

Aus dem LVR-Klinikum Düsseldorf  
Klinik und Poliklinik für Psychiatrie und Psychotherapie  
der Heinrich-Heine-Universität Düsseldorf  
Direktorin: Univ.-Prof. Dr. med. Eva Meisenzahl-Lechner

Prädiktion von Therapieerfolg bei Patienten mit  
Borderline Persönlichkeitsstörung mittels *Machine-*  
*Learning* basierter Ansätze - eine *real-world data*  
Analyse

Dissertation

zur Erlangung des Grades eines Doktors der Medizin  
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Furkan Demirbuga  
2025

Als Inauguraldissertation gedruckt mit der Genehmigung der  
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan/in: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter/in: PD. Dr. rer. nat. Susanne Neufang

Zweitgutachter/in: Prof. Dr. phil. Bettina Pollok

## Zusammenfassung (Deutsch)

Die Borderline Persönlichkeitsstörung stellt eine Untergruppe der emotional instabilen Persönlichkeitsstörungen dar und ist durch Symptome wie Identitätsstörung, suizidales Verhalten, instabile Beziehungen und Gefühle der Leere gekennzeichnet. In Anbetracht der hohen Prävalenz sowohl im ambulanten als auch im stationären Bereich, verursacht sie neben hohen funktionellen Beeinträchtigungen für die betroffenen Patienten eine erhebliche finanzielle Belastung des Gesundheitssystems.

Die vorliegende Arbeit stellt einen Versuch dar, das Therapieansprechen nach einer pharmakologischen und psychotherapeutischen Behandlung anhand von klinischen Routinedaten vorherzusagen. Zu diesem Zweck wurde ein Ansatz des maschinellen Lernens (ML) zur Klassifizierung von Patienten verwendet, die von der Therapie profitiert haben ([Teil]Responder), und von Patienten, die nicht davon profitiert haben (Non-Responder). Insgesamt wurden strukturelle MR-Bilder und klinisch-demografische Daten von 157 Patienten (126 [Teil]Responder und 31 Non-Responder) analysiert, die zwischen 2016 und 2020 im LVR-Klinikum Düsseldorf behandelt wurden. Insgesamt wurden fünf uni- und bimodale Klassifikatoren mit Hilfe von Support Vector Machines (SVM) ermittelt: das unimodale strukturelle MRT-Modell, das unimodale klinisch-demografische Modell, zwei bimodale MRT-klinisch-demografische Klassifikatoren sowie eine komprimierte Version der klinisch-demografischen Daten, die nur signifikante *features* verwendet. Die Bewertung der Klassifizierungen erfolgte insbesondere anhand der Kriterien Genauigkeit und Fläche unter der Kurve, und die Modellsignifikanz wurde anhand von Cohen's Kappa bestimmt.

Die Ergebnisse zeigten, dass nur das klinisch-demografische Originalmodell eine statistische Modellsignifikanz, eine Genauigkeit von 83,4 % und eine Fläche unter der Kurve von 0,73 aufwies. Die Analyse der *feature weights* ergab, dass insbesondere Medikamente und komorbide Störungen einen erheblichen Einfluss auf die Vorhersage haben. Die Plausibilität der Ergebnisse wurde durch die Übereinstimmung mit der vorhandenen wissenschaftlichen Literatur bestätigt. Mit vielversprechenden Werten in den relevanten Gütekriterien und plausiblen *feature weights* unterstreicht diese Dissertation die entscheidende Rolle von Routinedaten in der klinischen Forschung.

## **Zusammenfassung (Englisch)**

Borderline personality disorder is a subgroup of emotionally unstable personality disorders and is characterized by symptoms such as identity disorder, suicidal behavior, unstable relationships, and feelings of emptiness. Given its high prevalence in both the outpatient and inpatient sectors, it not only causes significant functional impairment for the patients affected, but also a considerable financial burden on the healthcare system. Initial studies have already been conducted to determine predictive characteristics for the response to therapy.

This present work constitutes an attempt to predict therapy response after pharmacological and psychotherapeutic treatment using a real-world data set. To do so, a Machine-Learning (ML) approach was used for classification into patients, who benefited from the therapy ([partial]responders) and those patients, who did not (non-responders).

In total, structural MR images and clinical-demographic data of 157 patients (126 (partial)responders and 31 non-responders) treated at LVR-Klinikum Düsseldorf between 2016 and 2020 was analyzed. In total, five uni- and bimodal classifiers were determined using Support Vector Machines (SVM), i.e., the unimodal structural MRI model, the unimodal clinical-demographic model, two bimodal MRI-clinical-demographic classifiers as well as a condensed version of the clinical-demographic data using significant features only. The evaluation of the classifications was particularly conducted through the metrics accuracy and area under the curve and model significance was determined using Cohen's Kappa.

Results showed that only the clinical-demographic original model reached a statistical model significance, an accuracy of 83.4 % and an area under the curve of 0.73. Analysis of the feature weights revealed that medication and comorbid disorders, in particular, had a significant impact on the prediction. Plausibility of the results was confirmed by consistency of most predictive features with the existing scientific literature. With promising performance measures and plausible feature weights this dissertation underpins the crucial role of real-world data sets in clinical research.

## Abkürzungsverzeichnis

AUC .....	<i>Area under the curve</i>
BDI .....	Beck-Depressions-Inventar
BPS .....	Borderline Persönlichkeitsstörung
BSL .....	Borderline Symptom Liste
Cat12 .....	<i>Computational Anatomy Toolbox</i>
CGI .....	<i>Clinical Global Impression</i>
CV .....	<i>Cross-validation folds</i>
DBT .....	Dialektisch Behaviorale Therapie
DSM-IV .....	<i>Diagnostic and Statistical Manual of Mental Disorders, fourth edition</i>
FA .....	<i>Flip angle</i>
FN .....	<i>False negative</i>
FP .....	<i>False positive</i>
GAF .....	<i>Global Assesment of Functioning</i>
GM .....	<i>Gray matter</i>
ICBM .....	<i>International Consortium for Brain Mapping</i>
ICD-10....	<i>International Statistical Classification of Diseases and Related Health Problems- 10</i>
KIS .....	Klinikinformationssystem
LVR .....	Lanschaftsverband Rheinland
MAR .....	<i>Missing at random</i>
MCAR .....	<i>Missing completely at random</i>
ML .....	<i>Machine-Learning</i>
mm .....	Millimeter
MPRAGE .....	<i>Magnetization Prepared Rapid Gradient Echo</i>
MRT .....	Magnetresonanztomographie

ms	Millisekunden
NaN	<i>Not a number</i>
NIfTI	<i>Neuroimaging Informatics Technology Initiative</i>
NLR	<i>Negative likelihood ratio</i>
NPV	<i>Negative predictive value</i>
PCA	<i>Principal Component Analysis</i>
PLR	<i>Positive likelihood ratio</i>
PPV	<i>Positive predictive value</i>
PsychKG	Psychisch-Kranken-Gesetz
PTBS	Posttraumatische Belastungsstörung
RCT	Randomisierte kontrollierte Studien
ROC	<i>Receiver operating characteristic</i>
ROI	<i>Region of interest</i>
RWD	<i>Real-world data set</i>
SPM12	<i>Software Statistical Parametric Mapping</i>
SVM	<i>Support Vector Machine</i>
TE	<i>Echo time</i>
TN	<i>True negative</i>
TP	<i>True positive</i>
TR	<i>Repetition time</i>
UKD	Uniklinikum Düsseldorf
ZAN-BPD	<i>Zanarini Rating Scale for Borderline Personality Disorder</i>

# Inhaltsverzeichnis

<b>1</b>	<b><i>Einleitung</i></b> .....	<b>1</b>
<b>1.1</b>	<b>Die emotional instabile Persönlichkeitsstörung vom Borderline Typ.....</b>	<b>1</b>
1.1.1	Begriffserklärung, Epidemiologie und Symptomatik .....	1
1.1.2	Therapie.....	3
1.1.3	Komorbidität mit Depression und anderen psychiatrischen Erkrankungen.....	4
1.1.4	Relevante Hirnregionen.....	5
<b>1.2</b>	<b>Exkurs: <i>Machine-Learning</i> basierte Klassifikation .....</b>	<b>6</b>
1.2.1	Datenmenge .....	8
1.2.2	SVM.....	10
1.2.3	<i>Ensemble learning</i> .....	11
1.2.4	Leistungsbewertung .....	12
1.2.5	Herausforderungen des <i>Machine-Learnings</i> .....	14
1.2.6	Kreuzvalidierung .....	15
1.2.7	Anwendungsbereiche des <i>Machine-Learnings</i> .....	17
<b>1.3</b>	<b>Anwendung des ML in der Psychiatrie und bei Patienten mit BPS.....</b>	<b>18</b>
<b>1.4</b>	<b>Zielsetzung.....</b>	<b>19</b>
<b>2</b>	<b><i>Material und Methoden</i> .....</b>	<b>20</b>
<b>2.1</b>	<b>Stichprobenbeschreibung.....</b>	<b>20</b>
<b>2.2</b>	<b>Datengewinnung.....</b>	<b>20</b>
<b>2.3</b>	<b>Analyse der klinisch-demographischen Daten.....</b>	<b>25</b>
<b>2.4</b>	<b>MRT-Sequenzparameter.....</b>	<b>26</b>
<b>2.5</b>	<b>Präprozessierung der MRT-Daten .....</b>	<b>26</b>
<b>2.6</b>	<b><i>Machine-Learning</i> an strukturellen Daten.....</b>	<b>28</b>
2.6.1	Algorithmusparameter .....	28
2.6.2	Vorverarbeitung der MRT-Daten.....	29
2.6.3	Kreuzvalidierungsschema .....	29
2.6.4	Leistungsbewertung .....	29

2.6.5	Visualisierung der Ergebnisse .....	30
<b>2.7</b>	<b>Machine-Learning an klinisch-demographischen Daten.....</b>	<b>31</b>
2.7.1	Algorithmusparameter .....	31
2.7.2	Vorverarbeitung.....	31
<b>2.8</b>	<b>Stacking.....</b>	<b>31</b>
<b>2.9</b>	<b>Bagging .....</b>	<b>31</b>
<b>2.10</b>	<b>Condensed version des klinisch-demographischen Klassifikators .....</b>	<b>32</b>
<b>3</b>	<b>Ergebnisse.....</b>	<b>33</b>
3.1	Statistische Unterschiede der Zielgruppen.....	33
3.2	Ergebnisse des <i>Machine-Learnings</i> .....	37
3.2.1	Ergebnisse des Originalmodells .....	39
3.2.2	Ergebnisse der <i>condensed version</i> .....	43
3.2.3	Signifikante Hirnareale .....	47
<b>4</b>	<b>Diskussion.....</b>	<b>50</b>
4.1	Bestimmung der Effizienz des Models .....	50
4.2	Relevante features.....	52
4.3	Nicht signifikante Klassifikationsmodelle.....	56
4.4	Limitationen .....	60
4.5	Schlussfolgerung und Ausblick .....	61
	<b>Literatur- und Quellenverzeichnis .....</b>	<b>64</b>
	<b>Anhang.....</b>	<b>79</b>

# 1 Einleitung

## 1.1 Die emotional instabile Persönlichkeitsstörung vom Borderline Typ

### 1.1.1 Begriffserklärung, Epidemiologie und Symptomatik

Die Borderline Persönlichkeitsstörung (BPS) stellt eine psychische Erkrankung dar, welche den Betroffenen selbst, die Angehörigen und das Gesundheitswesen vor große Herausforderungen stellt (Leichsenring et al., 2011; Lieb et al., 2004; Skodol, Gunderson, McGlashan, et al., 2002). Die Prävalenz der BPS in der Bevölkerung liegt zwischen 0,7 % und 4,5 % (Fiedler, 2018), wobei die durchschnittliche Prävalenz ca. 1,6 % beträgt (Grabe & Giertz, 2020). Etwa 12 % der Patienten in ambulanten psychiatrischen Einrichtungen und 22 % derer in stationären psychiatrischen Kliniken sind von der BPS betroffen (Ellison et al., 2018). Die Erkrankung betrifft in der Bevölkerung nahezu gleich viele Frauen und Männer (Grabe & Giertz, 2020). Dieses Verhältnis verschiebt sich in klinischen Stichproben in Richtung der Frauen (Paris, 2002).

BPS wird in der *International Statistical Classification of Diseases and Related Health Problems-10* (ICD-10) (Graubner, 2013) als Untergruppe der emotional instabilen Persönlichkeitsstörung kategorisiert. Die emotional instabilen Persönlichkeitsstörungen werden insbesondere durch die Impulsivität, Konfliktbereitschaft und emotionale Instabilität gekennzeichnet. Der Borderline-Typ beinhaltet ferner die Komponenten parasuizidale Handlungen und Suizidversuche, chronisches Gefühl der Leere, Störungen des Selbstbildes und intensive, aber inkonsistente zwischenmenschliche Beziehungen (Graubner, 2013). Das *Diagnostic and Statistical Manual of Mental Disorders, fourth edition* (DSM-IV) (American Psychiatric Association & Association, 1994) fügt diesen Kriterien die affektive Instabilität darunter auch übermäßige Wut, stressbedingte paranoide Ideen oder dissoziative Symptome und hektisches Bemühen zur Vermeidung des Verlassen Werdens hinzu. Nach DSM-IV ist das Vorhandensein von mindestens fünf der aufgeführten Kriterien für eine Diagnosestellung notwendig (Oldham, 2006; Skodol, Gunderson, Pfohl, et al., 2002). So können vielzählige Subtypen entstehen, die sich zusätzlich durch komorbide Erkrankungen differieren können, was

eine individuelle Untersuchung der Persönlichkeit und ursächlichen Pathologie notwendig macht (Oldham, 2006; Skodol, Gunderson, Pfohl, et al., 2002). Dies ist besonders im Hinblick auf die Tatsache relevant, dass sich bei BPS die Funktionsfähigkeit durch Verbesserung der Psychopathologie erhöhen kann (Skodol et al., 2005).

Die funktionelle Beeinträchtigung durch BPS manifestiert sich in diversen Lebensbereichen (Modestin & Villiger, 1989; Nakao et al., 1992; Swartz et al., 1990), insbesondere bei der Arbeit, in sozialen Beziehungen und in der Freizeit (Skodol, Gunderson, McGlashan, et al., 2002). Persönlichkeitsstörungen, und insbesondere die BPS, stellen eine erhebliche finanzielle Last für das Gesundheitswesen dar, unter anderem durch direkte medizinische Kosten wie die umfangreiche und kontinuierliche Nutzung von Psychotherapien und Psychopharmaka sowie anfallenden indirekten Kosten (Bender et al., 2001; Jackson & Burgess, 2004; Sansone et al., 2011; Soeteman et al., 2008). Demnach konsultieren Patienten mit BPS signifikant öfter Hausärzte und Fachärzte als Patienten ohne BPS (Sansone et al., 2011) sowie Patienten mit anderen Persönlichkeitsstörungen (Jackson & Burgess, 2004). Beispielsweise fanden Bender und Kollegen (2001) heraus, dass BPS bei der Einnahme von nahezu allen Medikamentengruppen zur Behandlung psychischer Störungen die meisten anderen psychiatrischen Erkrankungen übertrifft. Außerdem nahmen Patienten mit BPS psycho-soziale Therapien häufiger in Anspruch als Patienten mit anderen psychiatrischen Erkrankungen sowie anderen Persönlichkeitsstörungen (Bender et al., 2001). In einem 6-jährigen Zeitraum konnte gezeigt werden, dass der Großteil der BPS-Patienten ambulante Leistungen wie die Psychotherapie weiterhin in Anspruch nahmen und die Einnahme an Psychopharmaka fortführten (Hennen & Silk, 2004). Zusätzlich, jedoch, weisen BPS-Patienten eine hohe Therapieabbruchquote auf (Horner & Diamond, 1996; Kelly et al., 1992), welche die Herausforderung bei der Behandlung der Erkrankung erhöht.

Das für die Erkrankung charakteristische suizidale Verhalten trägt ebenfalls zu einer häufigeren Beanspruchung medizinischer Leistungen bei (American Psychiatric Association, 2001; Black et al., 2004; Oldham, 2006). Die Suizidrate wird zwischen 3 % (McGlashan, 1986) und ca. 10 % (Paris & Zweig-Frank, 2001) angegeben. Die Rate an unternommenen Suizidversuchen übersteigt die der

vollendeten Suizide deutlich (Soloff et al., 1994). Sexueller Missbrauch in der Kindheit stellt einen wichtigen Faktor dar, der das suizidale Verhalten im Erwachsenenalter steigert (Soloff et al., 2002). Weitere Prädiktoren wie das Fehlen von Inanspruchnahme ambulanter Leistungen oder ein niedriger soziökonomischer Status weisen auf ein erhöhtes Risiko für suizidales Verhalten hin (Soloff & Chiappetta, 2012).

### **1.1.2 Therapie**

Metaanalysen und Review-Artikel zeigten, dass Psychotherapeutische Ansätze, insbesondere die Dialektisch Behaviorale Therapie (DBT) (Linehan, 1993; Linehan et al., 1991), als effektivste Behandlungsstrategien für BPS gelten (Cristea et al., 2017; Kliem et al., 2010; Panos et al., 2014; Stoffers et al., 2012). Die ursprünglich für chronisch parasuizidale Patienten entwickelte DBT senkte in der von Linehan et al. (1991) durchgeführten randomisierten klinischen Studie sowohl die Anzahl und Schwere der Parasuizide (d.h. selbstverletzende Verhaltensweisen, welche eine nichttödliche Absicht haben) als auch die Anzahl an Tagen in der stationären psychiatrischen Behandlung. Zudem zeigte sich ein längerer Verbleib in der DBT-basierten Einzeltherapie im Vergleich zur Standardbehandlung (Linehan et al., 1991). Der signifikante Effekt von DBT auf Suizid und parasuizidales Verhalten wurde auch in einer Metaanalyse durch Panos et al. (2014) untermauert (Panos et al., 2014).

Die Behandlung des suizidalen Verhaltens steht an erster Stelle der DBT. In weiteren Schritten werden Faktoren, welche die Behandlung beeinträchtigen und die Lebensqualität einschränken, berücksichtigt. Schließlich mündet die Therapie in der Entwicklung und Verbesserung allgemeiner Bewältigungsfähigkeiten (Linehan et al., 1994).

Das Kernelement der DBT bildet die Balance und Umsetzung von Akzeptanz und Veränderung (Robins & Rosenthal, 2011; Swenson et al., 2001). Die Akzeptanz des Patienten wird unter anderem durch Bestätigung gefördert, indem der Therapeut dem Patienten die Sinnhaftigkeit seiner Handlungen und Reaktionen vergewissert. Zudem werden zur besseren Wahrnehmung des Erlebens Achtsamkestrainings angeboten (Robins & Rosenthal, 2011). Die Veränderung dysfunktionaler Verhaltens- und Denkmuster erfolgt auf der Grundlage von standardmäßigen

Kognitiven- und Verhaltenstherapien (Robins & Rosenthal, 2011). In unterschiedlichen Therapiesettings wie der Einzeltherapie, der Gruppentherapie und dem telefonischen Coaching wird das Erlernen neuer Fertigkeiten angestrebt, die im Alltag umgesetzt werden sollen (Kliem et al., 2010). Dysfunktionale Verhaltensweisen werden durch Behandlungsansätze wie der Förderung der Achtsamkeit, der Emotionsregulation, der Stressbewältigung und der zwischenmenschlichen Kompetenz behandelt (Swenson et al., 2001). Trotz der genannten positiven Effekte auf die BPS-Behandlung kann die DBT kurzfristig betrachtet insbesondere im Hinblick auf Kosten für die intensiven DBT-Interventionen möglicherweise teurer als die Standardtherapie sein (Priebe et al., 2012). Prädiktive Faktoren des Behandlungserfolgs eines Patienten (*therapy response*) von DBT könnten daher eine individualisierte Therapie effizienter planen und zu einer finanziellen Entlastung des Gesundheitssystems führen.

Eine standardisierte medikamentöse Behandlung der BPS existiert nicht, sondern orientiert sich an einzelnen Symptomen und komorbidien Erscheinungen. Die Einnahme ist somit individuell unterschiedlich (Stoffers et al., 2010). Zum Beispiel können Antidepressiva vor allem bei ausgeprägter komorbider Depression empfohlen werden (Jariani et al., 2010; Stoffers & Lieb, 2015). Weiterhin werden vor allem Antipsychotika der ersten und zweiten Generation, Stimmungsstabilisatoren, sowie Antikonvulsiva verschrieben (Abraham & Calabrese, 2008). Zudem sind Effekte der Omega 3 Fettsäuren, Opioidantagonisten, Alpha-2 adrenerge Agonisten und Oxytocin Gegenstand der Forschung (Stoffers & Lieb, 2015). In randomisierten kontrollierten Studien (RCT) wurden insbesondere positive Effekte der Antipsychotika zweiter Generation und der Stimmungsstabilisatoren festgestellt, wobei diese gemeinsam mit den Omega 3 Fettsäuren vor allem gegen affektive Symptome und die impulsive Komponente der BPS wirkungsvoll sein können (Bellino et al., 2011; Lieb et al., 2010; Stoffers et al., 2010) .

### **1.1.3 Komorbidität mit Depression und anderen psychiatrischen Erkrankungen**

Zanarini und Kollegen (1998, 2004) analysierten das komorbide Auftreten von psychiatrischen Erkrankungen des DSM-III (Association & Spitzer, 1980) bei BPS-Patienten verglichen mit deren Auftreten bei anderen

Persönlichkeitsstörungen (Zanarini et al., 1998; Zanarini et al., 2004). Die Ergebnisse wiesen einen signifikanten Unterschied hinsichtlich des Vorkommens von affektiven Störungen sowie Angststörungen darunter vor allem Posttraumatische Belastungsstörung (PTBS), Panikstörung und Sozialphobie auf, welche bei BPS-Patienten häufiger auftraten. Die Autoren stellten zusätzlich ein erhöhtes Auftreten von Essstörungen, sowie einfachen Phobien und schwere Depression bei BPS-Patienten fest (Zanarini et al., 1998). In weiteren Studien zeigten sich zudem erhöhte Komorbiditätsraten von BPS mit Substanzmissbrauch (McGlashan et al., 2000), Zwangsstörungen, somatoformen Störungen und bipolaren Störungen (Zimmerman & Mattia, 1999), sowie, nach Korrektur von soziodemographischen Variablen und anderen psychiatrischen Erkrankungen, mit bipolarer Störung, schizotypen- und narzisstischer Persönlichkeitsstörung (Grant et al., 2008).

Das gemeinsame Auftreten der BPS mit anderen psychiatrischen Erkrankungen kann bislang nicht vollständig erklärt werden (Gunderson & Elliott, 1985; Skodol, Gunderson, Pfohl, et al., 2002). Mögliche Annahmen sind, dass die Psychopathologie der BPS zusätzliche Erkrankungen verursacht, dass BPS typisches Verhalten durch deren Einfluss hervorgerufen wird, dass prinzipiell kein Zusammenhang besteht oder dass sie sich durch das gehäufte Auftreten in ihrer Ausprägung beeinflussen (Gunderson & Elliott, 1985; Skodol, Gunderson, Pfohl, et al., 2002).

#### **1.1.4 Relevante Hirnregionen**

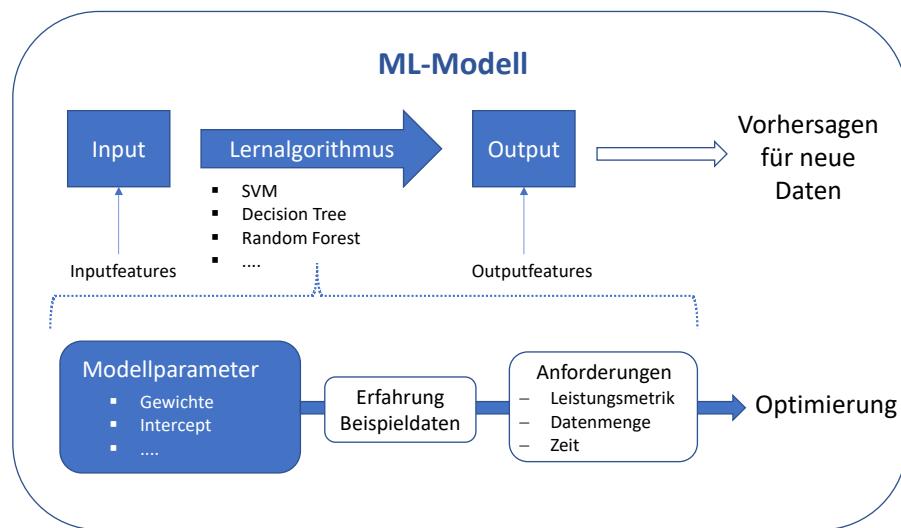
In einer Metaanalyse untersuchten Schulze et al. (2016) Hirnregionen, welche BPS-Patienten erfolgreich von gesunden Kontrollpersonen unterscheiden und somit spezifisch für diese Erkrankung zu sein scheinen (Schulze et al., 2016). Anhand spezifischen Einschlusskriterien unter anderem der Analyse der Grauen Substanz (GM) und des kompletten Hirns wurden 10 Studien integriert. Dabei wurde festgestellt, dass das GM-Volumen bei BPS insbesondere in den Regionen bilateral temporaler Gyrus, Pars opercularis des rechten unteren Frontalgyrus und rechter Hippocampus geringer war als bei der Kontrollgruppe. Dagegen zeigte sich in den Hirnregionen linkes Rolandisches Operculum, rechter mittlerer Frontalgyrus, rechtes Cerebellum und rechtes supplementär-motorisches Areal ein größeres GM-Volumen verglichen mit der Kontrollgruppe (Schulze et al., 2016).

Grecucci und Kollegen (2022) erforschten strukturelle Hirndaten, welche BPS von einer gesunden Kontrollgruppe und einer Kontrollgruppe mit Bipolarer Störung unterschieden. Hierbei zeigten sich das rechte Putamen, die rechte Amygdala, obere und mittlere Orbitalregionen, der Gyrus fusiformis und das linke 7b Areal des Cerebellum unterschiedlich zwischen BPS und beider Kontrollgruppen (Grecucci et al., 2022), u.a. der linke Thalamus, der Gyrus lingualis und das linke Pallidum waren unterschiedlich zwischen BPS Patienten und gesunden Kontrollprobanden.

## **1.2 Exkurs: *Machine-Learning* basierte Klassifikation**

*Machine-Learning* (ML), erstmals eingeführt durch (Samuel, 1959), bezeichnet die Optimierung und Steigerung der Leistung eines Computerprogramms bzw. eines sogenannten Algorithmus unter Zuhilfenahme der Erfahrung, welche an einer vorher bestimmten Leistungsmetrik gemessen wird (Mitchell, 1997). Das Lernen bezieht sich dabei auf die Optimierung der Parameter des Modells unter wiederholter Verwendung von Beispieldaten oder vorheriger Entscheidungen (Alpaydin, 2020; Bonacorso, 2017; El Naqa & Murphy, 2015). Dabei erkennt und lernt der Algorithmus eigenständig, relevante Beziehungen aus der Datenmenge zu extrahieren (Bishop & Nasrabadi, 2006). Grundlage für diesen Prozess stellt die Suche und Erkennung von Mustern innerhalb der eingehenden Daten dar (Fisher, 1938). Die Mustererkennung beschreibt das Vorgehen des Algorithmus bei der Filterung von nützlichen aus irrelevanten Informationen und deren optimale Beschreibung. Ziel ist ein selbst lernendes und problemlösendes System, welches Schlussfolgerungen aus einer Stichprobe zieht und schrittweise besser wird, indem es die Anforderungen an Leistung, Zeit und Datenmenge besser erfüllt (Alpaydin, 2020).

