankungen, Prof. Carsten Culmsee
- WP5 – Integrierte Analyse genetischer, epigenetischer und umweltbedingter Vulnerabilitätsfaktoren in affektiven Störungen, Prof. Marcella Rietschel, PD Axel Krug, Dr. Stephanie Witt, Prof. Markus Nöthen
- WP6 – Methodische Aspekte longitudinaler MRT Studien: Reliabilität, Qualitätssicherung statistische Genetik und Genetische Epidemiologie, Prof. Andreas Jansen, PD Astrid Dempfle
- CP1 – Zentralprojekt Biomaterialbank (BMB), PD Ina Pfefferle, Prof. Harald Renz, Prof. Dr. Thomas Schulze

Eingereichte Unterlagen:

1. Ihr Anschreiben vom 07.01.2014,
2. Ihre Stellungnahme vom 04.02.2014 zum Ethikkommissionsgutachten vom 24.01.2014,
3. Ethikantrag FOR 2107 – Version 2 vom 04.02.2014,
4. Studienprotokoll und Ablaufplan,
5. Detailinformationen zur Biomaterialienbank MACS/CP1 als Teil der Biobank Göttingen KFO241
6. Datenschutzrechtliche Stellungnahme zum Datenschutzkonzept für die Klinische Forschergruppe 241 Universitätsmedizin Göttingen vom 04.02.2014,
7. Probandeninformation und Einwilligung - Neuropsychologie und MRT - Version 2, 04.02.14,
8. Patienteninformation und Einwilligung – Neuropsychologie und MRT – Version 2, 04.02.14,
9. Probandeninformation und Einwilligung Reliabilitätsuntersuchung – Version 2, 04.02.14,
10. Forschergruppenantrag bei der Deutschen Forschungsgemeinschaft zum Thema „Neurobiology of affective disorders: A translational perspective on brain structure and function“,
11. Ethikvoten der Ethikkommission der Philipps-Universität Marburg zu entsprechenden Vorstudien,
12. Votum der Ethikkommission der Med. Fakultät Universitätsmedizin Göttingen vom 31.07.2013 zur Forschergruppe KFO241 „Genome-wide dissection of genotype-phenotype relationships in the longitudinal course of psychosis“,
13. Ethikvoten der Universitäten München, Bochum, Tübingen, Ulm und Graz zur Biomaterialsammlung im Rahmen der klinischen Forschergruppe 241 der Universitätsmedizin Göttingen,
14. CRF der Biomaterialbank Universitätsmedizin Göttingen (elektronische Eingabemaske),
15. Arbeitsanweisung Blutabnahme in Göttingen SOPBL-02, Version 02 v. 30.05.2012,
16. Datenschutzkonzept für die klinische Forschergruppe 241, Universitätsmedizin Göttingen vom 18.07.2012,
17. Benutzerhandbuch STARLIMS Klinische Forschergruppe 241 vom 15.05.2012,

**Sekretariat : Frau Raiss Montag – Donnerstag 8.00 – 12.00 Uhr, Freitag 8.00 – 11.00 Uhr
Frau Backes Montag – Donnerstag 14.00 – 16.00 Uhr**

Kommissionsmitglieder: Prof. Dr. med. R. Berger, Prof. Dr. jur. G. Freund, Prof. Dr. Dr. H. Brettel, Prof. Dr. med. J.-C. Krieg, Prof. Dr. med. Czubayko, Prof. Dr. med. G. Richter (Vorsitzender), PD Dr. C. Seifart, Prof. Dr. med. Uwe Wagner (stellvertretender Vorsitzender), Prof. Dr. med. R. Maier, Prof. Dr. med. N. Donner-Banzhoff, Dipl. Math. B. Greene Dr. rer. nat. Nina Timmesfeld, Prof. Dr. med. A. Neubauer, PD Dr. B. Tackenberg, Bettina Nieth, Dr. Thomas Neubert, cand. med. Christian Kreisel, cand. med. Martin Fries

- 2 -

18. Patienteninformation und Einwilligung – Biobank – Version 3, 07.02.2014,
19. Probandeninformation und Einwilligung – Biobank – Version 3, 07.02.2014,
20. Rekrutierungsflyer zur Studie.