Lernalgorithmen stellen ein mathematisches Konstrukt dar, welche aus Eingabedaten (*inputs*) Vorhersagen (*outputs*) erzeugen (Baloglu et al., 2021; Bi et al., 2019; Janiesch et al., 2021). In der Anwendung eines Algorithmus erfolgt initial die Dateneingabe: Die Daten werden bereinigt, vorverarbeitet und in einen Trainings- und einen Test-Datensatz aufgeteilt. Anhand des Lernalgorithmus wird ein repräsentatives und die Zusammenhänge in der Datenmenge beschreibendes Modell entwickelt, welches komplexe Muster und Beziehungen erkennt, ohne explizit für die individuelle Fragestellung programmiert zu sein (Bishop & Nasrabadi, 2006). Das Modell ist somit das Ergebnis des wiederholten Lernvorgangs durch den Algorithmus und spiegelt die eingegebenen Daten wider. Anschließend erfolgt die Optimierung der Modellparameter, sodass auf deren Grundlage Vorhersagen generiert werden können (Gollapudi, 2016) (s. **Abb.1**).



**Abb.1: Schematische Darstellung eines ML-Modells:** ML: *Machine-Learning*; SVM: *Support Vector Machine*; Der Lernalgorithmus erkennt Zusammenhänge und Beziehungen der vorgegebenen *input data* und der zugehörigen *output data*. Dabei bestimmen die Modellparameter die Fähigkeit, diese Zusammenhänge präzise festzustellen. Im Zuge des Lernprozesses, welcher die Optimierung der Parameter enthält, entsteht ein Modell, das Entscheidungen und Vorhersagen auf bisher unbekannten Daten generieren kann.

Beispielsweise ist die Optimierung des C- und des Gamma Parameters im Rahmen des *Support Vector Machine* Algorithmus (SVM) anzuführen. Hierbei reguliert der C-Parameter anhand der Gewichtung von Klassifikationsfehlern das Spannungsverhältnis zwischen einer optimalen Klassifizierung und der Verallgemeinerungsfähigkeit des Modells (Gaspar et al., 2012). Der Gamma- Parameter

bestimmt die Form der Grenze, anhand dessen die Datenpunkte voneinander getrennt werden (Al-Mejibli et al., 2020).

Geschieht die Eingabe der Daten in tabellarischer Form, so repräsentiert jede Zeile ein Beispiel oder *item*. Abhängig von der Art des Datensatzes können sie z.B. einzelne Patienten oder Bilder abbilden (Gollapudi, 2016; Mohri et al., 2018). Die Spalten werden als *features* oder Variablen bezeichnet. Diese stellen Eigenschaften und Merkmale der *items* dar und können als *in-* oder *output features* fungieren (Bi et al., 2019; Gollapudi, 2016; Mohri et al., 2018). *Input features* (unabhängige *features*) werden dem Lernalgorithmus zugeführt, um Vorhersagen in Form von *output features* (abhängigen *features*) zu generieren (Bi et al., 2019). Die *feature selection* zielt darauf ab, die kleinstmögliche Anzahl an Merkmalen auszuwählen ohne dabei die Generalisierbarkeit einzuschränken (Abe & Abe, 2010). Dadurch wird angestrebt, die Vorhersagegenauigkeit zu verbessern, die Rechenlast zu mindern und die Interpretation der Ergebnisse zu vereinfachen (Orru et al., 2012). Im Rahmen der *feature extraction* werden zusätzlich durch Hervorheben diskriminierender Informationen aus dem Pool der durch *feature selection* generierten *features* neue *features* gebildet (Goodfellow et al., 2016; Shah & Patel, 2016). Die Wichtigkeit der einzelnen *features* und deren Einfluss auf die Vorhersage des Modells wird anhand der *feature weights* ausgedrückt (Tian & Zalesky, 2021). Die Zusammenstellung von *items* mit zugewiesenen *features* wird als Datensatz bezeichnet (Gollapudi, 2016).

### 1.2.1 Datenmenge

Innerhalb von ML-basierten Analysen ist man in der Lage, verschiedene Datenmodalitäten zu verarbeiten und zu kombinieren. Algorithmen können neben numerischen Daten auch Bilder, strukturelle anatomische Daten, Texte und Zeitreihen zur Entwicklung von Analysemodellen verwenden (Janiesch et al., 2021). Dabei ist insbesondere darauf zu achten, dass die Daten kompatibel sind und vom ML-Algorithmus erkannt werden. Außerdem sollten sie von manuellen Fehlern bereinigt vorliegen (Gollapudi, 2016). Die eingegebenen Daten werden in Trainings- und Testdaten unterteilt. Anhand der Trainingsdaten wird ein analytisches Modell entwickelt, welches validiert und an Testdaten angewendet wird, um die Effizienz und Güte des Modells zu evaluieren, welches letztlich Vorhersagen auf unbekannte neue Daten generieren soll (Gollapudi, 2016).

ML-Analysen können als *supervised*- oder *unsupervised learning* durchgeführt werden. *Supervised learning* ist eine Unterform des ML, bei der die zu analysierenden Eingabedaten zu bereits gelabelten Ausgabedaten zugeordnet vorliegen (Bi et al., 2019). *Labels* kennzeichnen Zielvariablen, welche durch ML-Modelle vorhergesagt werden sollen. Dies geschieht durch spezifische Beschreibung der Rohdaten unter anderem durch Wertzuweisung oder kategoriale Einteilung (Bi et al., 2019; Mohri et al., 2018). Auf Grundlage der Ein- und Ausgabepaare im Datensatz kann der Lernalgorithmus hinsichtlich der Gütekriterien wiederholt trainiert und verbessert werden (Bonacorso, 2017; Mahesh, 2020). Die Aufgabe besteht in der Identifikation der Beziehung beider Daten, um für neue Eingaben die korrekten Ausgabedaten vorhersagen zu können (El Naqa & Murphy, 2015). Mithilfe dieser Lernstrategie können unter anderem Vorhersagen für eine numerischen Größe (Regression) oder für die Zugehörigkeit zu einer bestimmten Gruppe (Klassifikation) generiert werden (Bonacorso, 2017; Janiesch et al., 2021).

Beim *unsupervised learning* werden keine spezifischen Ausgabedaten bestimmt, sodass Zusammenhänge zwischen den Eingabedaten ermittelt werden (Sharma et al., 2021). Dabei werden Datenmengen aufgrund gehäuftem Auftreten von Merkmalen, der Gemeinsamkeit und Ähnlichkeit ihrer Eigenschaften separiert und in Gruppen („*Cluster*“) unterteilt (Alpaydin, 2020; Janiesch et al., 2021). Die Herausforderung besteht in der präzisen Trennung der Datensegmente (Bonacorso, 2017). Praktische Anwendung findet diese Methode beim *Clustering* und der Dimensionsreduktion (Bishop & Nasrabadi, 2006; Grira et al., 2004).

Die Dimensionsreduktion, auch als *feature reduction* bezeichnet, wird unter anderem vor der Eingabe in das *supervised learning* verwendet, um für das Lernen bedeutendere Dimensionen herauszuarbeiten, indem weniger wichtige reduziert werden (Gollapudi, 2016). Eine häufig genutzte Technik ist die Hauptkomponentenanalyse, eingeführt durch (Hotelling, 1933). Sie stellt die lineare Verknüpfung ursprünglicher Variablen zu neuen zueinander orthogonal stehenden Variablen, sogenannten Hauptkomponenten dar, sodass die Variation optimal erfasst wird (Kurita, 2019; Mahesh, 2020) .

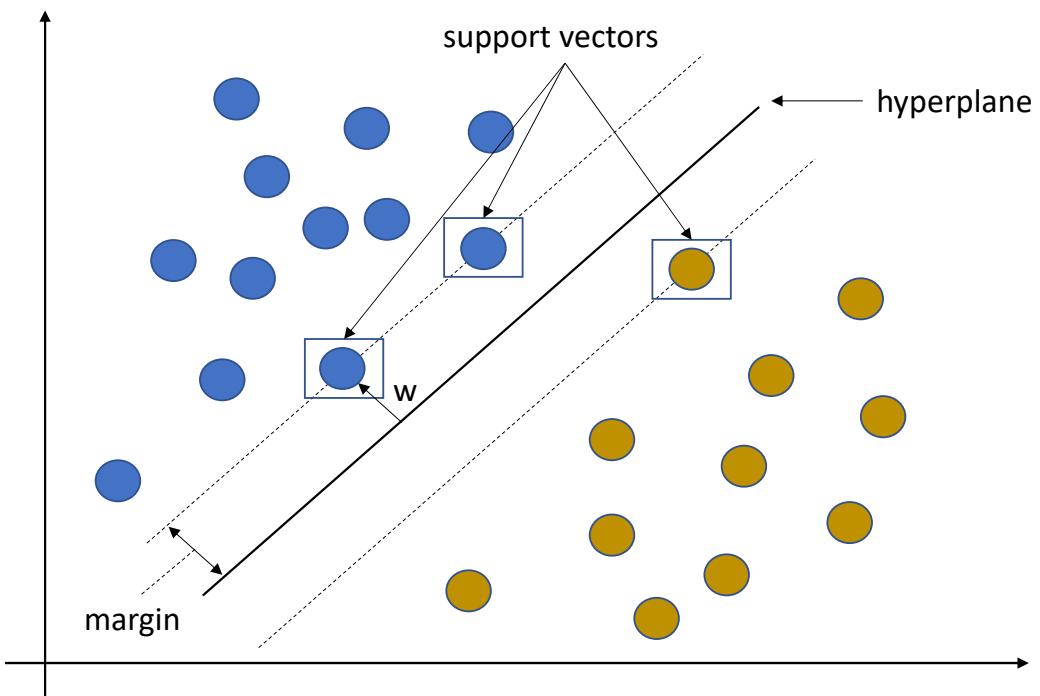
Das *semisupervised learning* kombiniert die oben beschriebenen Lernstrategien, indem Lernalgorithmen auf gelabelte und ungelabelte Datensätze appliziert werden (Gollapudi, 2016). Insbesondere bei bereits vorliegenden unbeschrifteten oder großen Datensätzen lässt sich somit der Aufwand vermeiden, die komplette Datenmenge zu labeln (Bi et al., 2019; Mahesh, 2020) Stattdessen wird ein an gelabelten Daten trainiertes Modell durch ungelabelte Daten ergänzt, um die Leistung zu optimieren (El Naqa & Murphy, 2015; Mohri et al., 2018)

### 1.2.2 SVM

SVM ist ein methodischer Ansatz zur Klassifikation. Das Ziel der Klassifikation besteht darin, Vorhersagen für eine kategoriale Zuordnung zu treffen, bei der jedes *item* im Datensatz zu einer im Vorhinein bestimmten Kategorie zugewiesen werden soll (Bi et al., 2019; Mohri et al., 2018). Die Trennung der Kategorien erfolgt mittels Aufstellung einer mathematischen Funktion, der sogenannten Diskriminante (Alpaydin, 2020). Dieser Diskriminierungsmechanismus hebt Unterscheidungsmerkmale hervor und stellt Beziehungen zwischen den Eingabedaten auf, welche die Gruppeneinteilung ausschlaggebend beeinflussen (Gollapudi, 2016).

Bei dem auf *supervised learning* basierenden SVM-Ansatz erfolgt die Abbildung aller *items* im Vektorraum, in dem die Aufstellung und Optimierung der Hyperebene (*hyperplane*), einer linearen Trennfunktion, stattfindet (Cortes & Vapnik, 1995; Hastie et al., 2009; James et al., 2013) (s. **Abb.2**). Die Flächen ober- und unterhalb der *hyperplane* repräsentieren die Zuordnung der *items* zu den Kategorien. Die der *hyperplane* nächstgelegene Vektoren der *items* werden als *support vectors* bezeichnet und bestimmen entscheidend deren Verlauf (Cortes & Vapnik, 1995). Das Ziel der SVM ist die Generierung des maximalen Abstands der *support vectors* zur *hyperplane* zur präzisen Trennung (Cortes & Vapnik, 1995; Cristianini & Shawe-Taylor, 2000; Orru et al., 2012). Durch den Einsatz von Kernels lassen sich auch nichtlinear verteilte *items* durch Transformation in höhere Dimensionen linear trennen (Hastie et al., 2009; James et al., 2013). Der C-Parameter reguliert dabei die Toleranz an Missklassifizierungen zur Vermeidung von *overfitting* und wird im Verlauf optimiert (Cortes & Vapnik, 1995; Dwyer et al., 2018). Die Funktion der *hyperplane* lässt sich durch einen Normalenvektor (Gewichtsvektor) und Bias (Verschiebungsparameter) beschreiben, welche im

Lernprozess angepasst werden (Hastie et al., 2009; Zarogianni et al., 2013). Ein wesentlicher Vorteil der SVM besteht darin, dass sie effizient mit hochdimensionalen Daten, die vor allem bei strukturellen Daten vorliegen, umgehen kann (Zarogianni et al., 2013).



**Abb.2: Schematische Darstellung der Funktionsweise einer SVM:** Quelle: In Anlehnung an (Zarogianni et al., 2013) und (Cortes & Vapnik, 1995). Dargestellt ist die Trennung der blauen und braunen Kreise, repräsentativ für zwei unterschiedliche Kategorien, durch *hyperplane*. Dabei bilden die Kreise einzelne *items* im Vektorraum ab. Die durch Rechtecke eingekreisten Kreise befinden sich am nächsten zum *hyperplane* und werden als *support vectors* bezeichnet. Sie bestimmen die Breite des *margin*. Zu jedem *item* existiert ein Gewichtsvektor *w*, welcher als Normalenvektor zum *hyperplane* fungiert.

### 1.2.3 Ensemble learning

Das *ensemble learning* bezeichnet die Verwendung und Kombination mehrerer Modelle, um eine Vorhersage zu generieren, dessen Vorhersagegenauigkeit dadurch verbessert wird (Sagi & Rokach, 2018). Es stehen unterschiedliche Methoden des ML zur Verfügung, darunter *stacking* und *bagging*.

Unter *stacking* („stapeln“) wird das Zusammenführen der *decision scores* z.B. unterschiedlicher Datenmodalitäten verstanden. Es werden Vorhersagen zweier oder mehrerer Modelle, in der vorliegenden Dissertation klinische und bildmorphologische, vereint und ein neues Modell kreiert, welches anhand der fusionierten Informationen lernt und dadurch eine verbesserte Klassifikation verspricht

(Wolpert, 1992). Konkret werden Vorhersagen der einzelnen Klassifikatoren von einem Meta-Klassifikator aufgenommen, welcher Zusammenhänge zwischen den Ausgaben der einzelnen Klassifikationen und den tatsächlich korrekten Gruppen/ Zuordnungen erkennt (Polikar, 2006).

Bei *bootstrap aggregation* oder *bagging* erfolgt die Kombination der *decision scores* aller vorher bereits analysierten Modalitäten. Hier liegen bereits *decision scores* vor, welche anschließend zusammengeführt werden. Der aggregierte *decision score* stellt den Durchschnitt der einzelnen *decision scores* dar und ermöglicht eine aussagekräftigere Gesamtbewertung. Dies geschieht, indem der komplette Trainingsdatensatz zufällig in Teilmengen unterteilt, diese durch mehrere Klassifikatoren (wie der SVM) trainiert und die Vorhersagen derer kombiniert wird (Breiman, 1996; Polikar, 2006).

#### 1.2.4 Leistungsbewertung

Die Leistung eines ML-Modells bezieht sich auf das Ausmaß, inwieweit die anfangs dargestellte Aufgabe, in dieser Dissertation die Klassifikation in (Teil)Responder und Non-Responder anhand prädiktiver *features*, erfüllt wurde (Gollapudi, 2016). Die Grundlage für die Evaluation der Modellleistung im Rahmen der Klassifikation lässt sich anhand einer 4-Felder Tafel darstellen (die sog. Konfusion-Matrix), die durch den Vergleich der Vorhersagen hinsichtlich der Zuordnung in die Zielgruppen mit der tatsächlichen Gruppenzugehörigkeit entsteht (s. **Tabelle 1**).

**Tabelle 1: Evaluationsmetriken von Klassifikationsmodellen**

	Positive Vorhersage	Negative Vorhersage
Tatsächlich Positiv	TP	FN
Tatsächlich Negativ	FP	TN

**Anmerkung:** TP: *True positive* (Richtig Positiv); FN: *False negative* (Falsch Negativ); FP: *False positive* (Falsch Positiv); TN: *True negative* (Richtig Negativ)

Dabei besitzen die einzelnen Evaluationsmetriken *true positive* (TP), *true negative* (TN), *false positive* (FP) und *false negative* (FN) eine unterschiedliche Bedeutung bezüglich der Ergebnisbeurteilung und sollten kontextabhängig bewertet werden (Gollapudi, 2016). Gängiges Gütekriterium eines ML-Modells, welche die

obigen Evaluationsmetriken zunutze macht, ist *accuracy* („Genauigkeit“) (Michie et al., 1994; Salton, 1989). *Accuracy* bezeichnet das Verhältnis der korrekten Vorhersagen zu allen generierten Vorhersagen und wird durch die Formel

$$(TP+TN)/ (TP+TN+FN+FP)$$

beschrieben (Jiao & Du, 2016; Sokolova et al., 2006). Insbesondere bei der Beurteilung und dem Vergleich der Klassifikationsergebnisse wird ein besonderes Augenmerk auf die *accuracy* der Vorhersagen gelegt.

*Area under the curve* (AUC) beschreibt die Fläche unter der *Receiver operating characteristic* (ROC) -Kurve, bei der die Sensitivität gegenüber der falsch positiven Rate aufgetragen ist (Bradley, 1997). Außerdem kann die Korrektheit der positiven Vorhersagen anhand des *positive predictive value* (PPV) und die negativen Vorhersagen mit der *negative predictive value* (NPV) bestimmt werden, um herauszufinden wie genau die Zuordnung zu bzw. Ausschluss aus der Zielgruppe erfolgte. Diese Genauigkeit wird mithilfe des *prognostic summary index* in Form von

$$PPV+ NPV- 100$$

zusammengefasst.

Die *positive likelihood ratio* (PLR)

$$\text{Sensitivität/ (1- Spezifität)}$$

und die *negative likelihood ratio* (NLR)

$$(1-\text{Sensitivität})/ \text{Spezifität}$$

dienen dem ganzheitlichen Verständnis der Testleistung (Shreffler & Huecker, 2022). Zur Evaluierung der Übereinstimmung zwischen der vorhergesagten Klassifikation und der tatsächlichen Klassenzugehörigkeit der *items* kann der robuste Cohen's Kappa Koeffizient verwendet werden (Cohen, 1960; Vieira et al., 2010). Demnach beträgt die Übereinstimmung durch Zufall gleich 0 und größere Werte weisen auf eine bessere Übereinstimmung hin. Unter Zuhilfenahme des p-Werts kann die Signifikanz der Übereinstimmung bewertet werden (Whitley & Ball, 2002).

Zur Ermittlung von *features*, welche sich signifikant auf die Klassifikation auswirken, kann die *sign-consistency based variable importance* oder *sign-based consistency*, die für eine erhöhte Reproduzierbarkeit und Klassifizierungsgenauigkeit plädiert, hinzugezogen werden (Gómez-Verdejo et al., 2019). Es reicht von 0 bis in den positiven Bereich, wobei positivere Werte eine bessere Konsistenz repräsentieren, dagegen 0 beschreibt, dass die Vorhersage gleichermaßen positiv wie negativ ist oder mit der Gewichtung 0 rausgelassen wurde (Gómez-Verdejo et al., 2019).

Unter Verwendung der *sign-based consistency* ist die Erstellung einer *condensed version* des Modells möglich. Dabei bilden signifikante Merkmale die Gesamtheit der Daten dieser neuen Klassifikation (Koutsouleris et al., 2021; Schmitgen et al., 2019). Die *condensed version* eines Klassifikators ermöglicht eine einfache Handhabung im klinischen Setting. Ziel der *condensed version* ist die schnellere Datenerhebung bei gleichzeitiger Erhaltung einer höchstmöglichen Genauigkeit der Vorhersage.

### **1.2.5 Herausforderungen des *Machine-Learnings***

Das Ziel des ML ist zunächst die optimale Beschreibung des Trainingssatzes, indem der Trainingsfehler minimiert wird und anschließend die Implementierung des Modells auf bisher unbeobachtete Daten mithilfe der Reduzierung des Testfehlers, auch als Generalisierung bezeichnet (Goodfellow et al., 2016; Mohri et al., 2018).

Bei *underfitting* ist der Trainingsfehler hoch, sodass die Trainingsdaten durch das Lernmodell ungenau abgebildet und ihre Eigenschaften unzureichend abdeckt werden (Bonacorso, 2017). Die Ursache für dessen Entstehung liegt in der niedrigen Komplexität des Modells und der ungenügenden Extraktion von Informationen aus dem Trainingsdatensatz. Die entstehende Fehlerart wird auch als *bias* bezeichnet (Gollapudi, 2016; Goodfellow et al., 2016).

Steigt die Modellkomplexität mit Vergrößerung des Datensatzes, ist der Trainingsfehler zwar niedrig, die Vorhersagen auf neuen Daten generieren jedoch zunehmend Fehler (Bonacorso, 2017; Gollapudi, 2016). Dadurch steigt die Varianz des Modells. *Overfitting* bezeichnet hierbei die erweiterte Schere zwischen der Trainings- und Testfehler (Goodfellow et al., 2016). Die Generalisierung ist

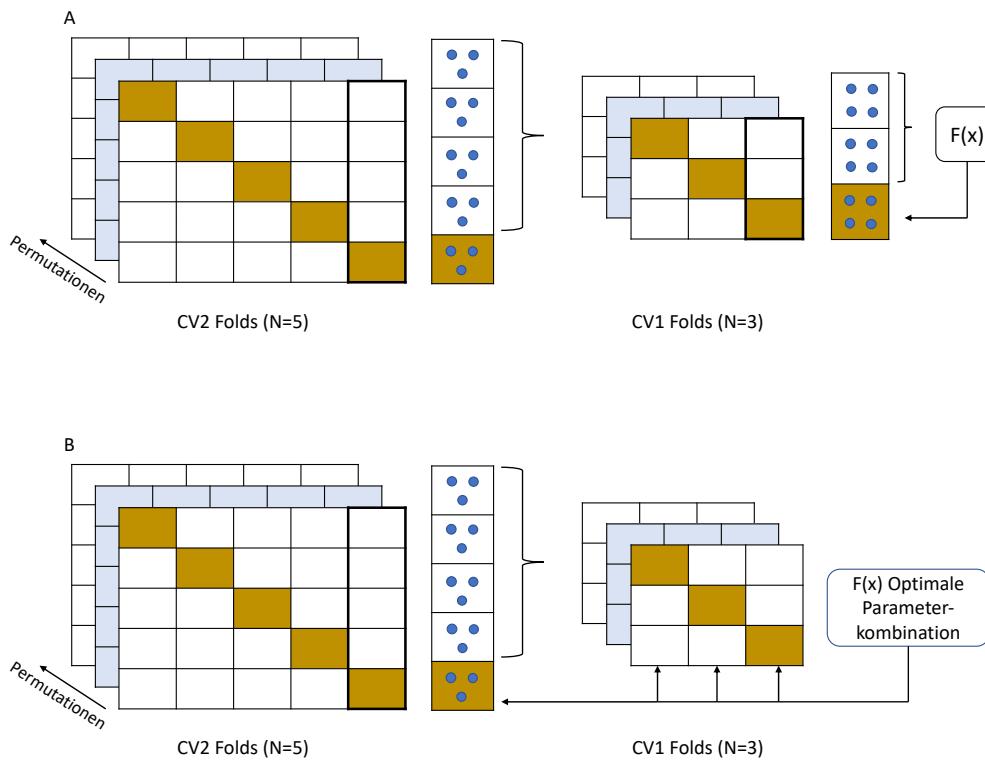
eingeschränkt, da bereits geringe zufällige Variationen der Trainingsdaten fälschlicherweise in die Mustererkennung integriert werden und das Lernmodell diesen gegenüber empfindlich reagiert (Alpaydin, 2020; Gollapudi, 2016).

### 1.2.6 Kreuzvalidierung

Zur Evaluation der Testfehler und Ermöglichung einer besseren Generalisierung und Verhinderung der Überanpassung kann die Kreuzvalidierung hinzugezogen werden, anhand dessen ein trainiertes Modell durch zufällige Aufteilung in Teilmengen an unbeobachtete Testdaten angewendet werden kann. Dies ist insbesondere bei vorliegenden kleineren Datensätzen vorteilhaft, um die Anzahl der Testdaten und somit die Sicherheit des Testfehlers zu erhöhen (Goodfellow et al., 2016). Durch die Implementierung auf verschiedene vorhandene Testdatensätze kann zudem die Robustheit der Modelle miteinander verglichen werden (Garcia & Herrera, 2008).

Das wiederholte, verschachtelte Kreuzvalidierungsschema (Stone, 1974; Varma & Simon, 2006), wie in dieser Arbeit angewendet, stellt eine spezifische Form der Kreuzvalidierung dar. Der grundlegende Vorteil gegenüber anderen Kreuzvalidierungsarten liegt in der Verbesserung der Zuverlässigkeit und Stabilität bezogen auf die Schätzung der Modelleistung. Grundsätzlich werden im inneren Zyklus der Kreuzvalidierung (*Cross-validation folds*, CV1) die einzelnen Modelle und der gesamte Prognose-Arbeitsablauf trainiert und optimiert. Der äußere Kreuzvalidierungszyklus arbeitet durch Ausschluss einer Validierungsstichprobe, um die optimierten Modelle daran zu trainieren, mit dem Ziel des Schutzes vor Überanpassung und der Generalisierbarkeit des Modells auf neue Datensätze (Koutsouleris et al., 2016; Krstajic et al., 2014). Ausgegangen von der Anzahl von 1500 Daten und  $k=3$  inneren und  $k=5$  äußeren *folds* würde die gesamte

Datenmenge zunächst in  $k$  äußere *folds* aufgeteilt und in jeder Runde einer dieser ausgeschlossen werden und als *testdata* fungieren (s. Abb.3).



**Abb.3: Wiederholtes verschachteltes Kreuzvalidierungsschema:** Quelle: In Anlehnung an (Koutsouleris et al., 2016). Zur einfacheren Veranschaulichung wurde eine kleine Datenmenge verwendet. A: CV1: Cross-validation folds 1; CV2: Cross-validation folds 2; Zur einfacheren Veranschaulichung wurde die CV1- und CV2- Foldanzahl auf 5 bzw. 3 bestimmt. Nach Ausschluss eines äußeren *folds* wandern die Daten in das innere Kreuzvalidierungszyklus und werden dort in 3 innere *folds* eingeteilt. Zwei dieser fungieren als *training data*, an denen Modelle mit jeglicher Parameterkombination entwickelt und auf die CV1-Testdaten angewendet werden. B: Für jeden CV-1 *fold* wird anhand der optimalen Parameterkombination die Trainings- und Testdaten erneut trainiert. Schließlich werden sie auf die ausgeschlossene CV2 *testdata* angewendet.

In diesem Beispiel entstünden 5 *folds* à 300 Daten. In das innere Kreuzvalidierungszyklus würden dann 1200 Daten einfließen. Auch hier findet eine Aufteilung in  $k$  innere *folds*, also 3 *folds* mit jeweils 400 Daten. Nach Ausschluss eines inneren *test fold* werden an den verbleibenden Datensätzen (hier 800 Daten) Modelle trainiert und auf das Testset implementiert, um so die optimale Parameterkombination zu ermitteln. Nachdem mit dieser erneut alle CV1 Test und Trainingsdaten trainiert werden, wird sie auf das nicht einbezogene CV2 *fold* angewendet (Varma & Simon, 2006). Eine weitere Möglichkeit ist die Auswahl an Wiederholungen, sogenannten Permutationen. Diese bewirkt, dass die Daten/

Personen vor dem Modellernen erneut durchmischt werden und sich somit die Zusammensetzung zugunsten der Verallgemeinerung der Vorhersage verändert.

### **1.2.7 Anwendungsbereiche des *Machine-Learnings***

Die Anwendung von ML hat sich im Alltag in verschiedenen Lebensbereichen etabliert und ist in vielen Domänen kaum noch wegzudenken. Bildverarbeitung, Vorhersagen, Informationssuche, Text- und die natürliche Sprachverarbeitung stellen unter anderem wichtige Einsatzbereiche dar und machen die Mustererkennung zunutze (Shinde & Shah, 2018). Beispielsweise ist die Handgestenerkennung im Kontext der Objekterkennung (Trigueiros et al., 2012) und die Textklassifikation als Teil der Klassifikationssysteme anzuführen (Mishu & Rafiuddin, 2016).