Sehr geehrter Herr Professor Kircher,

vielen Dank für Ihre ausführliche Stellungnahme sowie die Nachreichung fehlender Unterlagen und entsprechenden Modifikationen und Korrekturen in den eingesandten Unterlagen entsprechend unserem Gutachten vom 24.01.2014. Damit sind alle Auflagen erfüllt. Es bestehen nunmehr keine berufsethischen oder berufsrechtlichen Bedenken gegen die Durchführung der Studie.

Damit schließt die Ethikkommission des Fachbereichs Medizin der Philipps-Universität Marburg ihr Gutachten mit einem

positiven Ethikkommissionsvotum

ab.

Entsprechend der ausschließlich beratenden Funktion der Ethik-Kommission betrifft dieses Votum nur die ethische Beurteilung der Konzeption, der vorgesehenen Methoden, der Durchführung und Überwachung des betreffenden Projektes sowie der beabsichtigten Patientenaufklärung. Die ärztliche und juristische Verantwortung verbleibt jedoch uneingeschränkt beim Projektleiter und seinen Mitarbeitern.

Bitte geben Sie uns jede Änderung in der Protokolldurchführung bekannt. Es muss dann geklärt werden, ob das Votum der Ethik-Kommission noch Gültigkeit hat.

Über alle schwerwiegenden oder unerwarteten unerwünschten Ereignisse, die während der Studie auftreten und die Sicherheit der Studienteilnehmer oder die Durchführung der Studie beeinträchtigen könnten, muss der Vorsitzende der Ethik-Kommission unterrichtet werden.

Die Ethik-Kommission des Fachbereichs Humanmedizin der Philipps-Universität Marburg arbeitet gemäß den nationalen gesetzlichen Bestimmungen und den ICH-GCP-Richtlinien.

Außerdem benötigt die Ethik-Kommission einen Bericht nach Abschluss der Studie.

Mit freundlichen kollegialen Grüßen
für die Ethik-Kommission des
Fachbereichs Humanmedizin
der Philipps-Universität Marburg

Prof. Dr. med. G. Richter
(Vorsitzender Ethikkommission)

8.14 Votum der Ethik-Kommission am Studienstandort Münster

Im Folgenden ist das Ethik-Votum für die FOR 2107-Studie am Standort Münster dargestellt.