Auch im medizinischen Bereich expandiert die Anwendung des ML zum Beispiel im Rahmen der Objektdetektion die Identifikation von Brustkrebs (Shubair, 2016) und bei der *Computational Biology* zur Analyse von Proteinfunktionen abhängig von der Zusammensetzung oder die Identifikation funktionell relevanter Regionen (Caragea & Honavar, 2009; Mohri et al., 2018).

Eine zunehmend häufiger verwendete Nutzungsmöglichkeit des ML unter anderem in der Psychiatrie ist die Anwendung an ein *real-world data set* (RWD) (Chen et al., 2023; Franken et al., 2023; Ge et al., 2020; Hao et al., 2023; Mar et al., 2020; Walsh et al., 2018; Yang et al., 2023). Im medizinischen Kontext bezieht sich dies auf die Sammlung von Patientendaten, welche den Gesundheitszustand betreffen und auf Beobachtungen in der realen klinischen Praxis beruhen (Corrigan-Curay et al., 2018). Im Gegensatz zu streng kontrollierten klinischen Studien, deren untersuchte Populationen sich oft als nicht repräsentativ für die Gesellschaft und deren Umsetzung als aufwendig und teuer erweisen, profitiert ein RWD von einer einfachen Zugänglichkeit (Corrigan-Curay et al., 2018; Liu & Panagiotakos, 2022). ML-Ansätze können effektiv mit hochdimensionalen Datensätzen arbeiten und eignen sich daher besonders gut zur Anwendung auf RWD, in der unstrukturierte Daten mit zahlreichen *features* vorliegen (Liu & Panagiotakos, 2022). RWD können jedoch Verzerrungen unterliegen und inkonsistente Daten beinhalten. Sie sind oft unausgeglichen und unvollständig, was zulasten der Vorhersagegenauigkeit, Reproduzierbarkeit, Replizierbarkeit und

Auswertung der Ergebnisse geht (Liu & Panagiotakos, 2022; Muharemi et al., 2019). Im Rahmen der Vorverarbeitung wird beispielsweise durch Ergänzung fehlender Werte versucht, die Defizite der RWD zu minimieren.

### **1.3 Anwendung des ML in der Psychiatrie und bei Patienten mit BPS**

ML findet einen breiten Einsatz im Bereich der Psychiatrie und wird unter anderem zur Vorhersage von Therapieentscheidungen bzw. *therapy response* eingesetzt (Dwyer et al., 2018; Orru et al., 2012). Bislang ist der hauptsächliche Anwendungsbereich im Hinblick auf die Analyse der *therapy response* bei Depressionen (Chekroud et al., 2016; Costafreda et al., 2009; Gong et al., 2011; Iniesta, Malki, et al., 2016; Webb et al., 2020). Weitere Schwerpunkte liegen im Bereich der psychotischen Erkrankungen (Khodayari-Rostamabad et al., 2010; Koutsouleris et al., 2016; Tolmeijer et al., 2018), der Angststörungen (Doehrmann et al., 2013; Hahn et al., 2015; Whitfield-Gabrieli et al., 2016) sowie der PTSD (Stuke et al., 2021; Zhutovsky et al., 2019).

In der Anwendung bei BPS wurde ML bislang vornehmlich für Vorhersagen der Diagnose (Grecucci et al., 2022; Sato et al., 2012; Xu et al., 2014) und zur Abgrenzung zu anderen psychiatrischen Erkrankungen verwendet (Bayes et al., 2022; Bayes et al., 2021; Perez Arribas et al., 2018). Erste Studien zur Identifikation von prädiktiven *features* für die *therapy response* nach DBT fanden bereits statt (Herzog et al., 2020; Jiménez et al., 2022; Schmitgen et al., 2019; Yin et al., 2023). Die *therapy response* wurde u.a. anhand des BPS Schweregrads (Jiménez et al., 2022; Yin et al., 2023), der Emotionsdysregulation, der Nutzung von DBT-Fertigkeiten und funktionaler Beeinträchtigung definiert (Yin et al., 2023). Als Kriterium wurde u.a. auch die Punktzahl aus der Borderline Symptomliste (BSL) -95 (Bohus et al., 2007; Bohus et al., 2001) verwendet (Herzog et al., 2020). Diese stellt ein Selbstbewertungsfragebogen zur Quantifizierung der Symptomschwere bei Borderline Patienten dar und wird zur Evaluierung therapeutischer Veränderungen und neuer Therapieansätze eingesetzt (Bohus et al., 2007; Bohus et al., 2001).

In einem multimodalen Ansatz analysierten Schmitgen und Kollegen (2019) demographische und klinische Daten sowie Magnetresonanztomographie (MRT) -

Daten (Schmitgen et al., 2019). Die Responder wurden durch die Symptomreduktion in der *Zanarini Rating Scale for Borderline Personality Disorder* (ZAN-BPD) (Zanarini, 2003) Gesamtpunktzahl definiert. In den meisten Modellen zeichneten sich unter den klinisch-demographischen Daten die *features* ZAN-BPD-, BSL-95- und Beck-Depressions-Inventar (BDI; s. Kapitel 2.2) Gesamtpunktzahl und unter den strukturellen MRT-Daten das GM-Volumen der linken Amygdala mit der größten prädiktiven Vorhersagekraft aus. Bis auf das Modell mit nur klinisch-demographischen Daten lag die *accuracy* im Bereich von 73 bis 76 %, bei einer Sensitivität und Spezifität über 70 %.

## 1.4 Zielsetzung

Das Ziel der vorliegenden Arbeit ist daher durch einen SVM-Ansatz im Rahmen von ML mithilfe klinischer Routinedaten eine Klassifikation zwischen den Respondern und Non-Respondern bei Borderline Patienten bezüglich der *therapy response* zu generieren und dabei prädiktive *features* herauszufinden. Dabei werden klinisch-demographische und strukturelle MRT-Daten verwendet. Die ermittelten *features* sollen in der praktischen Anwendung Aufschluss darüber geben, wie hoch die Wahrscheinlichkeit ist, dass der Patient von der Therapie profitiert. Durch die Nutzung eines RWD soll eine realitätsnahe Vorstellung von der Korrelation der *features* und der *therapy response* gewonnen werden. Zudem soll durch die Erstellung einer *condensed version* der Prozess für Therapieentscheidungen beschleunigt werden.

Auf Basis der durchgeführten Literaturrecherche wird vermutet, dass sowohl signifikante klinische als auch strukturelle *features* existieren. In Bezug auf strukturelle MRT-Daten wird die Region um die Amygdala als prädiktiv erwartet. Weiterhin wird angenommen, dass insbesondere klinische *features*, welche die Symptomschwere und die komorbide Depression messen, eine hohe Vorhersagekraft besitzen.

Ausgehend von der von Schmitgen und Kollegen (2019) durchgeführten Studie ist zu erwarten, dass die *accuracy* des Modells mit den strukturellen Daten am höchsten sein wird (Schmitgen et al., 2019). Dann folgt das Modell mit der Kombination beider Modalitäten und schließlich das mit nur klinischen Daten.

## **2 Material und Methoden**

### **2.1 Stichprobenbeschreibung**

Untersucht wurden 157 Patienten im Alter von 18 bis 60 Jahren ( $M=29,5$  Jahre  $\pm 9,4$ ,  $SD$ ), welche im Zeitraum von 2016 bis 2020 im Landschaftsverband Rheinland (LVR) -Klinikum Düsseldorf mit der Diagnose einer BPS (nach ICD-10: F60.31) (World Health Organization, 2004) stationär behandelt wurden. Die durchschnittliche Dauer des Aufenthalts betrug  $M=48$  Tage  $\pm 23$ ,  $SD$ . Bei 57 Personen lag BPS als Hauptdiagnose vor. Patienten ohne BPS als Hauptdiagnose wurden hauptsächlich aufgrund einer komorbidens depressiven Symptomatik aufgenommen. Am häufigsten handelte es sich hierbei um eine schwere depressive Episode ohne psychotische Symptome (F32.2, N=37), gefolgt von der rezidivierenden depressiven Störung, gegenwärtig schwere Episode ohne psychotische Symptome (F33.2, N=33) Mehr als die Hälfte der Patienten (N=91) wurden auf einer Spezialstation für BPS (s. Kapitel 2.2) behandelt. Dort erfolgte die Aufnahme neben der Hauptdiagnose BPS (N=46) insbesondere aufgrund der komorbidens Diagnose F32.2 (N=21).

Der Großteil der Untersuchten waren per Einweisung durch den Hausarzt oder Psychiater vorstellig (N=123), wobei 22 notfallmäßig und 10 nach dem Psychisch-Kranken-Gesetz (PsychKG) in die stationäre Behandlung aufgenommen wurden. Eine reguläre Entlassung ließ sich bei 134 Patienten beobachten. 21 von ihnen verließen den Krankenhausaufenthalt gegen ärztlichen Rat. 43 der Patienten gingen einer Erwerbstätigkeit nach.

Neben einer medikamentösen Therapie nahmen 148 Personen an der Kognitiven Verhaltenstherapie teil. 18 Patienten wurden ausschließlich in Einzeltherapie behandelt, die übrigen in einer Kombination von Gruppen- und Einzelgesprächen. Vor Therapiebeginn ließen sich Selbstmordgedanken bei 99 von 115 Patienten eruieren.

### **2.2 Datengewinnung**

Die Studie wurde von der Ethikkommission der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf unter der Studiennummer 2021-1603 genehmigt.

Die klinisch-demographischen Daten wurden abhängig von vorher definierten Variablen (für eine vollständige Variablenliste siehe **Tabelle A1**) von der Klinikinformationssystem (KIS) -Datenbank (Nexus Deutschland GmbH, 2023) selektiert. Dokumentationen während des Klinikaufenthalts und Arztbriefe dienten hierbei als Informationsquelle. Es wurden lediglich Patienten mit Haupt- oder Nebendiagnose (Aufnahme aufgrund komorbider Erkrankung) BPS im Zeitraum von 2016 bis 2020 in Betracht gezogen. Der Großteil der Patienten wurde auf der für diese Erkrankung spezialisierten Station des LVR-Klinikums Düsseldorf im Zeitraum von 2018 bis 2020 aufgenommen. Dort wurden die Patienten umfänglich mit DBT behandelt; über Einzel- und Gruppengesprächen hinaus wurden soziale Fertigkeiten gefördert, Bewegungs- und Kreativtherapien angeboten und das Verständnis für die Erkrankung aufgebaut. Bei Nichtteilnahme an den Angeboten drohte ein Ausschluss aus der Therapie. In regelmäßigen Abschnitten erfolgte die Abfrage durch Fragebögen des Beck Depressionsinventar (BDI) (Beck et al., 1996; Hubley, 2021; Jackson-Koku, 2016) und der Kurzversion der Borderline Symptom Liste (BSL-95 / BSL-23) (Bohus et al., 2009).

Der BDI-Fragebogen umfasst 21 Fragen, jeder Frage sind 4 Antwortmöglichkeiten in einer Skala von 0 bis 3 zugeordnet, wobei eine höhere Ziffer die Schwere des depressiven Symptoms kennzeichnet. Die Antwort der Fragen bezieht sich auf die aktuelle Woche. Dabei werden zentrale Elemente der Depressionssymptome abgefragt: Gesundheit, Sexualität, Selbstwertgefühl und emotionale Zustände wie Hoffnung und Traurigkeit. In diesem Zusammenhang werden auch Gedanken und Einstellungen abgefragt, die von Schuldgefühlen bis hin zu Selbstmordgedanken reichen. Einen weiteren Aspekt bilden körperliche Symptome wie Schlaf, Appetit und Gewichtsabnahme (Beck et al., 1996).

Die Kurzversion der BSL-95 in Form der BSL-23 beinhaltet 23 Fragen in einer Skala von 0 bis 4, wobei eine höhere Ziffer eine stärkere Krankheitsausprägung repräsentiert. Dabei beziehen sich die Fragen auf das Befinden während des Verlaufs der letzten Woche. Die wichtigen Symptomkomplexe der BPS sind im Fragebogen aufgeführt wie zwischenmenschliche Beziehungen einschließlich Vertrauen gegenüber anderen und Kritikfähigkeit und die Selbstwahrnehmung wie Selbsthass und Selbstwert. Zudem werden affektive Symptome wie

Angstgefühl, innere Spannung und kognitive Symptome u.a. Konzentrationsfähigkeit, Kontrollverlust und Abwesenheit erfragt. Es sind außerdem Fragen zu Suizidgedanken und akustischen Halluzinationen integriert. Zusätzlich werden durch 11 Fragen das Verhalten eruiert: Hier ist die Sucht von Medikamenten, Drogen, Alkohol und Essen neben dem Sexualverhalten und selbstschädigendem Verhalten einschließlich Wutausbrüchen, Hochrisikoverhalten bis zu Suizidversuchen enthalten (Bohus et al., 2009).

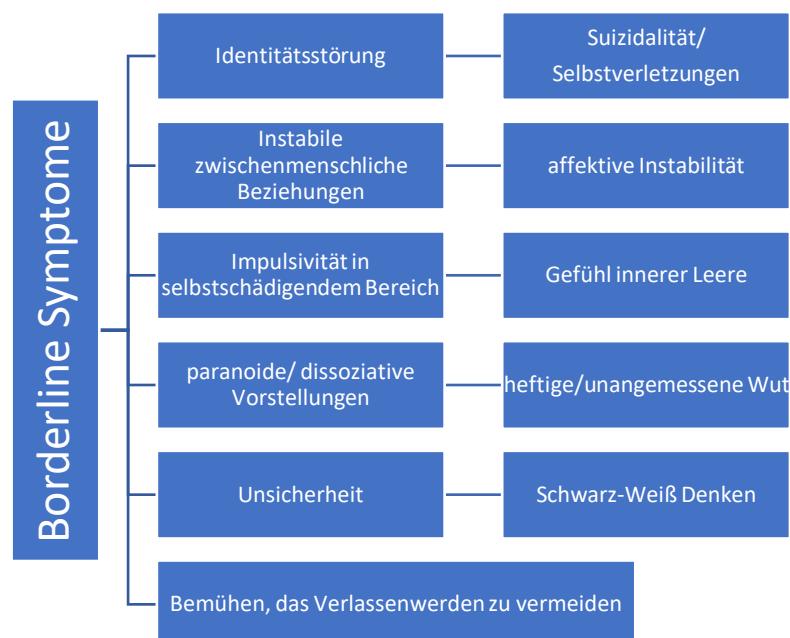
Des Weiteren wurde die Bewertungsskala *Clinical Global Impression* (CGI) (Guy, 1976) erhoben, welche nach Einbezug aller Patienteninformationen die therapeutische Einschätzung über den Schweregrad der psychischen Erkrankung und die Entwicklung im Verlauf wiedergibt. Anhand einer Skala von 1-7 wurde bei der Aufnahme und Entlassung die Schwere der psychischen Symptome bewertet, wobei höhere Werte auf eine geringere Symptombelastung hindeuten (Busner & Targum, 2007).

Die Funktionsfähigkeit wurde anhand der *Global Assessment of Functioning* (GAF) -Skala (Association, 2000) beurteilt. Diese beinhaltet neben psychischen Aspekten berufliche und soziale Faktoren. Die GAF-Skala wurde ebenfalls bei der Aufnahme und Entlassung bestimmt. Auf einer Skala von 1-100 stellen höhere Werte eine bessere Funktionsfähigkeit der Patienten dar (Aas, 2011).

Die Einteilung in die für die ML-basierten Analysen relevanten Klassen Responder und Non-Responder erfolgte anhand zweierlei Kriterien. Zur Bestimmung des Grenzwerts für die Response wurde insbesondere die Erfassung möglichst aller Patienten, die von der Therapie profitieren (hohe Sensitivität), priorisiert. Daher wurde der Grenzwert für die Response auf 30 % Symptomreduktion festgelegt, die sowohl Teilresponder und Responder umfasst (nachfolgend [Teil]Responder). Hierzu wurden die Werte aus den Fragebögen BDI und BSL-23 betrachtet. Als (Teil)Responder galten diejenigen, die einen Rückgang der Werte um mindestens 30 % zeigten (s. Kapitel 4.4). Umgekehrt folgte die Einteilung in die Gruppe Non-Responder bei Nichterreichen dieses Ziels. Falls diese Daten nicht vorhanden waren (N=83), wurde das Kriterium der qualitativen Therapieresponse hinzugezogen. Dazu wurden in der KIS-Datenbank Bewertungen des zuständigen Arztes bezüglich des Therapieansprechens kategorisiert. Eine „Teil- bis Vollremission“, „allmähliche bis rasche Verbesserung“ bzw. eine „Verbesserung in

mehr als drei Bereichen“ dienten als Indizien für ein Therapieansprechen. „Geringe Therapiemotivation“, „Exazerbation“ und „vorzeitige Entlassung“ wurden dagegen als Nicht-Ansprechen bewertet.

Für den Eingang in die Analysen war eine strukturelle Bildgebung des Schädelns anhand einer MRT-Aufnahme in der T1-Wichtung obligat. Das Vorliegen eines pathologischen Befunds resultierend aus der MRT-Untersuchung galt als Ausschlusskriterium, auch hirnorganische Beeinträchtigungen, ermittelt aus der qualitativen Dokumentation der betreuenden Ärzte, kennzeichneten einen Ausschluss der Betroffenen. Auf Grundlage von (Schneider et al., 2019) wurden komorbide körperliche Diagnosen für die BPS ermittelt und mit Diagnosen psychischer Störungen in die Variablenliste aufgenommen (s. **Tabelle A1**). Symptome für die BPS wurden gemäß DSM-IV (301.83) bestimmt (Saß, 2007). Die Variablen „Unsicherheit“ und „Schwarz-Weißdenken“ wurden diesem Symptomkomplex hinzugefügt (s. **Abb.4**).



**Abb. 4: Borderlinesymptome:** Die aufgeführten Elemente der Graphik sind Teil des Symptomkomplexes der Emotional instabilen Persönlichkeitsstörung vom Borderline-Typ. Die Datenbank wurde nach Vorhandensein dieser durchsucht. Die Anzahl der Symptome wurde summiert und stellt das *feature Borderlinesymptome* dar.

Neben der Gesamtpunktzahl im BDI-Fragebogen konnten einzelne Depressions-symptome aus der Datenbank extrahiert werden. Diese sind größtenteils in der

*Hamilton Depression Scale* (Hamilton & Guy, 1976) enthalten und in der **Tabelle 2** dargestellt.

**Tabelle 2: Depressionssymptome**

Depressionssymptome			
Pessimismus	Selbstvorwürfe	Konzentrations-schwierigkeiten	Depressive Stimmung
Gefühllosigkeit	Nervosität/ innere Spannung	Selbstmordgedanken	Suizidversuch
Psychische Angst	Enttäuschung/ Versagen	Verlust von Freude	Selbstablehnung/-hass
Somatische Angst	Schuldgefühle	Anfälle von Weinen	Unruhe
Interessenverlust	Entschlussunfähigkeit	Wertlosigkeit/ Insuffizienzgefühl	Reizbarkeit
Ermüdung	Hilflosigkeit	Einschlafstörungen	Durchschlafstörungen
Hemmung	Körperliche Symptome	Stimmungs-schwankungen	Depersonalisierung
Paranoia	Zwang	Berichtete Traurigkeit	Appetitverlust

**Anmerkung:** Aufgelistete Symptome können während einer depressiven Episode auftreten. Diese wurden bei jeder Person von der Datenbank extrahiert und unter Depressionssymptome in die Variablenliste aufgenommen; Körperliche Symptome beinhalten lediglich Genitalsymptome, Magen-Darmbeschwerden, auffälliges Elektroenzephalogramm, Kopfschmerzen. Für weitere körperliche Assoziationen siehe Anhang Variablenliste

Die Einnahme von Medikamenten wurde sowohl während der Aufnahme als auch bei der Entlassung erfasst und diente jeweils als eigenständige Variable für die Klassifikation. Dabei wurde die Menge der eingenommenen Medikamente berücksichtigt. Die Hauptwirkstoffklassen umfassten Antidepressiva und Antipsychotika, gefolgt von Schilddrüsen- und Vitamin-D-Tabletten (s. **Tabelle 3**). Zusätzlich wurden verschiedene Wirkstoffgruppen dokumentiert, darunter schlafanstoßende Medikamente, Antikonvulsiva (Antiepileptika), (pflanzliche) Beruhigungsmittel, Medikamente gegen Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung, Präparate gegen Narkolepsie sowie Mittel gegen Alkoholabhängigkeit (s. **Tabelle A1**).

**Tabelle 3: Hauptwirkstoffklassen der Medikamenteneinnahme mit den zu gehörigen Wirkstoffen/Medikamenten**

Antipsychotika		Antidepressiva		SD	Vitamin D
atypisch	typisch				
Quetiapin	Pipamperon	Mirtazapin	Fluoxetin	L-Thyroxin	Vigantol
Olanzapin	Prothipendyl	Sertralin	Citalopram	Thiamazol	Dekristol
Aripiprazol	Melperon	Trazodon	Escitalopram	Thyronajod	
Risperidon	Chlorprothixen	Venlafaxin	Duloxetin		
Paliperidon	Lithiumcarbonat		Agomelatin		
Cariprazin		Paroxetin	Trimipramin		
Amisulprid		Milnacipran	Clomipramin		
		Tianeptin	Amitriptylin		
		Doxepin	Fluvoxamin		
		Bupropionhydrochlorid			
		Opipramol			

**Anmerkung:** SD: Schilddrüsenmedikamente; Der Hauptwirkstoff von Vigantol und Dekristol ist Vitamin D3 (Colecalciferol)

## 2.3 Analyse der klinisch-demographischen Daten

(Teil)Responder und Non-Responder wurden hinsichtlich ausgewählter klinischer und demographischer Daten wie Alter, Geschlecht, Punktzahl bei erhobenen Fragebögen, Einnahme von Medikamenten aus den Hauptwirkstoffgruppen und Komorbiditäten miteinander verglichen. Anhand von Chi-Quadrat Tests und Univariaten Analysen wurden *a priori* Unterschiede zwischen den Gruppen untersucht. Die Analysen wurden mit *IBM SPSS Statistics for Mac* der Version 29.0.0.1.0 (171) (<https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-2901>; abgerufen am 10.Dezember 2023) durchgeführt.

Die für die ML-basierten Analysen verwendeten Variablen (s. **Anhang Tabelle A1**) besitzen, ausgenommen derer die zur Bestimmung der (Teil)Responder und

Non-Responder fungierten, einen Anteil von Maximum 26,8 % an fehlenden Daten. Im Schnitt betrug der Anteil  $1,1 \% \pm 4,2 \%$ . Die *Missing at random* (MAR) – und *Missing completely at random* (MCAR) -Mechanismen beschreiben dabei die Art und Weise der Entstehung fehlender Daten. Für die Erläuterung der Mechanismen fehlender Daten siehe Heitjan und Basu (1996) und Rubin (2018) (Heitjan & Basu, 1996; Rubin, 2018). Demnach liefert der SVM im MCAR-Mechanismus bei einem Anteil an fehlenden Daten zwischen 15 und 50 % und im MAR-Mechanismus (Verhältnis des fehlenden Anteils 1:2) zwischen 5-30 % ähnliche Werte. Das MAR-Mechanismus (Verhältnis des fehlenden Anteils 2:1) bringt sogar tendenziell bessere Werte mit Anstieg der fehlenden Daten hervor. Da der Mechanismus des Fehlens in dem vorliegenden RWD nicht eindeutig ist, wurde der Anteil an fehlenden Daten für die einzelnen Variablen sicherheitshalber auf unter 30 % festgelegt.

## 2.4 MRT-Sequenzparameter

Die verwendeten hochauflösende strukturelle Bilder wurden durch ein Siemens Sonata Scanner mit einer magnetischen Feldstärke von 1,5 Tesla des Instituts für Diagnostische und Interventionelle Radiologie des Uniklinikum Düsseldorf (UKD) generiert. Die gemessene strukturelle Sequenz war eine *Magnetization Prepared Rapid Gradient Echo* (MPRAGE) (Mugler III & Brookeman, 1990). Die räumliche Auflösung der Sequenz betrug in der sagittalen und koronaren Ebene jeweils  $0,5 * 0,5$  mm und in der axialen Ebene 1,04 mm. Die T1 gewichteten MRT-Aufnahmen wurden mit einer Echozeit (*Echo time*, TE) von 4,38 ms, einer Repetitionszeit (*Repetition time*, TR) von 1900 ms und einem Flipwinkel (*Flip angle*, FA) von 15° durchgeführt.

## 2.5 Präprozessierung der MRT-Daten

Mittels MATLAB, Version 2022b (The MathWorks, 2022) und der *Software Statistical Parametric Mapping* (SPM12, <https://www.fil.ion.ucl.ac.uk/spm/>; abgerufen am 30.November 2023) (Friston, 2003) wurde zunächst die durch den MR-Tomographen generierten DICOM-Dateien, welche zur standardisierten Speicherung von medizinischen Bildern dienen, in das *Neuroimaging Informatics Technology Initiative* (NifTI) -Format umgewandelt, die in Bezug auf die Analyse der strukturellen Dateien eine einfachere Struktur aufweisen (Flandin & Novak,

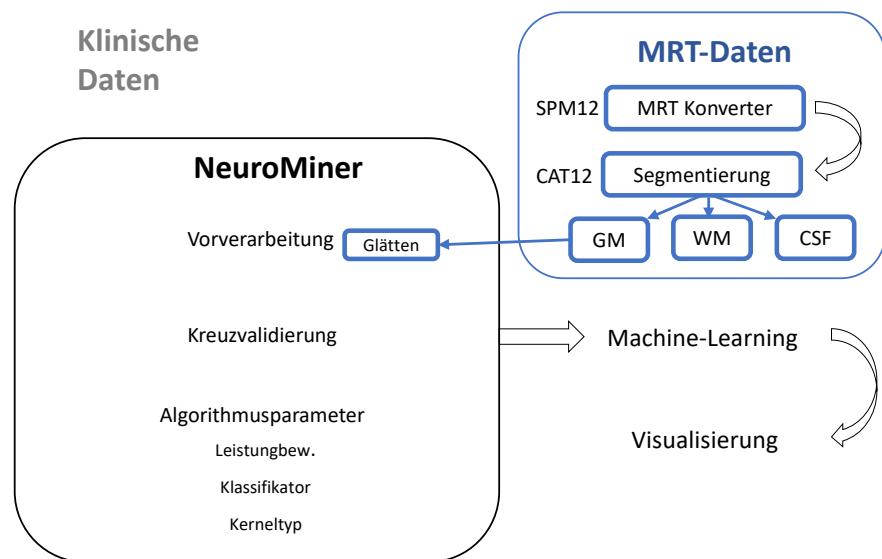
2020). Anschließend wurden die T1 gewichteten NifTI Dateien mit Hilfe der *Computational Anatomy Toolbox* (Cat12, <https://neuro-jena.github.io/cat/index.html#DOWNLOAD>; abgerufen am 30.November 2023) (Gaser et al., 2022) segmentiert. Die Hauptprozesse von Cat12 lassen sich in drei grundlegende Abschnitte unterteilen: Segmentierung, räumliche Registrierung und Analyse der Regionen von Interesse (ROI) auf Voxelbene (s. Cat 12 Manual) (Gaser & Kurth, 2021).

Vor der Segmentierung wird durch das *Skull Stripping* (Kalavathi & Prasath, 2016) das Schädelgewebe entfernt, um das Gehirn isoliert darzustellen. Anschließend werden die Bilder im Rahmen der *Affine Regularisation* durch Transformationen aneinander räumlich angepasst. Hier wurde die Vorlage *European Brains* der *International Consortium for Brain Mapping* (ICBM) -Gruppe (Mazziotta et al., 2001) verwendet.

In der Segmentierung wird ein 3D Bild des Gehirns in die Gewebearten Graue Substanz (GM), Weiße Substanz (WM) und Liquor (Cerebro-spinal fluid, CSF) unterteilt, wobei für die vorliegende Arbeit ausschließlich die GM genutzt wurde. Dieser Schritt stellt die Grundlage für die Untersuchungen kortikaler Veränderungen dar. Die Segmentierung verwendet zunächst *Tissue probability maps*, Karten zur Darstellung der Wahrscheinlichkeiten von Gewebetypen im Gehirn abhängig vom Ort. Die Intensitäten der GM werden durch lokal adaptive Segmentation angepasst.

Die räumliche Normalisierung wurde mittels *Geodesic Shooter* (Ashburner & Friston, 2011) umgesetzt. Sie beinhaltet eine Ausrichtung der Bilder auf einen standardisierten Raum zur besseren Vergleichbarkeit, indem sie auf bereits bestehende Vorlagen des MNI Raums (Mai & Majtanik, 2017) integriert werden. Um die absolute Menge an Gewebe vergleichen zu können, wurde die Normalisierung durch die Modulation ergänzt (Good et al., 2001). Dazu werden die mit der Vorlage ausgerichteten Gewebeklassenelemente mit der Jakobian-Determinante multipliziert.

Es erfolgte eine automatisierte sowie manuelle Überprüfung der Datenqualität. Weitere Präprozessierungsschritte der MRT-Daten und die ML-Analysen der strukturellen sowie klinischen Daten erfolgten mit NeuroMiner der Version 1.1 ([https://github.com/neurominer-git/NeuroMiner\\_1.1](https://github.com/neurominer-git/NeuroMiner_1.1); abgerufen am 1. Dezember 2023). Die grundlegenden MRT-Vorverarbeitungsschritte und wesentliche Komponenten des ML-Programms wurden anhand eines Schemas (s. **Abb.5**) für den Überblick veranschaulicht.



**Abb.5: Analysekomponenten des Machine-Learning Programms NeuroMiner und Vorverarbeitungsschritte der strukturellen MRT-Daten:** MRT: Magnetresonanztomographie; SPM12: *Software Statistical Parametric Mapping*; CAT12: *Computational Anatomy Toolbox*; GM: *Gray matter* („Graue Substanz“); WM: *White matter* („Weiße Substanz“); CSF: *Cerebrospinal fluid* („Liquor“); Leistungsbew: Leistungsbewertung; Blau hinterlegte Formen kennzeichnen die Verarbeitung der MRT-Daten, grau hinterlegte Formen die der klinischen Daten;

Die MRT-Daten wurden zunächst mit SPM12 in das NifTI-Format konvertiert und anschließend mittels CAT12 segmentiert. Für die weitere Analyse wurde ausschließlich die GM verwendet, die in NeuroMiner geglättet wurde. Zusammen mit den klinischen Daten dienten diese MRT-Daten in NeuroMiner zur Klassifikation. Schließlich wurden die Ergebnisse visualisiert.

## 2.6 Machine-Learning an strukturellen Daten

### 2.6.1 Algorithmusparameter

Die Klassifikation fand mit dem maschinellen Lernmodell SVM mit dem speziellen Typ C-Support Vector Classification (C-SVC) unter Nutzung von 11 verschiedenen C Parametern von  $2^{-6}$  bis  $2^4$  statt. Der Zugriff auf SVM erfolgte über LIBSVM

(Chang & Lin, 2007), es wurde ein linearer Kernel verwendet. Mithilfe der Einstellung *shrinking heuristics* erfolgte ein beschleunigter Trainingsprozess.

## 2.6.2 Vorverarbeitung der MRT-Daten

In der chronologischen Reihenfolge fand zunächst das Glätten (*smoothing*) durch eine Gaußsche Kernel mit Halbwertsbreite (*Full-Width-at-Half-Maximum*) von 8 mm statt. Dieser Schritt erhöht das Verhältnis zwischen dem gewünschten Signal und dem störenden Rauschen (Gaser & Kurth, 2021). Außerdem verringert er nach der Bildregistrierung entstehende Artefakte durch Neuberechnung (Mikl et al., 2008). Anschließend wurden die Merkmale des Datensatzes im Bereich von -1 bis 1 skaliert und unendliche Werte durch 0 ersetzt. Im nächsten Schritt fand die Entfernung von Merkmalen mit keiner Varianz, Merkmale mit fehlenden/undefinierten Werten (NaN) und Merkmale mit unendlichen Werten. Mittels der Teilkorrelationsanalyse (Marrelec et al., 2006) wurden die Daten um den Einfluss der Kovarianten Alter und Geschlecht bereinigt. Die Hauptkomponentenanalyse (PCA; s. Kapitel 1.2.1) (Van Der Maaten et al., 2009) kürzte die Dimension des Datensatzes auf 80 % der ursprünglichen Varianz. Zuletzt erfolgte erneut der Schritt der Skalierung.

## 2.6.3 Kreuzvalidierungsschema

NeuroMiner nutzt das wiederholte verschachtelte Kreuzvalidierungsschema (s. Kapitel 1.2.6). In der vorliegenden Arbeit wurden jeweils  $k=5$  für die Anzahl der inneren und äußeren *folds* gewählt. Die Wiederholungsrate betrug für den inneren sowie äußeren Zyklus jeweils drei Permutationen. Außerdem handelt es sich um eine gepoolte Kreuzvalidierung, wo die äußeren und inneren *folds* zufällig und automatisch generiert werden. Das aufgestellte Kreuzvalidierungsschema blieb für alle weiteren Analysen gleich.

## 2.6.4 Leistungsbewertung

Zur Leistungsbewertung für die Analyse der MRT-Datensätze wurde die *balanced accuracy* bestimmt, welche sich als

$$(\text{Sensitivität} + \text{Spezifität})/ 2$$

ausdrückt (Velez et al., 2007). Sensitivität stellt dabei die richtig positive (TP), Spezifität die richtig negative (TN) Rate dar. Grundlage dieser Entscheidung ist

die ungleiche Verteilung der Stichprobengröße (N=31 vs. N=126), welche zu unausgeglichenen Sensitivität und Spezifität führen kann (NeuroMiner Manual Version 1.4, [https://github.com/neurominer-git/Training\\_materials/blob/master/Neurominer\\_Manual.pdf](https://github.com/neurominer-git/Training_materials/blob/master/Neurominer_Manual.pdf); abgerufen am 2. Dezember 2023). In den **Tabellen 7 und 8** im Ergebnisteil sind weitere Leistungskennzahlen wie die *accuracy*, TP/TN, FP/FN, die AUC, positiv prädiktiver Wert, negativ prädiktiver Wert, *prognostic summary index*, *positive likelihood ratio* und *negative likelihood ratio* aufgeführt. Für die Erklärung der Leistungskennzahlen siehe in das Kapitel 1.2.4.

Es wurde die *sign-based consistency* verwendet, um die Relevanz der Daten, hier Voxel der für die Klassifikation signifikanten Hirnregionen, hinsichtlich der Vorhersagekraft des Modells zu bewerten (s. Kapitel 1.2.4). Zur Quantifizierung der Sicherheit/Robustheit einer Entscheidung werden die sogenannten *decision scores* für jeden Patienten berechnet (als Mittelwert der für diesen Patienten getroffenen Entscheidungen). Je höher dieser ausfällt, desto eindeutiger die Entscheidung, welcher Gruppe dieser Patient zuzuordnen ist. *Decision scores* um den Wert 0 bedeuten hingegen, dass die Patienten mit ähnlichen Wahrscheinlichkeiten der Gruppe (Teil)Responder und Non-Responder zugeordnet werden können.

Zur Bewertung der klinischen Datenanalyse wurden dieselben Leistungskennzahlen wie für die strukturellen Datensätze verwendet.

## 2.6.5 Visualisierung der Ergebnisse

Im Zuge der Vorverarbeitung werden die Daten in einen neuen Raum transformiert, in welcher die Vorhersagegenauigkeit und auch die *feature weights* (s. Kapitel 1.2.1) bestimmt werden. Um diese Ergebnisse richtig auswerten zu können, werden sie im Rahmen der Visualisierung in den ursprünglichen Raum zurückprojiziert. Matrixdaten, wie die klinisch-demographischen Daten, werden dabei anhand von Diagrammen und Graphen veranschaulicht. Neuroanatomische Bilder, wie strukturelle Daten aus dem Hirn, werden in den ursprünglichen Voxelraum zurückgeführt. Signifikante Variablen der klinischen Daten wurden mittels Balkendiagrammen dargestellt und relevante Hirnareale farblich hervorgehoben. Letztere wurden unter Zuhilfenahme von Standardatlanten identifiziert.

## **2.7 Machine-Learning an klinisch-demographischen Daten**

### **2.7.1 Algorithmusparameter**

*Machine-Learning* erfolgte anhand des SVM, welcher mit dem nonlinearen sigmoiden Kerneltyp arbeitete.

### **2.7.2 Vorverarbeitung**

Als erster und abschließender Schritt erfolgte eine Skalierung der Merkmale im Bereich von -1 und 1. Um den Effekt der als Kovariablen bestimmten Faktoren Alter und Geschlecht zu minimieren, wurde die Partialkorrelationsanalyse angewandt. Die Hauptkomponentenanalyse ermöglichte eine Reduktion des Orginaldatenvolumens auf 80 %. Fehlende Werte wurden durch die Hybrid-Technik bei Nutzung des *Statistics and Machine Learning Toolbox* von MATLAB ergänzt. HAMMING (Subasi et al., 2011) wurde für die ordinalen Werte in dem Schema des *dummy codings* ausgewählt und fehlende nominale bzw. kontinuierliche Werte durch EUCLIDEAN (Santos et al., 2020) vervollständigt.

## **2.8 Stacking**

Die Fusion der strukturellen und klinisch-demographischen Daten zu einem Meta-Modell zwecks der Verbesserung der Vorhersagegenauigkeit wurde durch die Verwendung der *Stacking*-Methode realisiert (s. Kapitel 1.2.3). Hier wurde zur Klassifikation der nonlineare Gaußsche Kernel des SVM verwendet. Präprozessierung fand in drei Schritten statt: Neben initialer und abschließender Skalierung der Modelle von -1 bis 1 und Ersetzung unendlicher Werte durch 0, wurden Elemente ohne Informationen ausgeschnitten. Diese beinhalteten Merkmale mit unendlichen Werten, Merkmale mit fehlenden/undefinierten Werten (NaN) und welche mit keiner Varianz. Beim *Stacking* wurden die Analysen lediglich mit der ersten Permutation des wiederholten verschachtelten Kreuzvalidierungsschemas generiert.

## **2.9 Bagging**

Es wurde die *late fusion* der *Bagging*-Methode eingesetzt, um die *decision scores* beider Datenmodalitäten zusammenzuführen (s. Kapitel 1.2.3). Die Eingabe der Einstellungen für die Algorithmusparameter, die Vorverarbeitung und das

Kreuzvalidierungsschema mussten separat für jede Modalität ausgeführt werden und wurden wie oben beschrieben unverändert übertragen.

## **2.10 Condensed version des klinisch-demographischen Klassifikators**

Anhand der signifikanten Variablen, ermittelt durch die *sign-based consistency*, wurde ein neuer kompakterer Datensatz aufgestellt (s. Kapitel 1.2.4). Es erfolgte die Klassifikation auf Grundlage dieser klinischen Daten. Anschließend wurde die *Stacking-* und *Bagging*-Methode auf dieses Modell und das für die strukturellen Daten angewendet. Aufgrund des reduzierten Datensatzes war keine Ersetzung der fehlenden Daten während der Vorverarbeitung notwendig. Bei dem *stacking* der klinischen Daten in Form der *condensed version* und der strukturellen Daten wurde auf den sigmoiden Kerneltyp gewechselt. Für alle anderen Einstellungen siehe obige Ausführungen.

### 3 Ergebnisse

#### 3.1 Statistische Unterschiede der Zielgruppen

Die durchgeföhrten *a priori* Analysen anhand des Chi-Quadrat Testes und der Univariaten Analyse zeigten signifikante Unterschiede ( $p<0.05$ ) der (Teil)Responder gegenüber den Non-Respondern in einigen demographischen und klinischen Bereichen: Non-Responder waren im Durchschnitt 3,6 Jahre älter als (Teil)Responder ( $p=0.015$ ). Außerdem zeigten die Untersuchungen einen Unterschied in dem Alter bei erster psychischer Behandlung, wobei Non-Responder signifikant älter bei Erstbehandlung waren als (Teil)Responder ( $p=0.029$ ) (s. Tabelle 4).

**Tabelle 4: Demographische und klinische Daten (Teil)Responder vs. Non-Responder**

	Resp [N=126]	Non-Resp [N=31]	Statistik	Effektstärke ( $\eta^2$ )
Geschlecht [m/w]	33/93	8/23	$X^2=0.0$ , $p=.965$	
<b>Alter</b>	<b>28.6(8.8)</b>	<b>33.2(11.2)</b>	<b><math>F_{(1,156)}=6.0</math>, <math>p=.015</math></b>	<b>.04</b>
Anzahl der Aufenthalte	4.0(4.9)	3.1(3.9)	$F_{(1,156)}=0.8$ , $p=.370$	.01
<b>Alter bei erster psychiatrischer Behandlung [<math>N_{\text{resp}}=123</math>, <math>N_{\text{non\_resp}}=30</math>]</b>	<b>24.1(8.7)</b>	<b>28.3(11.7)</b>	<b><math>F_{(1,156)}=4.9</math>, <math>P=.029</math></b>	<b>.03</b>
BDI_Aufnahme [ $N_{\text{resp}}=63$ , $N_{\text{non\_resp}}=26$ ]	36.5 (9.0)	33.8 (11.9)	$F_{(1,156)}=1.4$ , $p=.244$	.02
BSL-23 Aufnahme [ $N_{\text{resp}}=51$ , $N_{\text{non\_resp}}=13$ ]	59.1(26.3)	54.5 (35.1)	$F_{(1,156)}=0.3$ , $p=.603$	.00
CGI Aufnahme [ $N_{\text{resp}}=122$ , $N_{\text{non\_resp}}=31$ ]	2.79(0.66)	2.94(0.68)	$F_{(1,156)}=1.2$ , $p=.267$	.01
CGI Entlassung [ $N_{\text{resp}}=102$ , $N_{\text{non\_resp}}=29$ ]	4.28(0.71)	4.41(0.63)	$F_{(1,156)}=0.8$ , $p=.376$	.01
GAF Aufnahme [ $N_{\text{resp}}=122$ , $N_{\text{non\_resp}}=31$ ]	38.13(11.78)	39.52(10.53)	$F_{(1,156)}=0.4$ , $p=.552$	.00
GAF Entlassung [ $N_{\text{resp}}=102$ , $N_{\text{non\_resp}}=29$ ]	60.11(9.68)	58.38(11.38)	$F_{(1,156)}=0.7$ , $p=.416$	.01

Borderlinesymptome [N <sub>resp</sub> =114, N <sub>non_resp</sub> =25]	3.80(2.66)	4.12(2.79)	F <sub>(1,156)</sub> =0.3, p=.587	.00
Compliance (ja/nein) [N <sub>resp</sub> =125, N <sub>non_resp</sub> =31]	92/33	19/12	X <sup>2</sup> =1.8, p=.176	
Krankheitseinsicht (ja/nein) [N <sub>resp</sub> =110, N <sub>non_resp</sub> =24]	92/18	18/6	X <sup>2</sup> =1.0, p=.317	
Schule (kein Abschluss/ Hauptschulabschluss/ mittlere Reife/Abitur) [N <sub>resp</sub> =93, N <sub>non_resp</sub> =24]	7/12/28/46	1/4/9/10	X <sup>2</sup> =1.1, p=.781	
Beruf (keine Ausbildung/ Ausbildung/weiterführende Ausbildung/ Fachhochschulausbildung) [N <sub>resp</sub> =107, N <sub>non_resp</sub> =25]	43/43/2/19	6/16/1/2	X <sup>2</sup> =5.6 p=.132	

**Anmerkung:** m: männlich; w: weiblich; Resp: (Teil)Responder; Non-Resp: Non-Responder; BDI: Beck-Depressions-Inventar, BSL: Borderline Symptom Liste; CGI: Clinical Global Impression; GAF: Global Assessment of Functioning;  $\eta^2$ : Partial Eta Square  $\eta^2=0.01$  zeigt einen kleinen Effekt,  $\eta^2= 0.06$  einen mittleren Effekt und  $\eta^2=0.14$  einen großen Effekt; Signifikanzniveau  $p<0.05$ ; X<sup>2</sup>:Chi-Quadrat Test, F<sub>(1,156)</sub>: F-Verteilung der univariaten Analyse mit einem Freiheitsgrad im Zähler (Zwischen-Gruppen-Varianz) und 156 Freiheitsgrade im Nenner (innerhalb-Gruppen-Varianz) bei kontinuierlichen Variablen. Grenzwerte: BSL-23 ab einer Punktzahl von 43 hoher Schweregrad für Borderline Persönlichkeitsstörung; BDI ab 29 Punkten schwere Depression, CGI bei einer Punktzahl von 2 oder niedriger schwere psychische Störung; GAF bei 50 Punkten oder weniger ernsthafte Symptome oder Beeinträchtigung. Aufgrund einer hohen Anzahl an fehlenden Daten bei der Entlassung für BSL-23 (N=105) und BDI (N=84) sind sie in dieser Tabelle nicht angeführt, dienten aber als wichtiges Kriterium bei der Bewertung des Therapieansprechens.

Zusätzlich unterschieden sich die Non-Responder von den (Teil)Respondern durch eine höhere Antidepressiva Einnahme bei Aufnahme in die Behandlung ( $p=0.016$ ) (s. **Tabelle 5**).

**Tabelle 5: Medikamenteneinnahme (Teil)Responder vs. Non-Responder**

	Resp [N=126]	Non-Resp [N=31]	Statistik
Aufnahme_Neuroleptikum (ja/nein)	29/97	11/20	$\chi^2=2.0$ , $p=.154$
Entlassung_Neuroleptikum (ja/nein)	33/93	12/19	$\chi^2=1.9$ , $p=.167$
<b>Aufnahme_Antidepressivum (ja/nein)</b>	<b>51/75</b>	<b>20/11</b>	<b><math>\chi^2=5.8</math>, <math>p=.016</math></b>
Entlassung_Antidepressivum (ja/nein)	82/44	25/6	$\chi^2=2.8$ , $p=.096$

**Anmerkung:** Resp: (Teil)Responder; Non-Resp: Non-Responder; Signifikanzniveau  $p<0.05$ ;  $\chi^2$ :Chi-Quadrat Test; weiterführende Ausbildung meint z.B. den Meisterabschluss

Bezüglich des Auftretens von Komorbiditäten wiesen (Teil)Responder und Non-Responder keine Gruppenunterschiede auf (s. **Tabelle 6**).

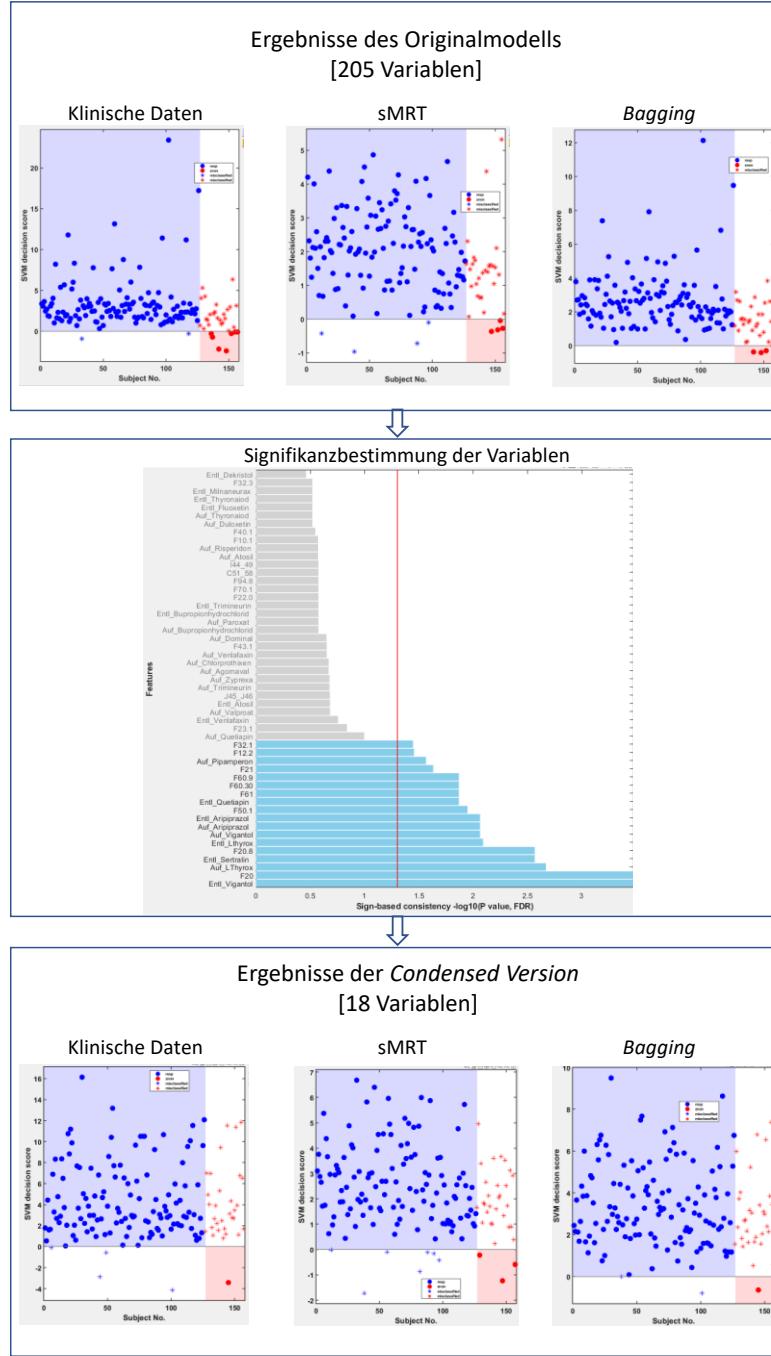
**Tabelle 6: Komorbiditäten (Teil)Responder vs. Non-Responder**

	Resp [N126]	Non-Resp [N31]	Statistik	Effektstärke ( $\eta^2$ )
Suizidalität während des Aufenthalts (Keine Suizidalität/ Suizidalität oder Suizidversuche/ Suizidalität und Suizidversuche) [N <sub>resp</sub> = 120, N <sub>non_resp</sub> = 31]	92/18/10	26/5/0	X <sup>2</sup> =2.8, p=.251	
Durchschnittliche Anzahl Suizidversuche [N <sub>resp</sub> = 105, N <sub>non_resp</sub> = 23]	0.81(1.13)	0.65(0.94)	F <sub>(1,156)</sub> =0.4 p=.534	.00
Depressive Stimmung [(keine/ leichte bis mittelgradige de- pressive Episode/ schwere depressive Episode) [N <sub>resp</sub> =117, N <sub>non_resp</sub> =28]	2/34/81	0/7/21	X <sup>2</sup> =0.7, p=.699	
Durchschnittliche Anzahl Depressions-symptome [Anzahl]	8.52(3.11)	8.10(2.52)	F <sub>(1,156)</sub> =0.5, p=.488	.00
Sucht (keine/ Nikotin oder Drogen/ Nikotin und Drogen) [N <sub>resp</sub> = 94, N <sub>non_resp</sub> = 24]	16/55/23	5/13/6	X <sup>2</sup> =0.2 p=.895	

**Anmerkung:** Resp: (Teil)Responder; Non-Resp: Non-Responder;  $\eta^2$ : Partial Eta Square  
 $\eta^2=0.01$  zeigt einen kleinen Effekt,  $\eta^2= 0.06$  einen mittleren Effekt und  $\eta^2=0.14$  einen großen Effekt; Signifikanzniveau  $p<0.05$ ; X<sup>2</sup>:Chi-Quadrat Test; F<sub>(1,156)</sub>: F-Verteilung der univariaten Analyse mit einem Freiheitsgrad im Zähler (Zwischen-Gruppen-Varianz) und 156 Freiheitsgrade im Nenner (innerhalb-Gruppen-Varianz) bei kontinuierlichen Variablen; ICD 10 Codes: leichte bis mittelgradige depressive Episode=F32.0-F32.1, schwere depressive Episode=F32.2-F32.3

## 3.2 Ergebnisse des *Machine-Learnings*

Die Ergebnisse der Analysen für das Originalmodell und der *condensed version* werden getrennt präsentiert. Die Streudiagramme der *decision scores* für das unimodale klinisch-demographische Originalmodell mit 205 Variablen (s. Anhang **Tabelle A1**), das unimodale Modell der MRT-Daten, sowie die bimodalen Modelle mit Datenfusionierung *Stacking* und *Bagging* wird zunächst dargestellt. Anschließend wird die nach Ermittlung signifikanter Variablen aufgestellte *condensed version* der klinisch-demographischen Daten mit 18 Variablen und die dadurch entstandene Verteilung der *decision scores* gezeigt. Für die zusammenfassende Veranschaulichung der *decision scores* und der signifikanten Variablen siehe in die schematische Darstellung (s. **Abb.6**). Alle auf der Schemadarstellung veranschaulichten Abbildungen werden nachfolgend näher erläutert.



**Abb. 6: Schematische Darstellung der Ergebnisse des Originalmodells sowie des Prozesses zur Ausarbeitung signifikanter Variablen für die *condensed version*, einschließlich dessen Ergebnisse:** Streudiagramme: Subject no.: Teilnehmernummer; resp: (Teil)Responder; znon: Non-Responder; misclassified: fehlklassifiziert; Auf\_: Aufnahme; Entl\_: Entlassung; Die Abkürzungen stellen ICD-10 Codes dar.

Diese Graphik visualisiert die chronologische Abfolge der Arbeitsschritte: Nach Beobachtung der Verteilung einzelner Entscheidungspunkte für die jeweiligen Datenmodelle (hier exemplarisch klinische Daten, MRT-Daten und das *Bagging*-Modell) wurden unter Zuhilfenahme der Visualisierung signifikante Variablen abgebildet. Diese wurde anschließend benutzt, um eine kompaktere Version zu entwickeln, dessen Verteilung unten auf der Graphik repräsentiert ist.

### 3.2.1 Ergebnisse des Originalmodells

Die Leistungskennzahlen der Klassifikation in die Zielgruppen bei den MRT-Daten zur strukturellen Erfassung, den klinischen Daten, der Fusion beider Modalitäten im Rahmen des *baggings* und dem durch *stacking* entwickelten Meta-Modell sind in der **Tabelle 7** aufgeführt. Unter den oben genannten Modellen und Modalitäten stellte sich die Vorhersage der klinisch-demographischen Daten als statistisch signifikant dar ( $p=0.013$ ). Die Genauigkeit (*accuracy*) der Schätzung betrug in allen Modellen über 80 %. Dabei schnitt sie bei den klinischen Daten am besten ab (83.4 %), gefolgt von der *Bagging*- (82.2 %) und *Stacking*-Technik (80.9 %).

Demnach erbrachte die Fusion der Datenmodalitäten keine genaueren Vorhersagen. Alle Modelle sind durch eine hohe Sensitivität bei relativ schwacher Spezifität gekennzeichnet. Dadurch sank folglich die *balanced accuracy*. Klinische Daten zeigten auch in diesem Bereich die beste Prädiktion (60.5 %).