 <p>ÄRZTEKAMMER WESTFALEN-LIPPE</p>	 <p>medizinische fakultät Westfälische Wilhelms-Universität Münster</p>	<p>ETHIK KOMMISSION der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität</p>
<p>Ethik-Kommission Münster · Gartenstraße 210 – 214 · 48147 Münster</p>		<p>Gartenstraße 210 – 214 48147 Münster, Germany Tel.: +49 (0)251 929 2460 Fax: +49 (0)251 929 2478 E-Mail: ethik-kommission@aekwl.de www.ethik-kommission.uni-muenster.de</p>
<p>Herrn Prof. Dr. med. Dr. phil Udo Dannlowski Klinik für Psychiatrie und Psychotherapie Philipps-Universität Marburg Rudolf-Bultmann-Straße 8 35039 Marburg</p>		<p>1. August 2014</p>
<p>Unser Aktenzeichen: 2014-422-b-S (bitte immer angeben!) Prüfplancode: DFG Gesch.z. DA 1151/5-1 Sponsor / Finanzierung: DFG Rahmenprogramm FOR 2107 Titel des Forschungsvorhabens: „Neurobiologie affektiver Störungen: Eine translationale Perspektive auf Hirnstruktur und -funktion“</p>		
<p>Votum</p>		
<p>Sehr geehrter Herr Prof. Dannlowski,</p>		
<p>für das oben genannte Forschungsvorhaben, welches bereits von einer anderen Ethik-Kommission positiv bewertet wurde, haben Sie mit Schreiben vom 26.07.2014 die Beratung durch die Ethik-Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster („Ethik-Kommission“) beantragt.</p>		
<p>Probanden für WP1 sollen auch in Münster am Universitätsklinikum rekrutiert werden. Das Biomaterial wird in die Biobank in Göttingen überführt.</p>		
<p>Die Ethik-Kommission hat die in Anlage 1 näher bezeichneten Unterlagen geprüft.</p>		
<p>Danach schließt sich unsere Ethik-Kommission dem Erstvotum der Ethik-Kommission des Fachbereichs Humanmedizin der Philipps-Universität Marburg (AZ:07/2014) an.</p>		
<p>Somit bestehen seitens der Ethik-Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster grundsätzlich weder ethische noch juristische Bedenken gegen Ihre Beteiligung / die Beteiligung des Universitätsklinikums Münster an der oben genannten Studie.</p>		
<p>Sie sind damit Ihrer Berufspflicht nach § 15 Abs. 1 der Berufsordnung der Ärztekammer Westfalen-Lippe nachgekommen.</p>		
<p>Wesentliche Änderungen des Studiendesigns (Amendments) und Meldungen schwerwiegender unerwünschter Ereignisse sollten uns – möglichst koordiniert durch den Projektleiter / die Studienzentrale / den Sponsor - mit entsprechender Stellungnahme der erstvotierenden Ethik-Kommission zugeleitet werden.</p>		
<p><small>Mitglieder: H.-W. Rojke (Vorsitzender), H. Pfeiffer (stellv. Vorsitzender), F. U. Müller, P. Scheuvel, R. Rapp-Engels, M. Focking, B. Huchlerbroch, J. Ritter, H. Schulze-Möcking, H.-D. Steinmeyer, D. Voß, M. Quante</small></p>		

Ethik-Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster
unser Az.: 2014-422-b-S
Schreiben vom: 1. August 2014

Für die Entscheidung der Ethik-Kommission erhebt die Ärztekammer Westfalen-Lippe Gebühren nach Maßgabe ihrer Verwaltungsgebührenordnung. Über die auf 50 % der Regelgebühr ermäßigten Gebühren erhalten Sie von der Ärztekammer einen gesonderten Bescheid.

Allgemeine Hinweise

Mit der vorliegenden Stellungnahme berät Sie die Ethik-Kommission zu den mit Ihrem Forschungsvorhaben verbundenen berufsethischen und berufsrechtlichen Fragen gemäß § 15 Abs. 1 Satz 1 Berufsordnung der Ärztekammer Westfalen-Lippe.

Die Einschätzung der Ethik-Kommission ist als ergebnisoffene Beratung für den Antragsteller nicht bindend. Die Ethik-Kommission weist darauf hin, dass unabhängig von der vorliegenden Stellungnahme die medizinische, ethische und rechtliche Verantwortung für die Durchführung des Forschungsvorhabens bei dessen Leiter und bei allen an dem Vorhaben teilnehmenden Ärzten bzw. Forschern verbleibt.

An der Beratung und Beschlussfassung haben keine Mitglieder der Ethik-Kommission teilgenommen, die selbst an dem Forschungsvorhaben mitwirken oder deren Interessen davon berührt werden.

Die Ethik-Kommission der Ärztekammer Westfalen-Lippe und der Medizinischen Fakultät der Westfälischen Wilhelms-Universität Münster ist organisiert und arbeitet gemäß den nationalen gesetzlichen Bestimmungen und den GCP-Richtlinien der ICH.

Mit freundlichen Grüßen


Univ.-Prof. Dr. med. Heidi Pfeiffer
Stellv. Vorsitzende der Ethik-Kommission