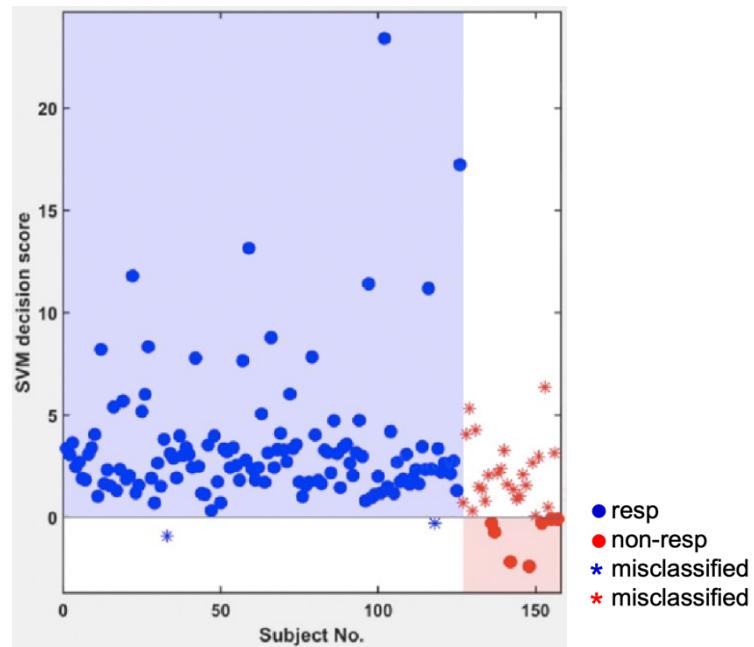
**Tabelle 7: Ergebnisse der Klassifikation zwischen (Teil)Respondern und Non-Respondern in dem Originalmodell**

Modell	TP/TN	FP/FN	ACC [%]	BAC [%]	Sens [%]	Spez [%]	AUC	PPV [%]	NPV [%]	PLR	NLR	PSI	Kappa	p <sub>FDR</sub>
sMRT	122/4	27/4	80.3	54.9	96.8	12.9	.70	81.9	50.0	1.1	0.2	31.9	1.1 p=.258	
klin	<b>124/7</b>	<b>24/2</b>	<b>83.4</b>	<b>60.5</b>	<b>98.4</b>	<b>22.6</b>	<b>.73</b>	<b>83.8</b>	<b>77.8</b>	<b>1.3</b>	<b>0.1</b>	<b>61.6</b>	<b>2.5</b> <b>p=.013</b>	
bagg	126/3	28/0	82.2	54.8	100.0	9.7	.77	81.8	100.0	1.1	0.0	81.8	1.0 p=.303	
stack	125/2	29/1	80.9	52.8	99.2	6.5	.62	81.2	66.7	1.1	0.1	47.8	0.6 p=.547	

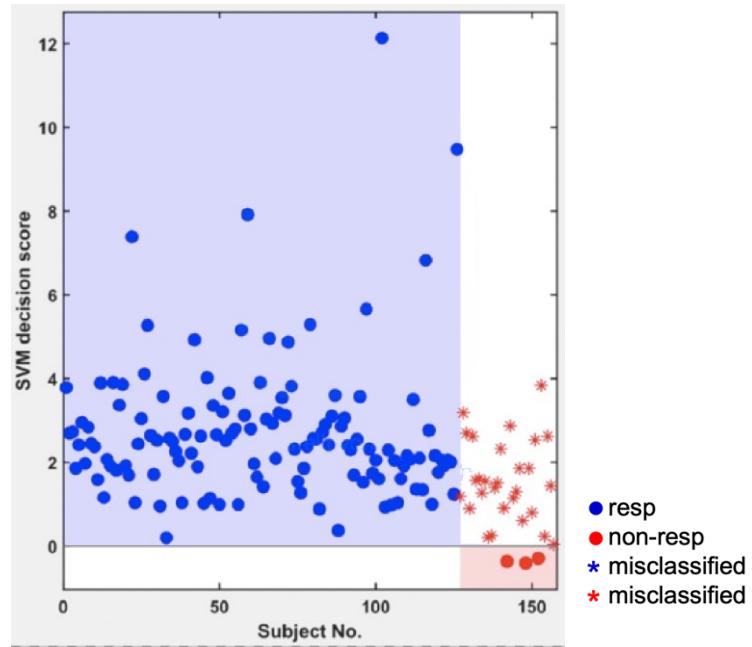
**Anmerkung:** sMRT: strukturelle MRT-Daten; klin: klinische Daten; bagg: *Bagging*; stack: *Stacking*; TP: *True positive rate*; TN: *True negative rate*; FP: *False positive rate*; FN: *False negative rate*; ACC: *Accuracy*; BAC: *Balanced accuracy*; Sens: *Sensitivität*; Spez: *Spezifität*; AUC: *Area under the curve*; PPV: *Positiv prädiktiver Wert*; NPV: *Negativ prädiktiver Wert*; PLR: *Positive likelihood ratio*; NPL: *Negative Likelihood ratio*, PSI: *Prognostic summary index*; pFDR: um die FDR (*false discovery rate*) korrigierter p Wert, Signifikanz ab  $p<0.05$

Die klinisch-demographischen Daten zeichneten sich durch das beste Verhältnis an Richtig Positiven ( $N=124$ ) gegenüber Falsch Positiven ( $N=24$ ) und damit auch durch den besten positiv prädiktiven Wert aus. Außerdem zeigten sie die niedrigste Anzahl an falschen Zuordnungen auf ( $N=26$ ) (s. **Abb.7**). Die durch *late fusion* generierten Daten wiesen keine falsch negative Klassifizierung auf und

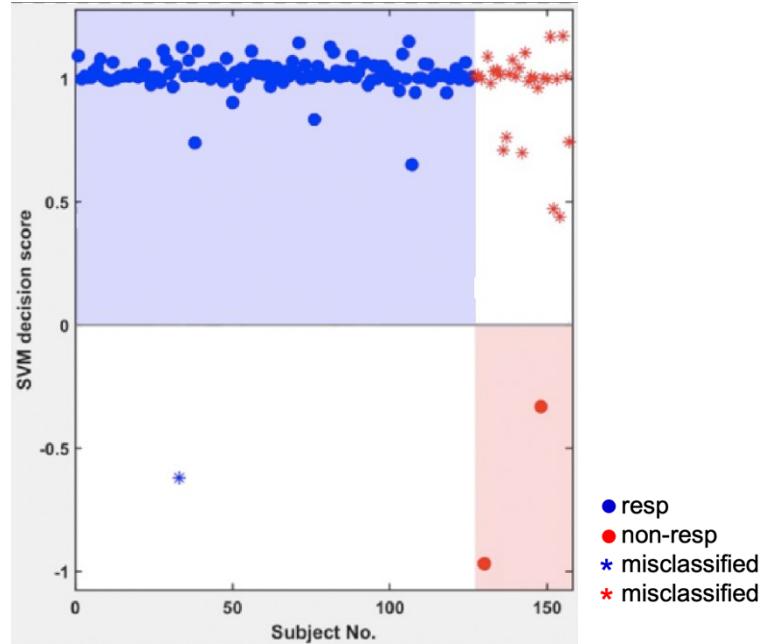
erhielten dadurch den maximalen negativ prädiktiven Wert (s. **Abb.8**). **Abb. 9** zeigt ebenfalls eine niedrige falsch negative Rate ( $N=1$ ) gegenüber einer höheren Anzahl an falsch positiv klassifizierten Personen in dem *Stacking*-Modell. Dieses Verhältnis war bei den Neuroimaging-Aufnahmen am höchsten (4:27) (s. **Abb.10**). Die Abbildungen 7 bis 10 zeigen die Verteilung der Patienten abhängig von den jeweiligen *decision scores* der SVM zur Klassifizierung in die (Teil)Responder und Non-Responder Gruppe.



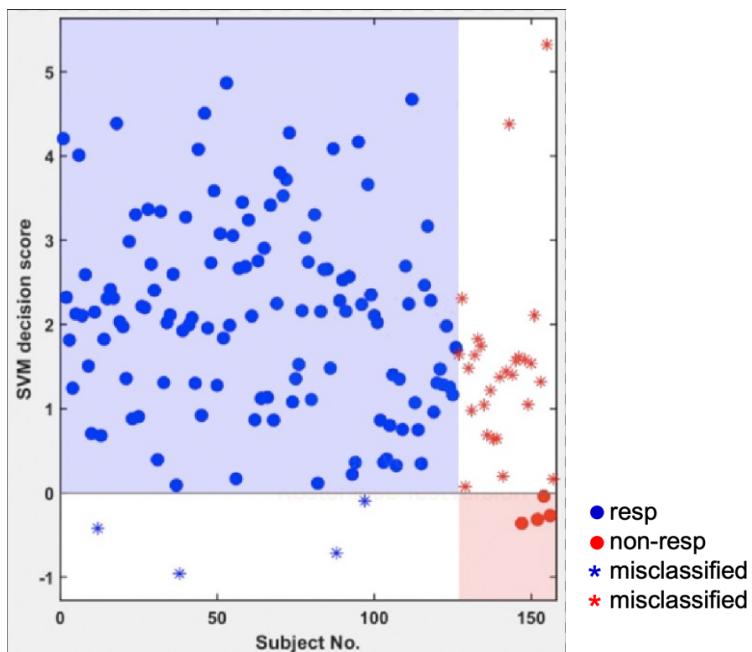
**Abb.7:** Verteilung der Patienten anhand der *decision scores* der SVM in den klinischen Daten: Subject no.: Teilnehmernummer; resp: (Teil)Responder; znon: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 23.4 wobei der minimale bei -2.4 liegt.



**Abb.8: Verteilung der Patienten anhand der *decision score* der SVM in dem *Bagging-Modell*:** *Subject no.:* Teilnehmernummer; *resp:* (Teil)Responder; *znon:* Non-Responder; *misclassified:* fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 12.1 wobei der minimale bei -0.4 liegt.

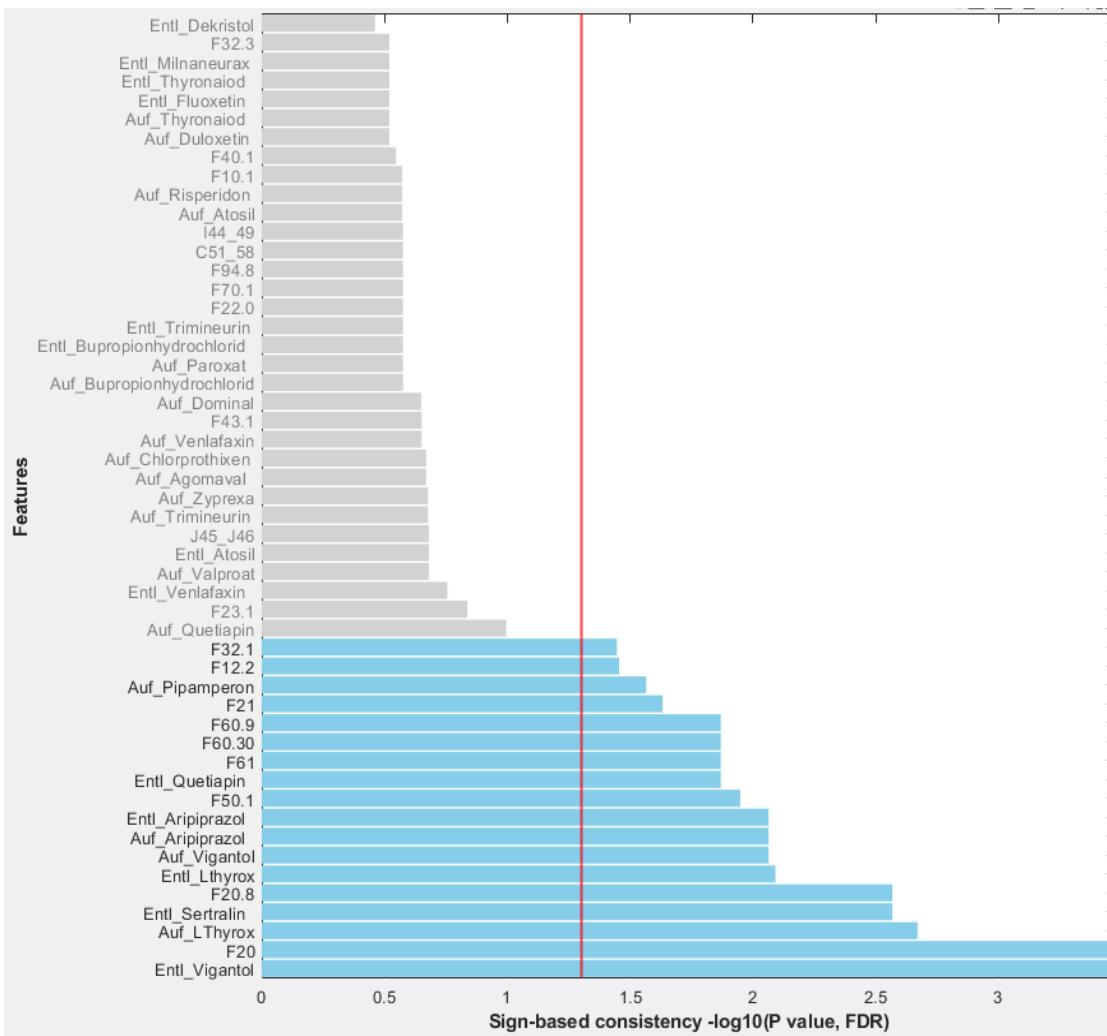


**Abb.9: Verteilung der Patienten anhand der *decision scores* der SVM im *Stacking-Modell*:** *Subject no.:* Teilnehmernummer; *resp:* (Teil)Responder; *znon:* Non-Responder; *misclassified:* fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 1.2 wobei der minimale bei -1.0 liegt.



**Abb.10: Verteilung der Patienten anhand der *decision scores* der SVM in den strukturellen MRT-Daten:** Subject no.: Teilnehmernummer; resp: (Teil)Responder; non: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 4.9 wobei der minimale bei -1.0 liegt.

Mittels Ergebnisvisualisierung wurden die *feature weights* bestimmt und anhand der *sign-based consistency* (s. Kapitel 1.2.4) auf statistische Signifikanz untersucht (s. **Abb.11**). Demnach beeinflussen die folgenden 18 *features* die Klassifikation maßgeblich: F32.1 (mittelgradige depressive Episode), F12.2 (Psychische und Verhaltensstörungen durch Cannabinoide Abhängigkeitssyndrom), Pipamperon-Einnahme während der Aufnahme, F21 (Schizotype Störung), F60.9 (Persönlichkeitsstörung, nicht näher bezeichnet), F60.30 (Emotional instabile Persönlichkeitsstörung-impulsiver Typ), F61 (Kombinierte und andere Persönlichkeitsstörungen), Quetiapin-Einnahme während der Entlassung, F50.1 (Atypische Anorexia nervosa), Aripiprazol-Einnahme während der Entlassung, Aripiprazol-Einnahme während der Aufnahme, Vigantol-Einnahme während der Aufnahme, L-Thyroxin-Einnahme während der Entlassung, F20.8 (Sonstige Schizophrenie), Sertralin-Einnahme während der Entlassung, L-Thyroxin-Einnahme während der Aufnahme, F20 (Schizophrenie) und letztlich Vigantol-Einnahme während der Entlassung.



**Abb.11:** Übersicht der sich als signifikant herausgestellten *features* bei Klassifikation in die Gruppen (Teil)Responder und Non-Responder abhängig von der *sign-based consistency*: Features: Variablen; Sign-based consistency -log10 (P value, FDR): Messung der Konsistenz von Vorzeichen unter Verwendung des an den *false discovery rate* angepassten p-Wertes, der durch -log10 skaliert wird, somit wird der dargestellte Wert umso größer je kleiner der p-Wert ist; Auf\_: Aufnahme; Entl\_: Entlassung; Die Abkürzungen stellen ICD-10 Codes dar.

### 3.2.2 Ergebnisse der *condensed version*

Die Leistungskennzahlen der Klassifikation in die Zielgruppen bei den MRT-Daten zur strukturellen Erfassung, den klinischen Daten, der Fusion beider Modalitäten im Rahmen des *baggings* und dem durch *stacking* entwickelten Meta-Modell sind in der **Tabelle 8** aufgeführt. Keines dieser Modelle ermöglichte eine statistisch signifikante Einordnung der Patienten in die (Teil)Responder und Non-Responder-Gruppen ( $p>0.05$ ). Zudem zeigten sie ähnliche *Accuracy*-Werte: Die Fusionstechnik *bagging* schnitt hier mit 79,6 % am besten ab, dem folgte das Metamodell *learning* gestützt auf das *stacking* (79,0 %). Das Gütekriterium

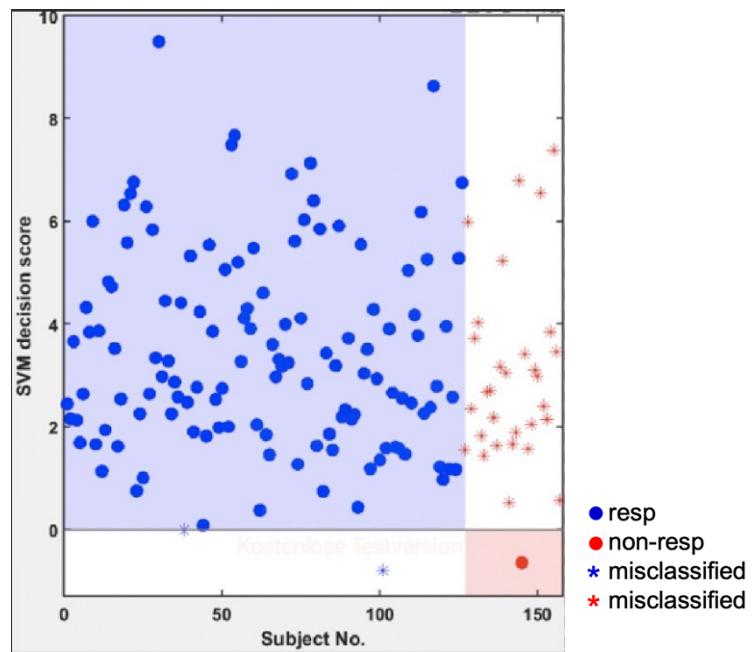
*balanced accuracy*, welche sowohl die Sensitivität als auch die Spezifität beinhaltet, war mit 52,1 % bei den bildgebenden MRT-Daten am höchsten. Die Analyse der *condensed version* zeigte somit auf, dass die Fusion der strukturellen und klinischen Datenmodalitäten in Bezug auf die Genauigkeit der Vorhersage einen Nutzen darstellte.

**Tabelle 8: Ergebnisse der Klassifikation zwischen (Teil)Respondern und Non-Respondern in der condensed version**

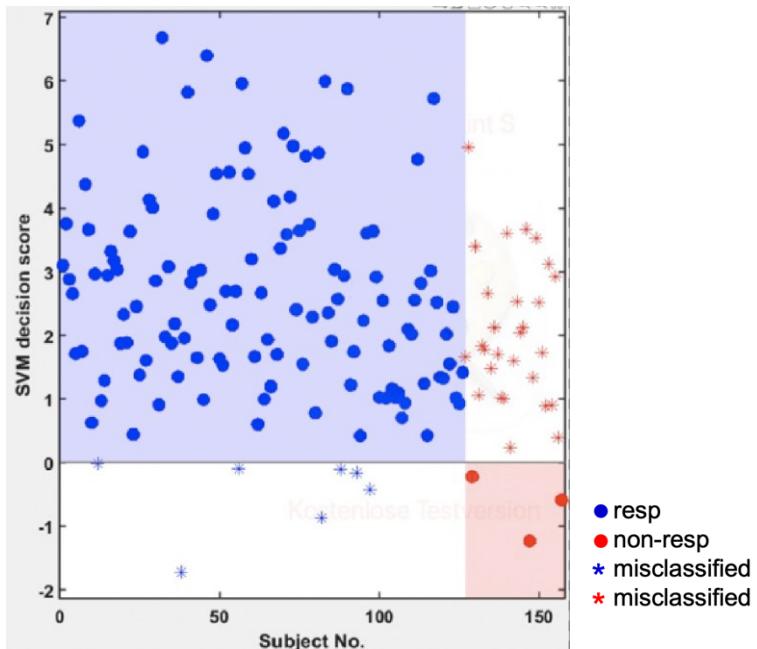
Modell	TP/TN	FP/FN	ACC [%]	BAC [%]	Sens [%]	Spez [%]	AUC	PPV [%]	NPV [%]	PLR	NLR	PSI	Kappa pFDR
sMRT	119/3	28/7	77.7	52.1	94.4	9.7	0.62	81.0	30.0	1.0	0.6	11.0	0.5 p=.6206
<b>klin</b>	122/1	30/4	78.3	50.0	96.8	3.2	0.51	80.3	20.0	1.0	1.0	0.3	0.01 p=.9955
bagg	124/1	30/2	79,6	50.8	98.4	3.2	0.57	80.5	33.3	1.0	0.5	13.9	0.2 P=.8615
stack	123/1	30/3	79.0	50.4	97.6	3.2	0.50	80.4	25.0	1.0	0.7	5.4	0.1 P=.9270

**Anmerkung:** sMRT: strukturelle MRT-Daten; klin: klinische Daten; bagg: *Bagging*; stack: *Stacking*; TP: *True positive rate*; TN: *True negative rate*; FP: *False positive rate*; FN: *False negative rate*; ACC: *Accuracy*; BAC: *Balanced accuracy*; Sens: *Sensitivität*; Spez: *Spezifität*; AUC: *Area under the curve*; PPV: *Positiv prädiktiver Wert*; NPV: *Negativ prädiktiver Wert*; PLR: *Positive likelihood ratio*; NPL: *Negative likelihood ratio*, PSI: *Prognostic summary index*; pFDR: um die FDR (*false discovery rate*) korrigierter p Wert, Signifikanz ab  $p<0.05$

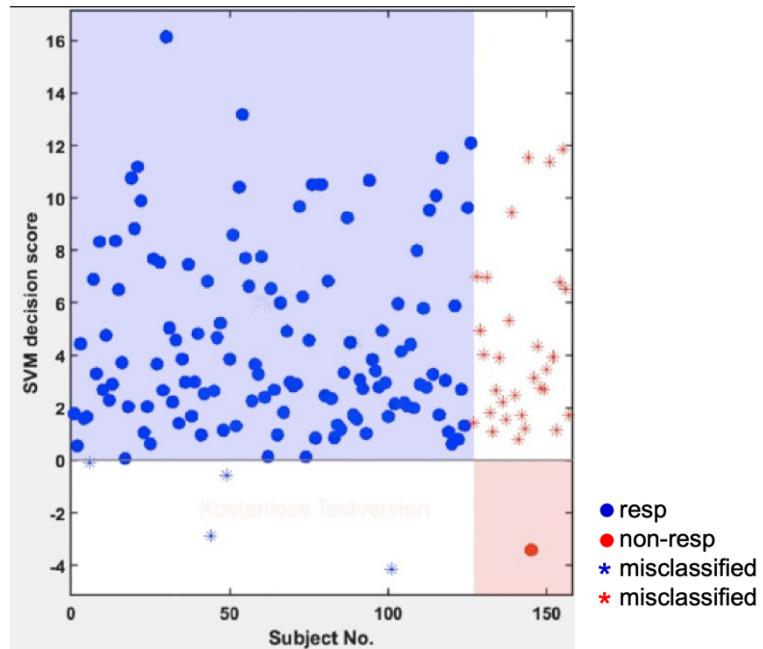
Unter allen trainierten Modellen wurde bei der *Bagging*-Strategie am seltensten Patienten in die falsche Gruppe eingeordnet (N=32) (s. **Abb.12**). Darüber hinaus wies es die höchste richtig positiven Rate und folglich auch die beste Sensitivität (98,4 %) auf. Die relativ niedrige Anzahl an falsch negativ eingeordneten Patienten (N=2) verbesserte den negativ prädiktiven Wert (33.3 %). Die strukturellen MRT-Daten ermöglichten die meiste richtige Zuordnung der Non-Responder, wodurch die Spezifität am höchsten war (s. **Abb.13**). Außerdem zeigten sie einen hohen positiv prädiktiven Wert mit 81 % durch unter anderem der niedrigen Anzahl der Falsch Positiven. Für die Verteilung bei dem klinischen Modell und der *Stacking*-Strategie siehe **Abb. 14** bzw. **Abb. 15**.



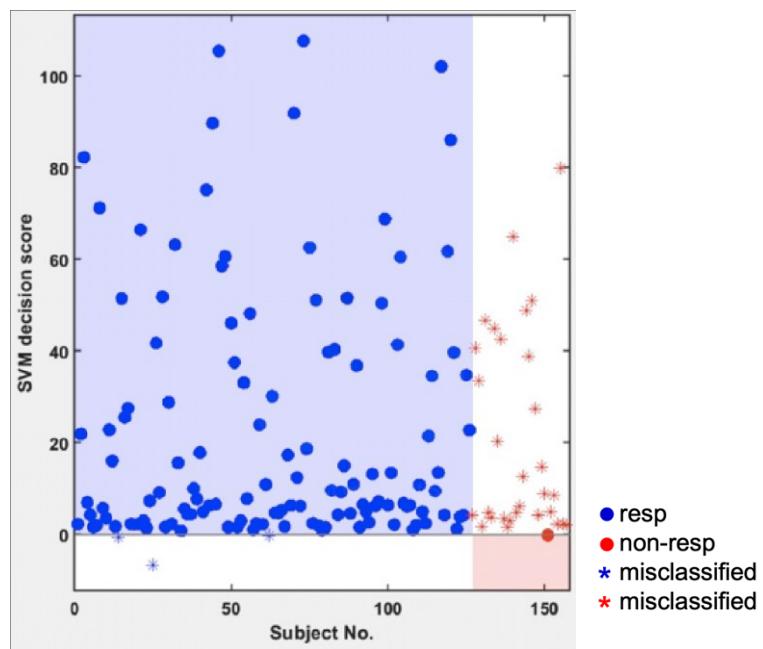
**Abb.12: Verteilung der Patienten anhand der *decision score* der SVM in dem Bagging-Modell für die condensed version:** Subject no.: Teilnehmernummer; resp: (Teil)Responder; znon: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 9.5 wobei der minimale bei -0.8 liegt.



**Abb.13: Verteilung der Patienten anhand der *decision scores* der SVM in den strukturellen MRT-Daten für die condensed version:** Subject no.: Teilnehmernummer; resp: (Teil)Responder; znon: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 6.7 wobei der minimale bei -1.7 liegt.



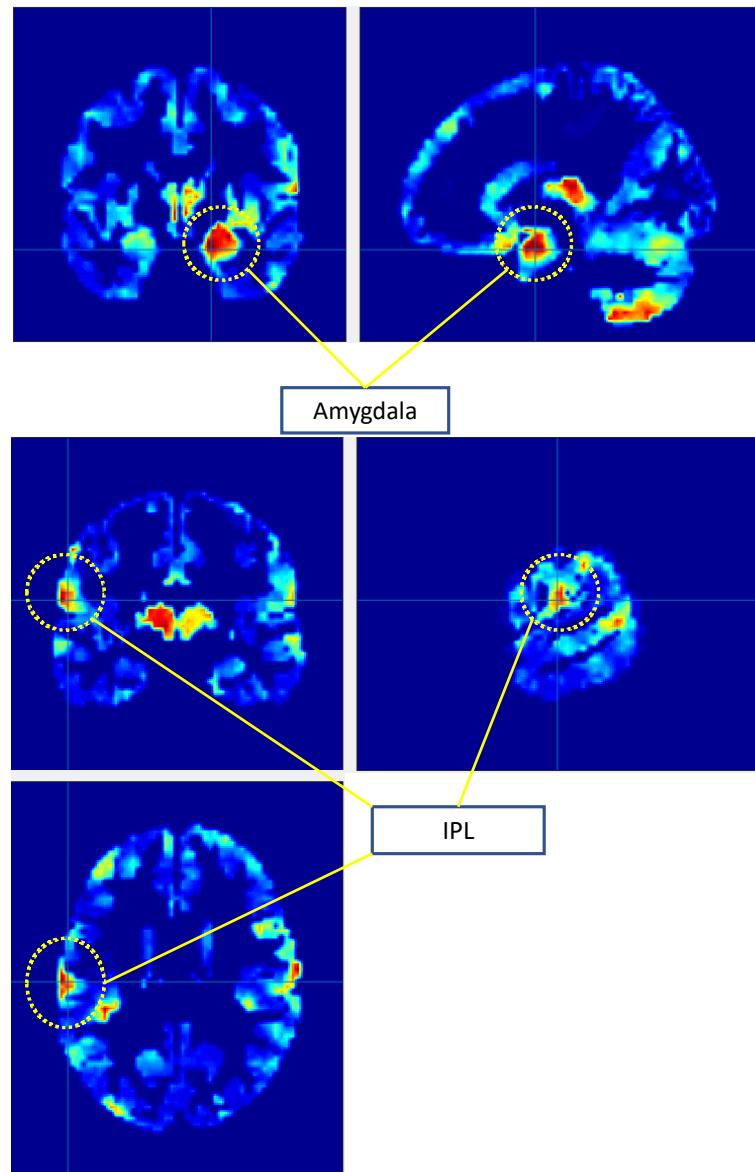
**Abb.14:** Verteilung der Patienten anhand der *decision score* der SVM in den klinischen Daten für die condensed version: Subject no.: Teilnehmernummer; resp: (Teil)Responder; non: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 16.1 wobei der minimale bei -4.1 liegt.

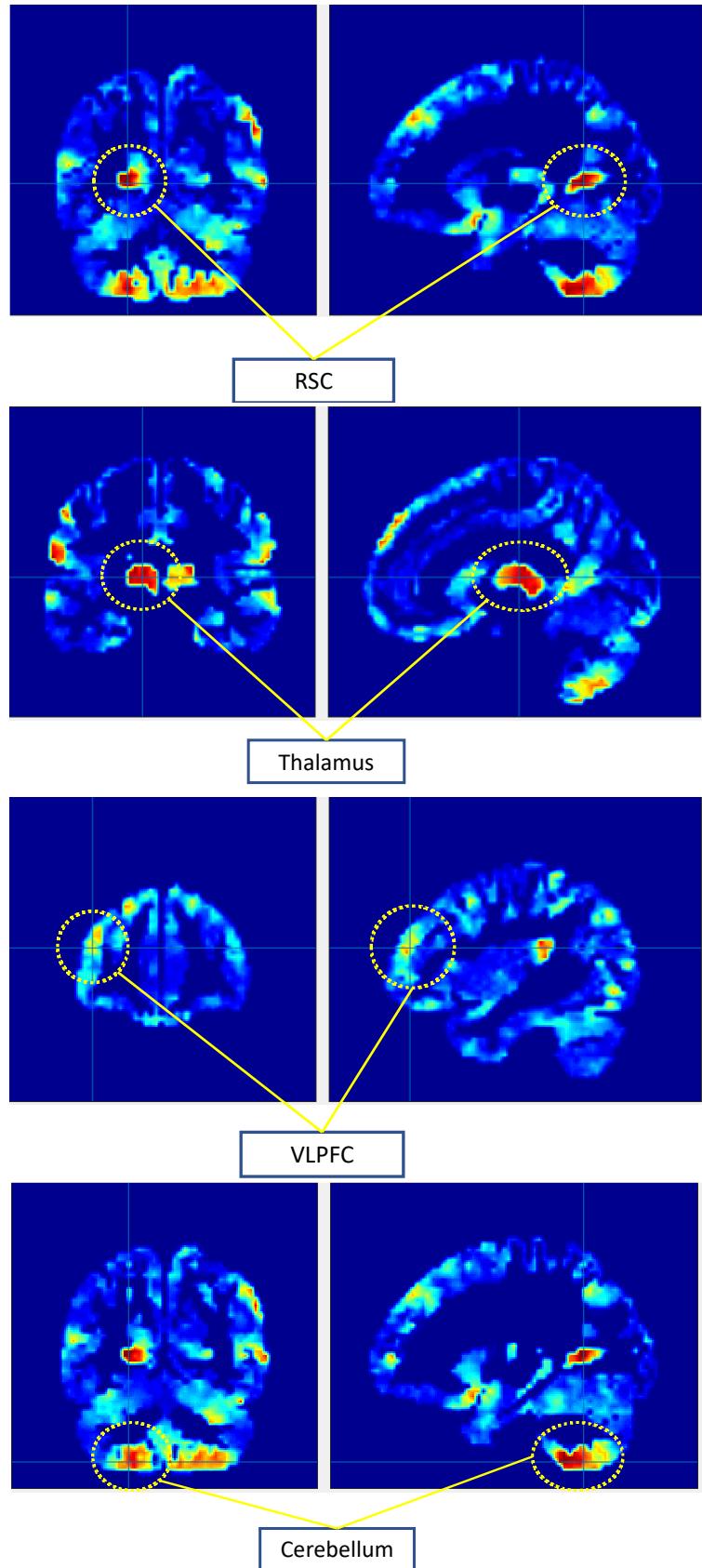


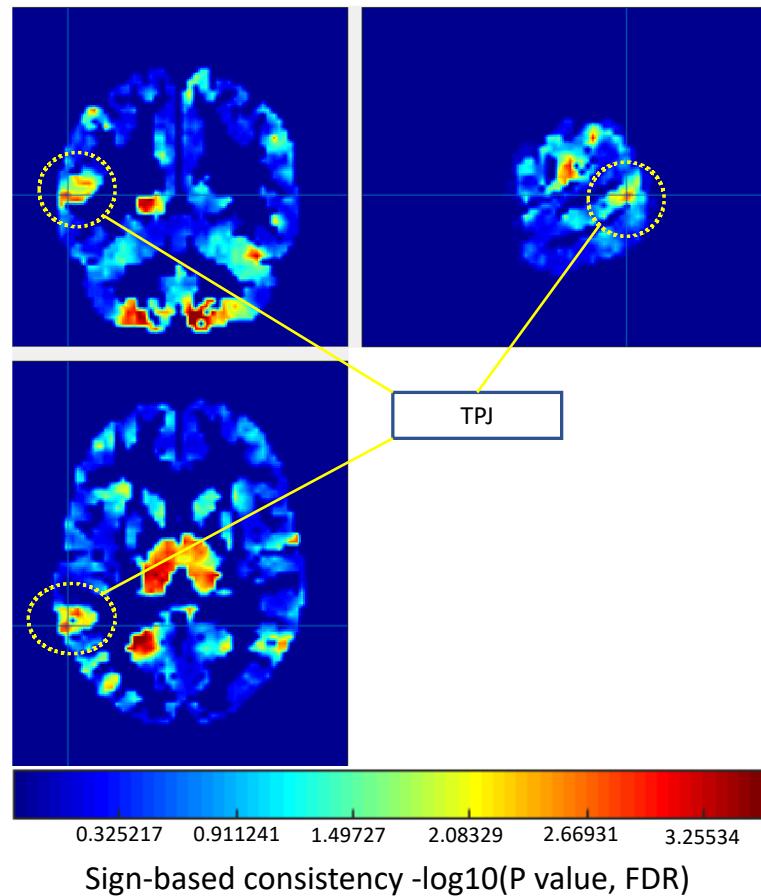
**Abb.15:** Verteilung der Patienten anhand der *decision score* der SVM im Stacking-Modell für die condensed version: Subject no.: Teilnehmernummer; resp: (Teil)Responder; non: Non-Responder; misclassified: fehlklassifiziert; aus den jeweiligen Zuordnungen ergibt sich die Rate an Richtig und Falsch Positiven bzw. Richtig und Falsch Negativen. Durch diese lassen sich weitere Leistungskennzahlen beschreiben. Der maximale *decision score* beträgt 107.6 wobei der minimale bei -6.6 liegt.

### 3.2.3 Signifikante Hirnareale

Bildgebende strukturelle Hirndaten wurden in Form von Voxeln in drei Hirnebenen visualisiert, wobei die statistisch relevanten Hirnabschnitte farbig hervorgehoben wurden (s. Abb.16). Nach Identifikation dieser Regionen unter Zuhilfenahme von Standardatlanten ergaben sich folgenden prädiktive Areale: Amygdala, Lobus parietalis inferior, retrosplenialer Cortex, Thalamus, temporo-parietaler Übergang, ventrolateraler präfrontaler Cortex und Cerebellum.







**Abb.16: Signifikante Hirnregionen in Bezug auf die Klassifikationen in die Zielgruppen (Teil)Responder und Non-Responder auf Grundlage der *sign-based consistency*:** *Sign-based consistency* -log10(P value, FDR): Messung der Konsistenz von Vorzeichen unter Verwendung des an den *False Discovery Rate* angepassten p-Wertes, der durch -log10 skaliert wird, somit wird der dargestellte Wert umso größer je kleiner der p-Wert ist; IPL: Lobus parietalis inferior; RSC: Retrosplenialer Cortex; TPJ: Temporoparietal junction („temporoparietale Übergangszone“), VLPFC: Ventrolateraler präfrontaler Cortex; Die Farbleiste unterhalb der Abbildung kennzeichnet die aufsteigende Signifikanz von der Farbe blau zur Farbe rot; Hirnareale in Standardschichten gezeigt: koronar (links), sagittal (rechts), axial (unten links)

## 4 Diskussion

In der vorliegenden Arbeit wurde ein multimodaler ML-Ansatz zur Vorhersage der *therapy response* bei Patienten mit BPS aufgestellt. Das Ziel war die Untersuchung von Prädiktoren, anhand derer potenziell die DBT-Therapie individuell optimiert werden kann, sowie im Rahmen der patientenzentrierten Forschung Therapieentscheidungen auf die individuellen Merkmale des Patienten angepasst werden können (Lambert, 2013). Dazu wurden 157 Patienten im Hinblick auf die Wirksamkeit der DBT- und pharmakologischen Therapie analysiert und Prädiktoren anhand klinischer, demographischer und MRT-Daten bestimmt. Unter den durchgeführten Modellen mit dem SVM-Lernalgorithmus erwies sich das klinisch-demographische Originalmodell als signifikant. Anhand der klinischen und demographischen RWD konnte die *therapy response* mit einer *accuracy* von 83.4 %, einer Sensitivität von 98.4 % und einer AUC von 0.73 vorhergesagt werden. Die Analyse der *feature weights* ergab, dass v.a. die Medikation und komorbide Erkrankungen einen signifikanten Einfluss auf die Prädiktion nahmen. Alle weiteren Klassifikatoren (basierend auf MRT-Daten, der Kombination der klinischen und strukturellen Daten anhand *bagging* und *stacking* sowie der *condensed version*) haben sich als nicht signifikant herausgestellt. Nach derzeitigem Wissensstand ist es die erste Arbeit, die basierend auf multimodalen RWD mit einem SVM die *therapy response* von Patienten mit BPS analysiert.

### 4.1 Bestimmung der Effizienz des Models

Die gefundenen Gütekriterien sind vergleichbar mit anderen Studienergebnissen: Insbesondere *accuracy*, Sensitivität, AUC, positiv prädiktiver-, negativ prädiktiver Wert und *prognostic summary index* schneiden verglichen mit Studien zur Vorhersage von *therapy response* in anderen psychiatrischen Erkrankungen unter Verwendung von klinisch-demographischen Daten gut ab (Hilbert et al., 2021; Koutsouleris et al., 2016; Lambert, 2013; Symons et al., 2020). Somit scheint auch die Dokumentation von Medikation und komorbiden Erkrankungen in RWDs eine verlässliche Datenquelle für Vorhersagen des Therapieerfolgs zu sein.

In der Referenzstudie von Schmitgen und Kollegen (2019) wurden ebenfalls klinische und hirnstrukturelle Daten verwendet und u.a. durch deren Kombination mehrere Modelle aufgestellt, welche den Lernalgorithmus *Random forest*

nutzten (Schmitgen et al., 2019). Durch *median permutation importance scores* im Zuge der Kreuzvalidierung wurden die prädiktivsten *features* für finale Modelle ausgewählt, vergleichbar mit der *condensed version* in dieser Arbeit. Auch das finale klinisch-demographische Modell von Schmitgen und Kollegen (2019) war nicht signifikant, und die *accuracy* war niedriger als im klinischen Modell der *condensed version* dieser Arbeit (68 % zu 78.3 %). Einfluss darauf könnte insbesondere die niedrige Anzahl an *features* (9 *features*) genommen haben (Schmitgen et al., 2019). Außerdem wurden weniger Patienten in die Analyse integriert (31 im Vergleich zu 157 in der vorliegenden Arbeit), womit potenziell wichtige prädiktive *features* durch die kleine Anzahl an *items* nicht abgedeckt wurden.

Besonders auffällig am klinischen Vorhersagemodell der vorliegenden Arbeit und allen anderen Modellen ist die geringe Spezifität, welche die richtige Identifikation von Non-Respondern darstellt. Dies kann darauf hinweisen, dass die ermittelten *features* nicht relevant für den (Miss)Erfolg in Bezug auf die Therapie waren. Einflussnehmende Faktoren wurden jedoch durch das Modell nicht erkannt. Bei dem Vergleich mit Schmitgen und Kollegen (2019) sieht man eine ausgewogene Verteilung von Sensitivität und Spezifität (Schmitgen et al., 2019). Dies könnte möglicherweise auch auf die ausgewogene Verteilung von Respondern (N=16) und Non-Respondern (N=16) zurückgeführt werden. Die unausgeglichene Verteilung der *items* zu (Teil)Respondern (N=126) und Non-Respondern (N=31) in dieser Arbeit bestärkt diese Hypothese. Die prädiktive *accuracy* kann zuungunsten der Minderheit sinken, weshalb die Anwendung einer *resampling* Technik eventuell nützlich sein könnte (Chakravarthy et al., 2019). Es muss darauf geachtet werden, dass diese Technik aufgrund der Gefahr des *overfitting* und der möglichen Hervorhebung außergewöhnlicher Merkmale der Stichprobe die Generalisierung einschränken kann (Yu, 2019).

Ferner muss diskutiert werden, welche Auswirkungen das Nichterkennen eines Non-Responders auf die Effizienz des Modells und auf den individuellen Patienten haben könnte. Wenn die Spezifität niedrig ist, steigt die Rate an falsch positiv eingeschätzten Patienten, d.h. sie würden hier eine Überversorgung erhalten. Dies würde sowohl unnötige Kosten generieren als auch den Patienten den Zugang zu alternativen Behandlungen verlangsamen oder verwehren. Dennoch ist dieser Zustand dem umgekehrten Szenario mit einer niedrigen Sensitivität

vorzuziehen, da durch die DBT und medikamentöse Therapie den Respondern gemäß aktueller Forschung effektiv geholfen werden kann und deren Nichterkennen für die individuellen Patienten bedeutet, dass sie die effektivste Behandlung nicht wahrnehmen können und möglicherweise durch wiederholte Arztbesuche die Gesundheitsversorgung erheblich finanziell belasten. Außerdem würde sich die gesundheitliche Belastung eventuell durch die Frustration über die fehlende Wirksamkeit der medizinischen Versorgung verschärfen. Ideal wäre das Szenario mit einer ausgewogenen Sensitivität und Spezifität, was eine hohe *balanced accuracy* zur Folge hätte.

## 4.2 Relevante features

In dem klinisch-demographischen Originalmodell waren folgende komorbide Erkrankungen signifikante *features* für die *therapy response*: F32.1 (mittelgradige depressive Episode), F12.2 (Psychische und Verhaltensstörungen durch Cannabinoide Abhängigkeitssyndrom), F21 (Schizotype Störung), F60.9 (Persönlichkeitsstörung, nicht näher bezeichnet), F61 (Kombinierte und andere Persönlichkeitsstörungen), F50.1 (Atypische Anorexia nervosa), F20.8 (Sonstige Schizophrenie), F20 (Schizophrenie) und eine vorangegangene Diagnostik mit F60.30 (Emotional instabile Persönlichkeitsstörung-impulsiver Typ).

In vorangegangen Studien ist bereits die Komorbidität der BPS mit Depressionen beschrieben (Zanarini et al., 1998). Bei den Patienten mit komorbider Depression könnte ein Fokus auf Aspekte der kognitiven Verhaltenstherapie im Rahmen der DBT zielführend sein. In vergleichbaren Studien wurde bereits die BDI-Gesamtpunktzahl (Schmitgen et al., 2019), als auch einzelne Bestandteile des BDI (Jiménez et al., 2022) als prädiktive *features* festgestellt, was den Einfluss der komorbidien Depression betont. In der Studie von Herzog und Kollegen (2020) wurde die komorbide rezidivierende depressive Störung prädiktiv für den Therapieabschluss festgestellt (Herzog et al., 2020). Daher ist auch ein fördernder Einfluss der Depression auf das Therapieansprechen denkbar. Es ist möglich, dass das Vorhandensein der Depression die Chance auf Verbesserung mit sich bringt, was sich positiv auf die *therapy response* auswirken kann.

Die zuvor erhobene Diagnose emotional instabile Persönlichkeitsstörung vom impulsiven Typ (F60.30) betont das Vorhandensein der emotionalen Instabilität

und mangelnde Impulskontrolle vor anderen Symptomen, die den Borderline-Typ definieren (Graubner, 2013). Die Impulsivität wird auch in der ZAN-BPD erfragt, dessen Gesamtpunktzahl eine prädiktive Stärke in Hinblick auf die Vorhersage der *therapy response* zugesprochen wird (Schmitgen et al., 2019). Die Affektstörung gemeinsam mit der Impulsivität sind auch Faktoren, welche die BPS am besten von anderen Persönlichkeitsstörungen unterscheiden (Zanarini et al., 1998). Es existiert scheinbar eine Untergruppe bei BPS-Patienten, die sich zunächst durch eine hohe Impulsivität und emotionale Instabilität zeigt, dessen Eigenschaften die *therapy response* signifikant beeinflussen.

Die Impulsivität im Rahmen der BPS beinhaltet u.a. Essstörungen und Substanzmissbrauch (American Psychiatric Association & Association, 1994). Die Komorbidität der BPS mit Essstörungen (Zanarini et al., 1998) und dem Substanzmissbrauch sind bereits bekannt (McGlashan et al., 2000). Außerdem wurde bereits festgestellt, dass ein signifikanter Unterschied zwischen weiblichen und männlichen BPS-Patienten für das Auftreten mancher Erkrankungen besteht (Zanarini et al., 1998). Demnach zeigen männliche BPS-Patienten eine höhere Rate an Substanzmissbrauch u.a. Drogenabhängigkeit, wogegen Essstörungen wie die Anorexia nervosa öfter bei weiblichen Patienten vorkommen. Diese Erkenntnis legt nahe, dass geschlechtsspezifische Prädiktoren Einfluss auf die *therapy response* nehmen können und gemeinsam die Impulsivität abbilden, welche ebenfalls prädiktiv zu sein scheint.

Die Vergabe der Diagnose Persönlichkeitsstörung, nicht näher bezeichnet (F60.9) erfolgt, falls die Zuordnung der Symptome zu einer spezifischen Persönlichkeitsstörung nicht möglich ist. Bei der Diagnose Kombinierte und andere Persönlichkeitsstörungen (F61) weisen die Patienten oft Symptome verschiedener Persönlichkeitsstörungen auf, ohne dass eine dominierende Persönlichkeitsstörung festgestellt werden kann (Graubner, 2013). Beide Diagnosen haben gemeinsam, dass eine eindeutige Zuordnung der Patienten nicht möglich ist, weshalb möglicherweise eine passende Therapie nicht angeboten werden kann und dies sich auf die *therapy response* auswirkt. Es gibt Hinweise dafür, dass Symptome verschiedener Persönlichkeitsstörungen Prädiktoren für die *therapy response* sind (Jiménez et al., 2022). Es erscheint daher sinnvoll, dass die Diagnose F61 eine große Vorhersagekraft besitzt.

Studien zeigen, dass die schizotypen Störung (F21) mit der BPS korreliert (Grant et al., 2008). Das ausgefallene Verhalten und die seltsamen Vorstellungen dieser Patienten hängen möglicherweise mit niedriger Therapiemitarbeit und folglich Einschränkungen im Therapieerfolg zusammen (Graubner, 2013). Die medikamentöse Therapie und Psychotherapie könnte bei diesen Patienten mit besonderem Fokus auf die bestehende schizotypen Störung ergänzt und angepasst werden. Psychotische Symptome bei Patienten mit BPS führen zu sozialer Entfremdung und Isolation (Barnow et al., 2010). Außerdem verschlechtern sie das therapeutische *outcome* (Slotema et al., 2018). So ist denkbar, dass eine komorbide Schizophrenie (F20) und sonstige Schizophrenie (F20.8) die Schwere der BPS erhöht und die Behandlung erschwert. Umgekehrt wurde gezeigt, dass eine komorbide BPS bei Patienten mit Schizophrenie die psychischen Symptome negativ beeinflusst (Bahorik & Eack, 2010).

Neben komorbiden Erkrankungen hat sich die Einnahme einiger Medikamente als signifikant herausgestellt: Pipamperon-Einnahme während der Aufnahme, Quetiapin-Einnahme während der Entlassung, Aripiprazol-Einnahme während der Aufnahme und Entlassung, Vigantol-Einnahme während der Aufnahme und Entlassung sowie L-Thyroxin-Einnahme während der Aufnahme und Entlassung sowie Sertralins-Einnahme während der Entlassung. Die Diskussion über die signifikanten Antipsychotika und Antidepressiva erfolgt auf Basis der Studien in der S3 Leitlinie Borderline Persönlichkeitsstörung Version 2.0 ([https://register.awmf.org/assets/guidelines/038-015I\\_S3\\_Borderline-Persoenlichkeitssstörungen\\_2022-11.pdf](https://register.awmf.org/assets/guidelines/038-015I_S3_Borderline-Persoenlichkeitssstörungen_2022-11.pdf); abgerufen am 9. Februar 2024).

Pipamperon gehört zu den klassischen Antipsychotika, welche typischerweise mit Sedierung und Somnolenz assoziiert sind (Eugene et al., 2021). Die angeführten Antipsychotika Quetiapin und Aripiprazol gehören zur Kategorie der atypischen Antipsychotika. Auch für diese Gruppe wurde gezeigt, dass sie ein erhöhtes Risiko für unerwünschte Nebenwirkungen besitzen (Gartlehner et al., 2021). Beispielsweise erhöht Quetiapin das Vorkommen unerwünschter Ereignisse wie die Mundtrockenheit, Sedierung und Gewichtszunahme (Black et al., 2014). Diese Einschränkungen könnten den Verlauf der Behandlung negativ lenken und das Ansprechen auf die Therapie beeinflussen.

Es besteht Evidenz dafür, dass u.a. Aripiprazol und Quetiapin Symptome der BPS signifikant senken (Ingenhoven, 2011). Beide Medikamente besitzen einen signifikant positiven Effekt auf globale Funktionsfähigkeit und kognitive Wahrnehmungssymptome (Ingenhoven, 2011; Nickel et al., 2006). Aripiprazol senkt zusätzlich impulsives Verhalten und affektive Dysregulation, deren potenziell entscheidende Relevanz in Bezug auf die *therapy response* oben beschrieben wurde (Nickel et al., 2006). Die Minderung u.a. der Aggressivität, Symptome bezüglich zwischenmenschlicher Beziehungen sowie des BPS-Schweregrads wurde bei Quetiapin beobachtet (Black et al., 2014). Somit werden viele Faktoren der BSL-23, welche für die Einteilung in (Teil)Responder und Non-Responder fungierten, und dadurch auch das *outcome* der Therapie durch diese Medikamente positiv beeinflusst.

Das Antidepressivum Sertraline gehört zur Gruppe der selektiven Serotonin-Wiederaufnahmehemmer (De Vane et al., 2002). Da die Verbesserung der depressiven Symptome, gemessen anhand der BDI-Punktzahl, u.a. die *therapy response* definiert, erscheint die Signifikanz des *feature* Sertraline als positiver Prädiktor nachvollziehbar. Es ist naheliegend, dass Sertraline in gegenseitiger Wechselwirkung mit dem *feature* mittelschwere depressiven Episode (F32.1) steht und eine bedeutende Vorhersagekraft aufweist. Außerdem wurde in dieser Arbeit ein statistischer Unterschied bezüglich der Antidepressiva Einnahme zwischen den (Teil)Respondern und Non-Respondern in der Aufnahme beobachtet (s. Kapitel 3.2). Die ausgeprägtere Einnahme von Antidepressiva durch die Non-Responder legt komorbide Depressionssymptome nahe, welche möglicherweise den Therapieverlauf negativ beeinflussen.

Es bestehen Hinweise für den Zusammenhang von BPS-Symptomen wie Wut und dem Vit-D Spiegel (Baykara et al., 2023). Die Einnahme von Medikamenten wie Vigantol, um diesen Spiegel zu erhöhen, könnte möglicherweise vorteilhafte Effekte auf die Wirksamkeit der Therapie haben.

Schilddrüsenerkrankungen stellen eine somatische Komorbidität von Menschen mit BPS dar (Schneider et al., 2019). Folglich lässt sich eine Assoziation zwischen der BPS und der Substitution von L-Thyroxin vermuten. Der Effekt von der Einnahme auf die *therapy response* muss näher erforscht werden. Zudem besteht Evidenz für einen Zusammenhang von Depressionen mit Vitamin-D Mangel

(Casseb et al., 2019; Parker et al., 2017) sowie mit einer Hypothyreose (Bode et al., 2021; Gold et al., 1981). Somit erscheint die Signifikanz der *features* Einnahme von Vigantol und L-Thyroxin bei der Aufnahme und Entlassung über den Einfluss auf die Depression indirekt auch auf die BPS plausibel.

Verglichen mit ähnlichen Studien fällt auf, dass die Schwere der Erkrankung im Gegensatz zur vorliegenden Arbeit einen prädiktiven Wert besaß. Sowohl die Schwere der Depression (Schmitgen et al., 2019), als auch die Ausprägung der BPS-Symptomatik wiesen einen hohen prädiktiven Wert auf (Schmitgen et al., 2019; Yin et al., 2023). Zudem scheint der Bildungsgrad großen Einfluss auf die Verbesserung der Symptome und den Therapieabschluss zu haben (Herzog et al., 2020; Schmitgen et al., 2019). Das funktionelle *outcome*, was bei dieser Arbeit nicht direkt erfasst wurde, kann u.a. durch demographische Variablen beeinflusst werden (Herzog et al., 2020). Einen wichtigen Faktor für die *therapy response* stellt außerdem die Therapiemotivation dar (Herzog et al., 2020; Yin et al., 2023), welche sich beispielsweise durch die Hausaufgabenabschlussrate und dem Verbleib in der Behandlung ausdrückt (Yin et al., 2023). In der Studie von Jiménez und Kollegen (2022) wurden spezifische *items* diagnostischer Fragebögen untersucht (Jiménez et al., 2022). Dabei zeigten sie die Relevanz einzelner *items* diverser klinischer Fragebögen, welche u.a. die Selbstwahrnehmung, den Umgang mit Emotionen und zugrundeliegende Persönlichkeitszüge näher eruierten.

### 4.3 Nicht signifikante Klassifikationsmodelle

Das Modell für die strukturellen MRT-Daten hat sich trotz einer *accuracy* über 80 % nicht als signifikant erwiesen. Die Vergleichsstudie von Schmitgen und Kollegen (2019) zeigte eine signifikante Vorhersage durch strukturelle MRT-Daten (Schmitgen et al., 2019). Dies kann auf die Nutzung eines anderen Lernalgorithmus, dem *Random forest* (Breiman, 2001) zurückgeführt werden, welchem eine hohe Vorhersage insbesondere in biologischen Daten zugesprochen wird (Qi et al., 2006). Zudem wurde in deren Arbeit ausschließlich Hirnregionen mit einer Assoziation zur affektiven Dysregulation analysiert (Schulze et al., 2016). In das endgültige Modell flossen außerdem nur die prädiktivsten Hirnregionen ein. Eine unterschiedliche Vorgehensweise bestand auch bei der Bestimmung prädiktiver *features* unter Zuhilfenahme der Permutationswichtigkeit (Genuer et al., 2010).

Die Stichprobe war mit 31 Patienten wesentlich kleiner als in der vorliegenden Arbeit und es wurden nur weibliche Teilnehmer rekrutiert. So kann die Generalisierbarkeit in Frage gestellt werden, da die Ergebnisse möglicherweise nur auf Frauen oder die Stichprobe angepasst sind. Außerdem wiesen Responder und Non-Responder statistisch signifikante Unterschiede in der Schwere der BPS-Ausprägung (BSL, ZAN-BPD) und dem Ausmaß an Angst auf, welche potenziell mit signifikanten Hirnregionen korrelieren (Schmitgen et al., 2019).

Obwohl das Modell nicht signifikant ist, sind die herausgefundenen prädiktiven Hirnregionen im Hinblick auf den Zusammenhang mit BPS und *der therapy response* plausibel. Die ermittelten Hirnregionen sind: Amygdala, Lobus parietalis inferior (IPL), retrosplenialer Cortex (RSC), Thalamus, temporoparietaler Übergang (TPJ), ventrolateraler präfrontaler Cortex (VLPFC) und Cerebellum.

Die strukturelle Eigenschaft der Amygdala wird zur Diskriminierung der BPS gegenüber Kontrollgruppen und anderen psychiatrischen Erkrankungen verwendet (Grecucci et al., 2022). Darüber hinaus wurde die Vorhersagekraft auf die *therapy response* gezeigt (Schmitgen et al., 2019). Eine Einschränkung der Funktion der Amygdala bei der Regulierung von Emotionen und den Einfluss auf soziale Interaktionen (Whalen & Phelps, 2009) könnte einige Kernsymptome der BPS wie emotionale Instabilität und instabile zwischenmenschliche Beziehungen erklären.

Wichtige Funktionen des VLPFC im Kontext psychiatrischer Erkrankungen sind Reaktionsplanung und Hemmung von ungeeigneten Handlungen (Aron et al., 2014) sowie die emotionale Verarbeitung (Chick et al., 2020; He et al., 2018). So spielt der VLPFC eine Rolle bei der Impulsivitätskontrolle (Goya-Maldonado et al., 2010). Außerdem erstreckt sich der VLPFC u.a. im Pars opercularis des rechten unteren Frontalgyrus, für dessen GM-Volumen eine Assoziation mit BPS festgestellt wurde (O'Reilly, 2010; Schulze et al., 2016).

TPJ und IPL sind zwei miteinander zusammenhängende Hirnregionen, welche gemeinsam an der Regulierung diverser höherer kognitiver Funktionen beteiligt sind (Igelström & Graziano, 2017). Es wurde in funktionellen Analysen beobachtet, dass sie insbesondere im Rahmen einer reaktiven Aufmerksamkeitssteuerung, in den Prozessen der Selbstwahrnehmung und sozialer Interaktion wie z.B. das Hineinversetzen in die andere Person aktiviert werden (Igelström & Graziano, 2017). Eine beeinträchtigte Funktion in diesen Bereichen könnte

möglicherweise in Impulsivität durch Probleme bei der Aufmerksamkeitssteuerung, Identitätsstörung durch gestörte Selbstwahrnehmung und Schwierigkeiten in zwischenmenschlichen Beziehungen aufgrund fehlender sozialer Kompetenzen resultieren. Der Zusammenhang zwischen der BPS und strukturellen Hirndaten muss weiter erforscht werden.

Grecucci und Kollegen (2022) stellten ein erniedrigtes GM-Volumen für den Thalamus bei der Diskriminierung zwischen BPS-Patienten und der Kontrollgruppe fest (Grecucci et al., 2022). Strukturelle Auffälligkeiten in dieser Hirnregion wurden auch durch andere Studien berichtet, jedoch mit einem erhöhten GM-Volumen (Sampedro et al., 2021). Eine Veränderung des Volumens scheint bei BPS vorzuliegen, die genauen Umstände zum Anstieg und Abfall müssen eruiert werden. Der Thalamus spielt eine wichtige Rolle bei der Weiterleitung von sensorischen, motorischen sowie emotionalen Informationen an höhergelegene Hirnstrukturen (Jones, 1997). So könnte diese Hirnregion eventuell durch ihren Einfluss auf kognitive Funktionen wie Aufmerksamkeit und Bewusstsein sowie emotionale Funktionen mit der BPS zusammenhängen (Arend et al., 2015; Ward, 2013).

Die Assoziation des Cerebellums in strukturellen MRT-Daten mit der BPS und die Diskriminierungsfähigkeit gegenüber Kontrollgruppen sowie anderen psychiatrischen Erkrankungen ist bekannt (Grecucci et al., 2022; Schulze et al., 2016). Neben gängigen Funktionen wie der Koordinierung von motorischen Abläufen gibt es Hinweise für einen Einfluss des Cerebellums auf die Emotionsregulierung (Schutter & Van Honk, 2005, 2009) und Impulsivität (Miquel et al., 2019). Somit würden Veränderungen des Cerebellums im direkten Zusammenhang mit der Psychopathologie des BPS stehen.

Vor allem die Beteiligung am episodischen Gedächtnis und die Verarbeitung emotionaler Informationen durch den RSC legen einen Zusammenhang mit der BPS nahe (Maddock, 1999; Vann et al., 2009). So wurde in einer Studie eine Aktivierung des RSC beim Abruf von sowohl episodischen als auch semantischen Gedächtnisinhalten beobachtet (Mensebach et al., 2009). Außerdem fanden Kluetsch und Kollegen (2012) heraus, dass der RSC eine reduzierte Beteiligung an der Schmerzverarbeitung bei BPS-Patienten aufweist und dies möglicherweise Selbstverletzungen fördert (Kluetsch et al., 2012; Krause-Utz et al.,

2014). Bisher fehlen Untersuchungen struktureller Hirndaten bzgl. der Korrelation mit der BPS.

Die Kombination der beiden Modalitäten, klinisch-demographische – und strukturelle Daten, anhand von *bagging* und *stacking* hat sich trotz einer *accuracy* von mindestens 80 % nicht als signifikant erwiesen. Das wirft die Hypothese auf, dass die Kombination mehrerer Modalitäten möglicherweise die vollständige Leistungsfähigkeit eines Modells einschränken kann und daher kritisch überprüft werden sollte (Schmitgen et al., 2019). In der vorliegenden Arbeit sank die *accuracy* des klinischen Modells zugunsten der des strukturellen Modells und die Signifikanz bestand nicht mehr. Vielmehr sollte versucht werden, Vorteile und Stärken beider Modelle zu vereinen, um einen Mehrwert durch die Kombination zu erzielen.

Die Erstellung eines neuen Modells mithilfe signifikanter *features*, die *condensed version*, war nicht signifikant im Hinblick auf die Ermittlung der *therapy response*. Einen Einfluss darauf könnte die Nutzung eines praxisnahen RWD mit fehlenden Daten genommen haben. Es war auffällig, dass die Kombination beider Modalitäten die *accuracy* steigerte und somit vorteilhafter in Bezug auf eine ganzheitlichere Betrachtung des Patienten und eine genauere Schätzung des *therapy response* ist. Ein wesentlicher Vorteil der *condensed version* ist die schnelle Erhebung der notwendigen *features*, die im Vergleich zum Originalmodell erheblich reduziert sind. Sofern die *condensed version* weiterentwickelt wird und eine statistische Signifikanz zeigt, bietet sie die Chance, Patienten einfach in die Zielgruppen einteilen zu können. Auch bei der Vergleichsstudie von Schmitgen und Kollegen (2019) stellte sich die *condensed version* für die klinisch-demographischen Daten nicht als signifikant dar, wobei die für die strukturellen MRT-Daten signifikant war (Schmitgen et al., 2019).

Die anfangs aufgestellten Hypothesen haben sich bis auf das Fehlen von signifikanten strukturellen *features* und die *accuracy* der Modelle größtenteils bestätigt. Trotz Nicht-Signifikanz der strukturellen Daten decken sich prädiktive Hirnregionen mit denen der Vorgängerstudien. Außerdem besteht ein Zusammenhang zwischen der sich als signifikant herausgestellten komorbidien Erkrankungen wie die Depression mit den prädiktiven *features* aus Vergleichsstudien. Die Schwere der BPS hat sich im Vergleich zu anderen Studien nicht direkt als prädiktives

*feature* herausgestellt. Bei der Wahl des optimalen Modells muss zwischen der Präzision der Vorhersagen und der einfachen Umsetzbarkeit im klinischen Setting abgewogen werden. Präferiert man das signifikante Modell mit der höheren *accuracy*, kann man potenzielle Kosten und Frustration durch falsche Zuordnung der Patienten vermeiden. Dank der beschleunigten Therapieentscheidung bei einer weiterentwickelten *condensed version* könnte die Therapie schneller eingeleitet werden, was die Dauer des Aufenthalts verkürzen und finanziellen Nutzen mit sich bringen kann.

#### 4.4 Limitationen

In der vorliegenden Arbeit fehlte eine detaillierte und systematische Erhebung der individuellen Symptomatik anhand standardisierter Fragebögen in der klinischen Routine, die in Studienanalysen als relevant identifiziert wurden. Beispielsweise könnte durch Abfrage der einzelnen *items* des BSL-23, BDI, ZAN-BPD (Schmitgen et al., 2019) sowie Fragebögen wie etwa bzgl. des Umgangs mit den Gefühlen oder der Persönlichkeitsstruktur (Jiménez et al., 2022) die Psychopathologie eruiert werden. So hätten sie eventuell Prädiktoren für den Misserfolg aufdecken können. Stattdessen wurde in dieser Arbeit nur die Gesamtpunktzahl einiger Fragebögen als *feature* aufgeführt.

Obwohl die verwendete Stichprobengröße mit ähnlichen Studien vergleichbar war, konnte keine getrennte interne Validierung der Klassifikatoren an eine Teilmenge des Datensatzes durchgeführt werden. Es fand zwar eine wiederholte verschachtelte Kreuzvalidierung mit Einteilung in Trainings- und Testdaten statt, es sind jedoch alle Daten in die Analysen eingegangen. Außerdem fehlt die Implementierung einer externen Validierung auf einen unabhängigen Datensatz, um die Generalisierung zu überprüfen (Coutanche & Hallion, 2019). Dies ist insbesondere bei kleineren Stichprobengrößen besonders wichtig, da hier Fehlschätzungen der Vorhersagegenauigkeit häufig auftreten können (Poldrack et al., 2020). Außerdem wurden relativ viele *features* verwendet, was das Risiko des *overfitting* erhöhen kann (Coutanche & Hallion, 2019).

Die unterschiedliche Stichprobengröße seitens der (Teil)Responder und Non-Responder wurde bereits angesprochen. Die Tatsache, dass nicht genug Prädiktoren für den Misserfolg gefunden wurden, wird durch das Argument

unterstützt, dass eine kleinere Stichprobengröße weniger Prädiktoren generieren kann (Herzog et al., 2020). Die gewählte Gruppierung in (Teil)Responder ermöglichte eine erweiterte Betrachtung von Therapieansprechen, könnte jedoch dazu geführt haben, dass eine unterschiedliche Stichprobengröße generiert wurde. Darüber hinaus erfolgte keine Differenzierung bzgl. BPS als Haupt- oder Nebendiagnose.

Es muss beachtet werden, dass die Ergebnisse der vorliegenden Arbeit als Unterstützung bei der klinischen Entscheidungsfindung fungieren. Sie sollen diese nicht komplett ersetzen. Die Wahl der Therapie sollte weiterhin durch den Entscheidungsträger mit seiner Fachkompetenz getroffen werden, welcher die Patienten ganzheitlich betrachtet und sie in den Entscheidungsprozess involviert. Sonst droht eine Arzt-Patienten Beziehung, welche die Autonomie der Patienten stark einschränkt (Shutzberg, 2021).

## 4.5 Schlussfolgerung und Ausblick

In der vorliegenden Arbeit wurde ein ML-Ansatz zur Ermittlung prädiktiver *features* hinsichtlich der *therapy response* bei Patienten mit BPS beschrieben. Sie stellte einen Versuch dar, durch Auswertung multimodaler Daten das Fundament für eine Behandlungsstrategie zu legen, welche (Teil)Respondern eine passende Therapie anbietet und potenziellen Nicht-Respondern durch alternative Therapievorschläge anfallende Kosten und Frustration erspart (American Psychiatric Association, 2001). Durch die Verwendung von ML war es möglich, einen hochdimensionalen Datensatz mit zahlreichen *features* zu analysieren. Die Besonderheit lag an der Nutzung eines RWD mit fehlenden Daten, womit die Umstände in der realen klinischen Praxis abgebildet wurden. Es ist bemerkenswert, wie gut die aufgestellten ML-Modelle im Hinblick auf objektive Parameter wie Medikation, Diagnosestellung komorbider Erkrankungen sowie MRT-Daten mit Studienergebnissen vergleichbar waren. Die Plausibilität der Ergebnisse bestätigte sich durch Feststellung der meisten prädiktiven *features* in der bestehenden wissenschaftlichen Literatur. Spezifische klinische *features* bzgl. verwendeter Medikation und komorbider Erkrankungen sagten auf signifikantem Niveau die *therapy response* vorher. Angesichts erfolgreicher Ergebnisse in der Vergleichsstudie (Schmitgen et al., 2019) ist die Nutzung von strukturellen MRT-Daten trotz der Nicht-Signifikanz in vorliegender Arbeit zu empfehlen. Es könnte zielführend sein,

einen anderen Lernalgorithmus und alternative Analysestrategien zu erproben. Die Erstellung eines neuen Modells anhand der prädiktiven Hirnregionen, wie durch Schmitgen und Kollegen (2019) realisiert, würde potenziell eine bessere und signifikante Vorhersage erlauben (Schmitgen et al., 2019). Unter Umständen stünden die *features* als Komplex miteinander zusammen. Bei der Nutzung der *condensed version* muss der Nutzer zwischen einer präzisen Vorhersage und einer schnellen Entscheidungsfindung abwägen. Dessen Effektivität könnte durch zukünftige Studienanalysen optimiert werden. Prädiktoren für den Misserfolg sind in dem verwendeten Datensatz nicht ausreichend erfasst, möglicherweise könnten sie durch Erhebung der Psychopathologie identifiziert werden. Außerdem sollte eine externe Validierung mit einem unabhängigen, möglichst unterschiedlichen Testdatensatz durchgeführt werden.

Zukünftige prospektive longitudinale Studien unter kontrollierten Bedingungen könnten als Grundlage Ergebnisse dieser Arbeit verwenden, um so den Zusammenhang der Prädiktoren mit der BPS eingehender zu untersuchen. Auch die Erforschung kausaler Zusammenhänge sollte beispielsweise mittels randomisierter kontrollierter Studien angestrebt werden. Angesichts insbesondere der komorbidien Erkrankungen lässt sich vermuten, dass gewisse Subgruppen unter den Patienten mit BPS existieren, womit die Komplexität der BPS präziser aufgeschlüsselt werden kann. In weiteren Studien könnte durch Einteilung der Patienten in Subgruppen versucht werden, dem Bedarf dieser Patientengruppen angepasste Behandlungen anzubieten (Iniesta, Stahl, et al., 2016). Dadurch würden neue Prädiktoren entdeckt werden, u.a. solche, die den Misserfolg auch vorhersagen. Die verwendeten Datenmodalitäten könnte man durch die Nutzung von beispielsweise funktionellen MRT-Daten, anderen Gewichtungen der MRT-Bilder und EEG-Aufnahmen erweitern. Durch die Kombination derer würde man neue Erkenntnisse über die Interaktion ihrer *features* gewinnen. Dies könnte in der Zukunft das Vorgehen im klinischen Alltag prägen.

Zur weitergehenden Analyse der Effektivität der Studienergebnisse wäre der Vergleich zwischen einer konventionellen Therapieentscheidung und einer Patientenauswahl auf Grundlage des ML-Ansatzes sinnvoll. So könnte man die Präzision bei der Wahl potenziell geeigneter Patienten durch beispielsweise den

Einfluss auf direkte und indirekte Kosten, Dauer des Aufenthalts sowie Zufriedenheit der Patienten und Ärzte messen.

## Literatur- und Quellenverzeichnis

- Aas, I. M. (2011). Guidelines for rating global assessment of functioning (GAF). *Annals of general psychiatry*, 10, 1-11.
- Abe, S., & Abe, S. (2010). Feature selection and extraction. *Support vector machines for pattern classification*, 331-341.
- Abraham, P. F., & Calabrese, J. R. (2008). Evidenced-based pharmacologic treatment of borderline personality disorder: a shift from SSRIs to anticonvulsants and atypical antipsychotics? *Journal of Affective Disorders*, 111(1), 21-30.
- Al-Mejibli, I. S., Alwan, J. K., & Abd Dhafar, H. (2020). The effect of gamma value on support vector machine performance with different kernels. *International Journal of Electrical and Computer Engineering*, 10(5), 5497.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- American Psychiatric Association. (2001). Practice guideline for the treatment of patients with borderline personality disorder. American Psychiatric Association. *The American Journal of Psychiatry*, 158(10 Suppl), 1-52.
- American Psychiatric Association, A., & Association, A. P. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV* (Vol. 4). American psychiatric association Washington, DC.
- Arend, I., Henik, A., & Okon-Singer, H. (2015). Dissociating emotion and attention functions in the pulvinar nucleus of the thalamus. *Neuropsychology*, 29(2), 191.
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in cognitive sciences*, 18(4), 177-185.
- Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage*, 55(3), 954-967. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2010.12.049>
- Association, A. P. (2000). Diagnostic and statistical manual of mental disorders. *Text revision*.
- Association, A. P., & Spitzer, R. L. (1980). *Diagnostic and statistical manual of mental disorders*. American psychiatric association.
- Bahorik, A. L., & Eack, S. M. (2010). Examining the course and outcome of individuals diagnosed with schizophrenia and comorbid borderline personality disorder. *Schizophrenia Research*, 124(1), 29-35. <https://doi.org/https://doi.org/10.1016/j.schres.2010.09.005>
- Baloglu, O., Latifi, S. Q., & Nazha, A. (2021). What is machine learning? *Archives of Disease in Childhood-Education and Practice*.
- Barnow, S., Arens, E. A., Sieswerda, S., Dinu-Biringer, R., Spitzer, C., & Lang, S. (2010). Borderline Personality Disorder and Psychosis: A Review. *Current psychiatry reports*, 12(3), 186-195. <https://doi.org/10.1007/s11920-010-0107-9>
- Bayes, A., Spoelma, M., & Parker, G. (2022). Comorbid bipolar disorder and borderline personality disorder: Diagnosis using machine learning. *Journal of psychiatric research*, 152, 1-6.
- Bayes, A., Spoelma, M. J., Hadzi-Pavlovic, D., & Parker, G. (2021). Differentiation of bipolar disorder versus borderline personality disorder: A machine learning approach. *Journal of Affective Disorders*, 288, 68-73.

- Baykara, S., Karakaş, Ş., Şuheda, K., TAŞCI, G., & Selda, T. (2023). Relationship between impulsivity, aggression and blood ghrelin, vitamin D, lipids levels in borderline personality disorder. *Cukurova Medical Journal*, 48(2), 644-653.
- Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory-II. *Psychological assessment*.
- Bellino, S., Rinaldi, C., Bozzatello, P., & Bogetto, F. (2011). Pharmacotherapy of borderline personality disorder: a systematic review for publication purpose. *Curr Med Chem*, 18(22), 3322-3329. <https://doi.org/10.2174/092986711796504682>
- Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., Shea, M. T., Zanarini, M. C., Oldham, J. M., & Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of Psychiatry*, 158(2), 295-302.
- Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, 188(12), 2222-2239. <https://doi.org/10.1093/aje/kwz189>
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.
- Black, D., Zanarini, M., Romine, A., Shaw, M., Allen, J., & Schulz, S. (2014). Comparison of Low and Moderate Dosages of Extended-Release Quetiapine in Borderline Personality Disorder: A Randomized, Double-Blind, Placebo-Controlled Trial. *The American Journal of Psychiatry*, 171. <https://doi.org/10.1176/appi.ajp.2014.13101348>
- Black, D. W., Blum, N., Pfohl, B., & Hale, N. (2004). Suicidal behavior in borderline personality disorder: prevalence, risk factors, prediction, and prevention. *Journal of personality disorders*, 18(3: Special issue), 226-239.
- Bode, H., Ivens, B., Bschor, T., Schwarzer, G., Henssler, J., & Baethge, C. (2021). Association of hypothyroidism and clinical depression: a systematic review and meta-analysis. *JAMA Psychiatry*, 78(12), 1375-1383.
- Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R.-D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The short version of the Borderline Symptom List (BSL-23): development and initial data on psychometric properties. *Psychopathology*, 42(1), 32-39.
- Bohus, M., Limberger, M. F., Frank, U., Chapman, A. L., Kühler, T., & Stieglitz, R. D. (2007). Psychometric properties of the Borderline Symptom List (BSL). *Psychopathology*, 40(2), 126-132. <https://doi.org/10.1159/000098493>
- Bohus, M., Limberger, M. F., Frank, U., Sender, I., Gratwohl, T., & Stieglitz, R. D. (2001). [Development of the Borderline Symptom List]. *Psychother Psychosom Med Psychol*, 51(5), 201-211. <https://doi.org/10.1055/s-2001-13281> (Entwicklung der Borderline-Symptom-Liste.)
- Bonacorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145-1159. [https://doi.org/https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/https://doi.org/10.1016/S0031-3203(96)00142-2)
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24, 123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

- Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*, 4(7), 28-37.
- Caragea, C., & Honavar, V. G. (2009). Machine Learning in Computational Biology. In.
- Casseb, G. A., Kaster, M. P., & Rodrigues, A. L. S. (2019). Potential role of vitamin D for the management of depression and anxiety. *CNS drugs*, 33(7), 619-637.
- Chakravarthy, A. D., Bonthu, S., Chen, Z., & Zhu, Q. (2019). Predictive models with resampling: A comparative study of machine learning algorithms and their performances on handling imbalanced datasets. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA),
- Chang, C.-C., & Lin, C.-J. (2007). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2.
- Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, 3(3), 243-250.
- Chen, Y.-M., Chen, P.-C., Lin, W.-C., Hung, K.-C., Chen, Y.-C. B., Hung, C.-F., Wang, L.-J., Wu, C.-N., Hsu, C.-W., & Kao, H.-Y. (2023). Predicting new-onset post-stroke depression from real-world data using machine learning algorithm. *Frontiers in Psychiatry*, 14, 1195586.
- Chick, C. F., Rolle, C., Trivedi, H. M., Monuszko, K., & Etkin, A. (2020). Transcranial magnetic stimulation demonstrates a role for the ventrolateral prefrontal cortex in emotion perception. *Psychiatry Research*, 284, 112515. <https://doi.org/https://doi.org/10.1016/j.psychres.2019.112515>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.
- Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-world evidence and real-world data for evaluating drug safety and effectiveness. *Jama*, 320(9), 867-868.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. (2009). Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS One*, 4(7), e6353. <https://doi.org/10.1371/journal.pone.0006353>
- Coutanche, M. N., & Hallion, L. S. (2019). Machine learning for clinical psychology and clinical neuroscience.
- Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *JAMA Psychiatry*, 74(4), 319-328.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.
- De Vane, C. L., Liston, H. L., & Markowitz, J. S. (2002). Clinical pharmacokinetics of sertraline. *Clinical pharmacokinetics*, 41, 1247-1266.
- Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., Triantafyllou, C., Saygin, Z. M., Whitfield-Gabrieli, S., & Hofmann, S. G. (2013). Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry*, 70(1), 87-97.
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91-118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>

- El Naqa, I., & Murphy, M. J. (2015). *What is machine learning?* Springer.
- Ellison, W. D., Rosenstein, L. K., Morgan, T. A., & Zimmerman, M. (2018). Community and clinical epidemiology of borderline personality disorder. *Psychiatric Clinics*, 41(4), 561-573.
- Eugene, A., Eugene, B., Masiak, M., & Masiak, J. (2021). Head-to-Head Comparison of Sedation and Somnolence Among 37 Antipsychotics in Schizophrenia, Bipolar Disorder, Major Depression, Autism Spectrum Disorders, Delirium, and Repurposed in COVID-19, Infectious Diseases, and Oncology From the FAERS, 2004–2020. *Frontiers in Pharmacology*, 12. <https://doi.org/10.3389/fphar.2021.621691>
- Fiedler, P. (2018). Epidemiologie und Verlauf von Persönlichkeitsstörungen. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*.
- Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, 8(4), 376-386.
- Flandin, G., & Novak, M. J. U. (2020). fMRI Data Analysis Using SPM. In S. Ulmer & O. Jansen (Eds.), *fMRI: Basics and Clinical Applications* (pp. 89-116). Springer International Publishing. [https://doi.org/10.1007/978-3-030-41874-8\\_8](https://doi.org/10.1007/978-3-030-41874-8_8)
- Franken, K., Ten Klooster, P., Bohlmeijer, E., Westerhof, G., & Kraiss, J. (2023). Predicting non-improvement of symptoms in daily mental healthcare practice using routinely collected patient-level data: a machine learning approach. *Frontiers in Psychiatry*, 14, 1236551.
- Friston, K. J. (2003). Statistical Parametric Mapping. In R. Kötter (Ed.), *Neuroscience Databases: A Practical Guide* (pp. 237-250). Springer US. [https://doi.org/10.1007/978-1-4615-1079-6\\_16](https://doi.org/10.1007/978-1-4615-1079-6_16)
- Garcia, S., & Herrera, F. (2008). An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of machine learning research*, 9(12).
- Gartlehner, G., Crotty, K., Kennedy, S., Edlund, M., Ali, R., Siddiqui, M., Fortman, R., Wines, R., Persad, E., & Viswanathan, M. (2021). Pharmacological Treatments for Borderline Personality Disorder: A Systematic Review and Meta-Analysis. *CNS drugs*, 35. <https://doi.org/10.1007/s40263-021-00855-4>
- Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., Luders, E., & Initiative, A. s. D. N. (2022). CAT—A computational anatomy toolbox for the analysis of structural MRI data. *bioRxiv*, 2022.2006. 2011.495736.
- Gaser, C., & Kurth, F. (2021). *Manual Computational Anatomy Toolbox - CAT12*. <http://dbm.neuro.uni-jena.de/cat12/CAT12-Manual-old.pdf> [Abgerufen am 30. November 2023]
- Gaspar, P., Carbonell, J., & Oliveira, J. L. (2012). On the parameter optimization of Support Vector Machines for binary classification. *Journal of Integrative Bioinformatics*, 9(3), 33-43. <https://doi.org/doi:10.1515/jib-2012-201>
- Ge, F., Jiang, J., Wang, Y., Yuan, C., & Zhang, W. (2020). Identifying suicidal ideation among Chinese patients with major depressive disorder: evidence from a real-world hospital-based study in China. *Neuropsychiatric Disease and Treatment*, 665-672.
- Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225-2236.

- Gold, M. S., Pottash, A., & Extein, I. (1981). Hypothyroidism and depression: evidence from complete thyroid function evaluation. *Jama*, 245(19), 1919-1922.
- Gollapudi, S. (2016). *Practical machine learning*. Packt Publishing Ltd.
- Gómez-Verdejo, V., Parrado-Hernández, E., Tohka, J., & Alzheimer's Disease Neuroimaging, I. (2019). Sign-Consistency Based Variable Importance for Machine Learning in Brain Imaging. *Neuroinformatics*, 17(4), 593-609. <https://doi.org/10.1007/s12021-019-9415-3>
- Gong, Q., Wu, Q., Scarpazza, C., Lui, S., Jia, Z., Marquand, A., Huang, X., McGuire, P., & Mechelli, A. (2011). Prognostic prediction of therapeutic response in depression using high-field MR imaging. *NeuroImage*, 55(4), 1497-1503.
- Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *NeuroImage*, 14(1), 21-36. <https://doi.org/https://doi.org/10.1006/nimg.2001.0786>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Goya-Maldonado, R., Walther, S., Simon, J., Stippich, C., Weisbrod, M., & Kaiser, S. (2010). Motor impulsivity and the ventrolateral prefrontal cortex. *Psychiatry Research: Neuroimaging*, 183(1), 89-91.
- Grabe, H. J., & Giertz, K. (2020). Die Borderline-Persönlichkeitsstörung in den psychosozialen, psychotherapeutischen und psychiatrischen Versorgungssystemen von Deutschland. Psychotherapie Forum,
- Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., Smith, S. M., Dawson, D. A., Pulay, A. J., & Pickering, R. P. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of clinical psychiatry*, 69(4), 533.
- Graubner, B. (2013). *ICD-10-GM 2014 Systematisches Verzeichnis: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme 11. Revision-German Modification Version 2014*. Deutscher Ärzteverlag.
- Grecucci, A., Lapomarda, G., Messina, I., Monachesi, B., Sorella, S., & Siugzdaite, R. (2022). Structural features related to affective instability correctly classify patients with borderline personality disorder. A supervised machine learning approach. *Frontiers in Psychiatry*, 13, 804440.
- Grira, N., Crucianu, M., & Boujema, N. (2004). Unsupervised and semi-supervised clustering: a brief survey. *A review of machine learning techniques for processing multimedia content*, 1(2004), 9-16.
- Gunderson, J. G., & Elliott, G. R. (1985). The interface between borderline personality disorder and affective disorder. *The American Journal of Psychiatry*, 142(3), 277-288.
- Guy, W. (1976). *ECDEU assessment manual for psychopharmacology*. US Department of Health, Education, and Welfare, Public Health Service ....
- Hahn, T., Kircher, T., Straube, B., Wittchen, H.-U., Konrad, C., Ströhle, A., Wittmann, A., Pfleiderer, B., Reif, A., & Arolt, V. (2015). Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry*, 72(1), 68-74.
- Hamilton, M., & Guy, W. (1976). Hamilton depression scale. *Group*, 1(4).

- Hao, Y., Zhang, J., Yang, L., Zhou, C., Yu, Z., Gao, F., Hao, X., Pang, X., & Yu, J. (2023). A machine learning model for predicting blood concentration of quetiapine in patients with schizophrenia and depression based on real-world data. *British Journal of Clinical Pharmacology*, 89(9), 2714-2725.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
- He, Z., Lin, Y., Xia, L., Liu, Z., Zhang, D., & Elliott, R. (2018). Critical role of the right VLPFC in emotional regulation of social exclusion: a tDCS study. *Social Cognitive and Affective Neuroscience*, 13(4), 357-366. <https://doi.org/10.1093/scan/nsy026>
- Heitjan, D. F., & Basu, S. (1996). Distinguishing "Missing at Random" and "Missing Completely at Random". *The American Statistician*, 50(3), 207-213. <https://doi.org/10.1080/00031305.1996.10474381>
- Hennen, J., & Silk, K. R. (2004). Mental health service utilization by borderline personality disorder patients and Axis II comparison subjects followed prospectively for 6 years. *J Clin Psychiatry*, 65(1), 28-36.
- Herzog, P., Feldmann, M., Voderholzer, U., Gärtnner, T., Armbrust, M., Rauh, E., Doerr, R., Rief, W., & Brakemeier, E.-L. (2020). Drawing the borderline: Predicting treatment outcomes in patients with borderline personality disorder. *Behaviour Research and Therapy*, 133, 103692.
- Hilbert, K., Jacobi, T., Kunas, S. L., Elsner, B., Reuter, B., Lueken, U., & Kathmann, N. (2021). Identifying CBT non-response among OCD outpatients: A machine-learning approach. *Psychotherapy Research*, 31(1), 52-62.
- Horner, M. S., & Diamond, D. (1996). Object relations development and psychotherapy dropout in borderline outpatients. *Psychoanalytic Psychology*, 13(2), 205.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417.
- Hubley, A. M. (2021). Beck depression inventory. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1-11). Springer.
- Igelström, K. M., & Graziano, M. S. (2017). The inferior parietal lobule and temporoparietal junction: a network perspective. *Neuropsychologia*, 105, 70-83.
- Ingenhoven, T. (2011). Differential Effectiveness of Antipsychotics in Borderline Personality Disorder Meta-Analyses of Placebo-Controlled, Randomized Clinical Trials on Symptomatic Outcome Domains. *Journal of clinical psychopharmacology*, 31, 489-496. <https://doi.org/10.1097/JCP.0b013e3182217a69>
- Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., & Stahl, D. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of psychiatric research*, 78, 94-102.
- Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological medicine*, 46(12), 2455-2465.
- Jackson, H. J., & Burgess, P. M. (2004). Personality disorders in the community: Results from the Australian National Survey of Mental Health and Well-Being Part III: Relationships between specific type of personality disorder, Axis 1 mental disorders and physical conditions with disability and health consultations. *Social Psychiatry and Psychiatric Epidemiology*, 39, 765-776.

- Jackson-Koku, G. (2016). Beck Depression Inventory. *Occupational Medicine*, 66(2), 174-175. <https://doi.org/10.1093/occmed/kqv087>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695. <https://doi.org/10.1007/s12525-021-00475-2>
- Jariani, M., Saaki, M., Nazari, H., & Birjandi, M. (2010). The effect of Olanzapine and Sertraline on personality disorder in patients with methadone maintenance therapy. *Psychiatria Danubina*, 22(4.), 544-547.
- Jiao, Y., & Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology*, 4(4), 320-330. <https://doi.org/10.1007/s40484-016-0081-2>
- Jiménez, S., Angeles-Valdez, D., Rodríguez-Delgado, A., Fresán, A., Miranda, E., Alcalá-Lozano, R., Duque-Alarcón, X., de Montis, I. A., & Garza-Villarreal, E. A. (2022). Machine learning detects predictors of symptom severity and impulsivity after dialectical behavior therapy skills training group in borderline personality disorder. *Journal of psychiatric research*, 151, 42-49.
- Jones, E. G. (1997). Cortical development and thalamic pathology in schizophrenia. *Schizophrenia bulletin*, 23(3), 483-501.
- Kalavathi, P., & Prasath, V. B. S. (2016). Methods on Skull Stripping of MRI Head Scan Images—a Review. *Journal of Digital Imaging*, 29(3), 365-379. <https://doi.org/10.1007/s10278-015-9847-8>
- Kelly, T., Soloff, P. H., Cornelius, J., George, A., Lis, J. A., & Ulrich, R. (1992). Can we study (treat) borderline patients? Attrition from research and open treatment. *Journal of personality disorders*, 6(4), 417-433.
- Khodayari-Rostamabad, A., Hasey, G. M., MacCrimmon, D. J., Reilly, J. P., & de Bruin, H. (2010). A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. *Clinical Neurophysiology*, 121(12), 1998-2006.
- Kliem, S., Kröger, C., & Kosfelder, J. (2010). Dialectical behavior therapy for borderline personality disorder: a meta-analysis using mixed-effects modeling. *Journal of Consulting and Clinical Psychology*, 78(6), 936.
- Kluetsch, R. C., Schmahl, C., Niedtfeld, I., Densmore, M., Calhoun, V. D., Daniels, J., Kraus, A., Ludaescher, P., Bohus, M., & Lanius, R. A. (2012). Alterations in default mode network connectivity during pain processing in borderline personality disorder. *Archives of General Psychiatry*, 69(10), 993-1002.
- Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., . . . Meisenzahl, E. (2021). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*, 78(2), 195-209. <https://doi.org/10.1001/jamapsychiatry.2020.3604>
- Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derk, E. M., Fleischhacker, W. W., & Hasan, A. (2016). Multisite prediction of 4-week

- and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *The Lancet Psychiatry*, 3(10), 935-946.
- Krause-Utz, A., Winter, D., Niedtfeld, I., & Schmahl, C. (2014). The Latest Neuroimaging Findings in Borderline Personality Disorder. *Current psychiatry reports*, 16(3), 438. <https://doi.org/10.1007/s11920-014-0438-z>
- Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*, 6(1), 1-15.
- Kurita, T. (2019). Principal component analysis (PCA). *Computer Vision: A Reference Guide*, 1-4.
- Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change*. John Wiley & Sons.
- Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, 377(9759), 74-84. [https://doi.org/https://doi.org/10.1016/S0140-6736\(10\)61422-5](https://doi.org/https://doi.org/10.1016/S0140-6736(10)61422-5)
- Lieb, K., Völlm, B., Rücker, G., Timmer, A., & Stoffers, J. M. (2010). Pharmacotherapy for borderline personality disorder: Cochrane systematic review of randomised trials. *The British Journal of Psychiatry*, 196(1), 4-12.
- Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *The Lancet*, 364(9432), 453-461.
- Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press.
- Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, 48(12), 1060-1064.
- Linehan, M. M., Tutek, D. A., Heard, H. L., & Armstrong, H. E. (1994). Interpersonal outcome of cognitive behavioral treatment for chronically suicidal borderline patients. *American Journal of Psychiatry*, 151(12), 1771-1775.
- Liu, F., & Panagiotakos, D. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, 22(1), 287.
- Maddock, R. J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in neurosciences*, 22(7), 310-316.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1), 381-386.
- Mai, J. K., & Majtanik, M. (2017). *Human Brain in Standard MNI Space: A Comprehensive Pocket Atlas*. Elsevier.
- Mar, J., Gorostiza, A., Ibarroondo, O., Cernuda, C., Arrospide, A., Iruin, Á., Larrañaga, I., Tainta, M., Ezpeleta, E., & Alberdi, A. (2020). Validation of machine learning models to predict dementia-related neuropsychiatric symptoms in real-world data.
- Marrelec, G., Krainik, A., Duffau, H., Péligrini-Issac, M., Lehéricy, S., Doyon, J., & Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1), 228-237. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2005.12.057>

- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., . . . Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*, 356(1412), 1293-1322. <https://doi.org/10.1098/rstb.2001.0915>
- McGlashan, T. H. (1986). The Chestnut Lodge follow-up study: III. Long-term outcome of borderline personalities. *Archives of General Psychiatry*, 43(1), 20-30.
- McGlashan, T. H., Grilo, C. M., Skodol, A. E., Gunderson, J. G., Shea, M. T., Morey, L. C., Zanarini, M. C., & Stout, R. L. (2000). The Collaborative Longitudinal Personality Disorders Study: baseline Axis I/II and II/II diagnostic co-occurrence. *Acta Psychiatrica Scandinavica*, 102(4), 256-264. <https://doi.org/https://doi.org/10.1034/j.1600-0447.2000.102004256.x>
- Mensebach, C., Beblo, T., Driessen, M., Wingenfeld, K., Mertens, M., Rullkoetter, N., Lange, W., Markowitsch, H. J., Ollech, I., & Saveedra, A. S. (2009). Neural correlates of episodic and semantic memory retrieval in borderline personality disorder: An fMRI study. *Psychiatry Research: Neuroimaging*, 171(2), 94-105.
- Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification.
- Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., Brázdl, M., & Krupa, P. (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, 26(4), 490-503. <https://doi.org/https://doi.org/10.1016/j.mri.2007.08.006>
- Miquel, M., Nicola, S. M., Gil-Miravet, I., Guarque-Chabreria, J., & Sanchez-Hernandez, A. (2019). A working hypothesis for the role of the cerebellum in impulsivity and compulsivity. *Frontiers in behavioral neuroscience*, 13, 99.
- Mishu, S. Z., & Rafiuddin, S. M. (2016). Performance analysis of supervised machine learning algorithms for text classification. *2016 19th International Conference on Computer and Information Technology (ICCIT)*, 409-413.
- Mitchell, T. M. (1997). Machine learning. In: McGraw-hill.
- Modestin, J., & Villiger, C. (1989). Follow-up study on borderline versus nonborderline personality disorders. *Comprehensive psychiatry*, 30(3), 236-244.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Mugler III, J. P., & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15(1), 152-157. <https://doi.org/https://doi.org/10.1002/mrm.1910150117>
- Muharemi, F., Logofătu, D., & Leon, F. (2019). Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecommunication*, 3(3), 294-307.
- Nakao, K., Gunderson, J. G., Phillips, K. A., Tanaka, N., Yorifuji, K., Takaishi, J., & Nishimura, T. (1992). Functional impairment in personality disorders. *Journal of personality disorders*, 6(1), 24-33.
- Nexus Deutschland GmbH. (2023). *Nexus / KIS-Container 9.10.00.0000* [Online Datenbank]. Informationen unter <https://www.nexus-ag.de/klinik/klinikinformationssystem> [Abgerufen am 28.Okttober 2023]

- Nickel, M. K., Mühlbacher, M., Nickel, C., Kettler, C., Gil, F. P., Bachler, E., Buschmann, W., Rother, N., Fartacek, R., Egger, C., Anvar, J., Rother, W. K., Loew, T., & Kaplan, P. (2006). Aripiprazole in the treatment of patients with borderline personality disorder: A double-blind, placebo-controlled study. *American Journal of Psychiatry*, 163, 833-838. <https://doi.org/10.1176/ajp.2006.163.5.833>
- O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends Neurosci*, 33(8), 355-361. <https://doi.org/10.1016/j.tins.2010.05.002>
- Oldham, J. M. (2006). Borderline personality disorder and suicidality. *American Journal of Psychiatry*, 163(1), 20-26.
- Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, 36(4), 1140-1152.
- Panos, P. T., Jackson, J. W., Hasan, O., & Panos, A. (2014). Meta-analysis and systematic review assessing the efficacy of dialectical behavior therapy (DBT). *Research on Social Work Practice*, 24(2), 213-223.
- Paris, J. (2002). Implications of long-term outcome research for the management of patients with borderline personality disorder. *Harvard Review of Psychiatry*, 10(6), 315-323.
- Paris, J., & Zweig-Frank, H. (2001). A 27-year follow-up of patients with borderline personality disorder. *Comprehensive psychiatry*, 42(6), 482-487.
- Parker, G. B., Brotchie, H., & Graham, R. K. (2017). Vitamin D and depression. *Journal of Affective Disorders*, 208, 56-61.
- Perez Arribas, I., Goodwin, G. M., Geddes, J. R., Lyons, T., & Saunders, K. E. (2018). A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational Psychiatry*, 8(1), 274.
- Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, 77(5), 534-540. <https://doi.org/10.1001/jamapsychiatry.2019.3671>
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21-45. <https://doi.org/10.1109/MCAS.2006.1688199>
- Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., Molosankwe, I., McCrone, P., & Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: a pragmatic randomised controlled trial. *Psychotherapy and psychosomatics*, 81(6), 356-365.
- Qi, Y., Bar-Joseph, Z., & Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins: Structure, Function, and Bioinformatics*, 63(3), 490-500.
- Robins, C. J., & Rosenthal, M. Z. (2011). Dialectical behavior therapy. *Acceptance and mindfulness in cognitive behavior therapy: Understanding and applying the new therapies*, 164-192.
- Rubin, D. B. (2018). Multiple imputation. In *Flexible Imputation of Missing Data, Second Edition* (pp. 29-62). Chapman and Hall/CRC.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249.

- Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc.
- Sampedro, F., Farrés, C. C. i., Soler, J., Elices, M., Schmidt, C., Corripio, I., Domínguez-Clavé, E., Pomarol-Clotet, E., Salvador, R., & Pascual, J. C. (2021). Structural brain abnormalities in borderline personality disorder correlate with clinical severity and predict psychotherapy response. *Brain imaging and behavior*, 15, 2502-2512.
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229. <https://doi.org/10.1147/rd.33.0210>
- Sansone, R. A., Farukhi, S., & Wiederman, M. W. (2011). Utilization of primary care physicians in borderline personality. *General Hospital Psychiatry*, 33(4), 343-346.
- Santos, M. S., Abreu, P. H., Wilk, S., & Santos, J. (2020). How distance metrics influence missing data imputation with k-nearest neighbours. *Pattern Recognition Letters*, 136, 111-119. <https://doi.org/https://doi.org/10.1016/j.patrec.2020.05.032>
- Saß, H. (2007). *Diagnostisches und statistisches Manual psychischer Störungen*.
- Sato, J. R., de Araujo Filho, G. M., de Araujo, T. B., Bressan, R. A., de Oliveira, P. P., & Jackowski, A. P. (2012). Can neuroimaging be used as a support to diagnosis of borderline personality disorder? An approach based on computational neuroanatomy and machine learning. *Journal of psychiatric research*, 46(9), 1126-1132.
- Schmitgen, M. M., Niedtfeld, I., Schmitt, R., Mancke, F., Winter, D., Schmahl, C., & Herpertz, S. C. (2019). Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging. *Brain and behavior*, 9(9), e01384.
- Schneider, F., Erhart, M., Hewer, W., Loeffler, L. A., & Jacobi, F. (2019). Mortality and medical comorbidity in the severely mentally ill—a German registry study. *Dtsch Arztbl International*, 116, 405-411. <https://doi.org/10.3238/arztebl.2019.0405>
- Schulze, L., Schmahl, C., & Niedtfeld, I. (2016). Neural correlates of disturbed emotion processing in borderline personality disorder: a multimodal meta-analysis. *Biological psychiatry*, 79(2), 97-106.
- Schutter, D. J., & Van Honk, J. (2005). The cerebellum on the rise in human emotion. *The Cerebellum*, 4, 290-294.
- Schutter, D. J., & van Honk, J. (2009). The cerebellum in emotion regulation: a repetitive transcranial magnetic stimulation study. *The Cerebellum*, 8, 28-34.
- Shah, F. P., & Patel, V. (2016). A review on feature selection and feature extraction for text classification. 2016 international conference on wireless communications, signal processing and networking (WiSPNET),
- Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, 2(1), 24-28. <https://doi.org/https://doi.org/10.1016/j.gltp.2021.01.004>
- Shinde, P. P., & Shah, S. (2018, 16-18 Aug. 2018). A Review of Machine Learning and Deep Learning Applications. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),
- Shreffler, J., & Huecker, M. R. (2022). *Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios*. StatPearls Publishing, Treasure Island (FL). <http://europepmc.org/abstract/MED/32491423>

<http://europepmc.org/books/NBK557491>

<https://www.ncbi.nlm.nih.gov/books/NBK557491>

Shubair, R. (2016). *Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis*. <https://doi.org/10.1109/ICEDSA.2016.7818560>

Shutzberg, M. (2021). The doctor as parent, partner, provider... or comrade? Distribution of power in past and present models of the doctor–patient relationship. *Health Care Analysis*, 29(3), 231-248.

Skodol, A. E., Gunderson, J. G., McGlashan, T. H., Dyck, I. R., Stout, R. L., Bender, D. S., Grilo, C. M., Shea, M. T., Zanarini, M. C., & Morey, L. C. (2002). Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder. *American Journal of Psychiatry*, 159(2), 276-283.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: psychopathology, comorbidity, and personality structure. *Biological psychiatry*, 51(12), 936-950.

Skodol, A. E., Pagano, M. E., Bender, D. S., Shea, M. T., Gunderson, J. G., Yen, S., Stout, R. L., Morey, L. C., Sanislow, C. A., & Grilo, C. M. (2005). Stability of functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder over two years. *Psychological medicine*, 35(3), 443-451.

Slotema, C., Blom, J. D., Niemantsverdriet, M. B., Deen, M., & Sommer, I. E. (2018). Comorbid diagnosis of psychotic disorders in borderline personality disorder: Prevalence and influence on outcome. *Frontiers in Psychiatry*, 9, 84.

Soeteman, D. I., Roijen, L. H.-v., Verheul, R., & Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *Journal of clinical psychiatry*, 69(2), 259.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). *Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation* (Vol. Vol. 4304). [https://doi.org/10.1007/11941439\\_114](https://doi.org/10.1007/11941439_114)

Soloff, P. H., & Chiappetta, L. (2012). Prospective predictors of suicidal behavior in borderline personality disorder at 6-year follow-up. *American Journal of Psychiatry*, 169(5), 484-490.

Soloff, P. H., Lis, J. A., Kelly, T., Cornelius, J., & Ulrich, R. (1994). Risk factors for suicidal behavior in borderline personality disorder. *The American Journal of Psychiatry*, 151(9), 1316-1323.

Soloff, P. H., Lynch, K. G., & Kelly, T. M. (2002). Childhood abuse as a risk factor for suicidal behavior in borderline personality disorder. *Journal of personality disorders*, 16(3), 201-214.

Stoffers, J., Völlm, B. A., Rücker, G., Timmer, A., Huband, N., & Lieb, K. (2010). Pharmacological interventions for borderline personality disorder. *Cochrane Database of Systematic Reviews*(6).

Stoffers, J. M., & Lieb, K. (2015). Pharmacotherapy for borderline personality disorder—current evidence and recent trends. *Current psychiatry reports*, 17, 1-11.

Stoffers, J. M., Völlm, B. A., Rücker, G., Timmer, A., Huband, N., & Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *Cochrane Database Syst Rev*, 2012(8), Cd005652. <https://doi.org/10.1002/14651858.CD005652.pub2>

- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2), 111-133.
- Stuke, H., Schoofs, N., Johanssen, H., Bermpohl, F., Ülsmann, D., Schulte-Herbrüggen, O., & Priebe, K. (2021). Predicting outcome of daycare cognitive behavioural therapy in a naturalistic sample of patients with PTSD: a machine learning approach. *European Journal of Psychotraumatology*, 12(1), 1958471.
- Subasi, M. M., Subasi, E., Anthony, M., & Hammer, P. L. (2011). A new imputation method for incomplete binary data. *Discrete Applied Mathematics*, 159(10), 1040-1047. <https://doi.org/https://doi.org/10.1016/j.dam.2011.01.024>
- Swartz, M., Blazer, D., George, L., & Winfield, I. (1990). Estimating the prevalence of borderline personality disorder in the community. *Journal of personality disorders*, 4(3), 257-272.
- Swenson, C. R., Sanderson, C., Dulit, R. A., & Linehan, M. M. (2001). The application of dialectical behavior therapy for patients with borderline personality disorder on inpatient units. *Psychiatric Quarterly*, 72, 307-324.
- Symons, M., Feeney, G. F., Gallagher, M. R., Young, R. M., & Connor, J. P. (2020). Predicting alcohol dependence treatment outcomes: a prospective comparative study of clinical psychologists versus 'trained'machine learning models. *Addiction*, 115(11), 2164-2175.
- The MathWorks, Inc. (2022). *Matlab*. In (Version R2022b) <https://www.mathworks.com/products/matlab.html> [Abgerufen am 30. November 2023]
- Tian, Y., & Zalesky, A. (2021). Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? *NeuroImage*, 245, 118648.
- Tolmeijer, E., Kumari, V., Peters, E., Williams, S. C., & Mason, L. (2018). Using fMRI and machine learning to predict symptom improvement following cognitive behavioural therapy for psychosis. *NeuroImage: Clinical*, 20, 1053-1061.
- Trigueiros, P., Ribeiro, F., & Reis, L. (2012). *A comparison of machine learning algorithms applied to hand gesture recognition*.
- Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10(66-71).
- Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature reviews neuroscience*, 10(11), 792-802.
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1), 1-8.
- Velez, D. R., White, B. C., Motinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M., & Moore, J. H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidemiology*, 31(4), 306-315. <https://doi.org/https://doi.org/10.1002/gepi.20211>
- Vieira, S. M., Kaymak, U., & Sousa, J. M. (2010). Cohen's kappa coefficient as a performance measure for feature selection. International conference on fuzzy systems,
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of child psychology and psychiatry*, 59(12), 1261-1270.

- Ward, L. M. (2013). The thalamus: gateway to the mind. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(6), 609-622.
- Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, 88(1), 25.
- Whalen, P. J., & Phelps, E. A. (2009). *The human amygdala*. Guilford Press.
- Whitfield-Gabrieli, S., Ghosh, S., Nieto-Castanon, A., Saygin, Z., Doehrmann, O., Chai, X., Reynolds, G., Hofmann, S., Pollack, M., & Gabrieli, J. (2016). Brain connectomics predict response to treatment in social anxiety disorder. *Molecular psychiatry*, 21(5), 680-685.
- Whitley, E., & Ball, J. (2002). Statistics review 3: hypothesis testing and P values. *Critical Care*, 6, 1-4.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241-259.  
[https://doi.org/https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/https://doi.org/10.1016/S0893-6080(05)80023-1)
- World Health Organization. (2004). *International Statistical Classification of Diseases and related health problems: Alphabetical index* (Vol. 3). World Health Organization.
- Xu, T., Cullen, K. R., Hour, A., Lim, K. O., Schulz, S. C., & Parhi, K. K. (2014). Classification of borderline personality disorder based on spectral power of resting-state fMRI. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,
- Yang, L., Zhang, J., Yu, J., Yu, Z., Hao, X., Gao, F., & Zhou, C. (2023). Predicting plasma concentration of quetiapine in patients with depression using machine learning techniques based on real-world evidence. *Expert Review of Clinical Pharmacology*, 16(8), 741-750.
- Yin, Q., Stern, M., Kleiman, E. M., & Rizvi, S. L. (2023). Investigating predictors of treatment response in Dialectical Behavior Therapy for borderline personality disorder using LASSO regression. *Psychotherapy Research*, 33(4), 455-467.
- Yu, C. H. (2019). Resampling methods: concepts, applications, and justification. *Practical Assessment, Research, and Evaluation*, 8(1), 19.
- Zanarini, M. C. (2003). Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. *Journal of personality disorders*, 17(3), 233-242.
- Zanarini, M. C., Frankenburg, F. R., Dubo, E. D., Sickel, A. E., Trikha, A., Levin, A., & Reynolds, V. (1998). Axis I comorbidity of borderline personality disorder. *American Journal of Psychiatry*, 155(12), 1733-1739.
- Zanarini, M. C., Frankenburg, F. R., Hennen, J., Reich, D. B., & Silk, K. R. (2004). Axis I comorbidity in patients with borderline personality disorder: 6-year follow-up and prediction of time to remission. *American Journal of Psychiatry*, 161(11), 2108-2114.
- Zarogianni, E., Moorhead, T. W., & Lawrie, S. M. (2013). Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. *NeuroImage: Clinical*, 3, 279-289.
- Zhutovsky, P., Thomas, R. M., Olff, M., van Rooij, S. J., Kennis, M., van Wingen, G. A., & Geuze, E. (2019). Individual prediction of psychotherapy outcome in

- posttraumatic stress disorder using neuroimaging data. *Translational Psychiatry*, 9(1), 326.
- Zimmerman, M., & Mattia, J. I. (1999). Axis I diagnostic comorbidity and borderline personality disorder. *Comprehensive psychiatry*, 40(4), 245-252.

## Anhang

**Tabelle A1.** Variablenliste und deren Anteil an fehlenden Daten

Name Variable	Fehlende Daten in %	Name Variable	Fehlende Daten in %
BDI_50%Rückgang	53,5	Aufnahme_Risperidon	0,0
BDI_30%Rückgang	53,5	Aufnahme_Methylphenidat	0,0
BSL-23_50%Rückgang	66,9	Aufnahme_Sedonium	0,0
BSL-23_30%Rückgang	66,9	Aufnahme_Xeplion	0,0
Qualitatives Therapiean-sprechen	3,2	Aufnahme_Tianeurax	0,0
Anzahl der Aufenthalte	0,0	Aufnahme_Reagila	0,0
Aufnahme	1,2	Aufnahme_Fluvoxamin	0,0
Entlassung	1,2	Aufnahme_Doxepin	0,0
Schule	25,5	Aufnahme_Agomelatin	0,0
Erwerbstätigkeit	7,6	Aufnahme_Milnacirpran	0,0
Beruf	15,9	Aufnahme_Opipramol	0,0
Aufnahme_Zopiclon	0,0	Aufnahme_Olanzapin	0,0
Aufnahme_Valproat	0,0	Aufnahme_Amisulprid	0,0
Aufnahme_Quetiapin	0,0	Aufnahme_Clozapin	0,0
Aufnahme_Mirtazapin	0,0	Sucht_Nikotin und Drogen	24,8
Aufnahme_Sertraline	0,0	Kognitive Verhaltensthera-pie	5,7
Aufnahme_Pipamperon	0,0	Entlassung_Zopiclon	0,0
Aufnahme_Vigantol	0,0	Entlassung_Valproat	0,0
Aufnahme_Dekristol	0,0	Entlassung_Quetiapin	0,0
Aufnahme_Trazodon	0,0	Entlassung_Mirtazapin	0,0
Aufnahme_Dominal	0,0	Entlassung_Sertraline	0,0

Aufnahme_Venlafaxin	0,0	Entlassung_Pipamperon	0,0
Aufnahme_Bupropionhydrochlorid	0,0	Entlassung_Vigantol	0,0
Aufnahme_Fluoxetin	0,0	Entlassung_Dekristol	0,0
Aufnahme_Citalopram	0,0	Entlassung_Trazodon	0,0
Aufnahme_Escitalopram	0,0	Entlassung_Dominal	0,0
Aufnahme_Duloxetin	0,0	Entlassung_Venlafaxin	0,0
Aufnahme_Agomaval	0,0	Entlassung_Bupropionhydrochlorid	0,0
Aufnahme_L-Tyroxin	0,0	Entlassung_Fluoxetin	0,0
Aufnahme_Thyronaiod	0,0	Entlassung_Citalopram	0,0
Aufnahme_Thiamazol	0,0	Entlassung_Escitalopram	0,0
Aufnahme_Lasea	0,0	Entlassung_Duloxetin	0,0
Aufnahme_Paroxat	0,0	Entlassung_Agomaval	0,0
Aufnahme_Trimineurin	0,0	Entlassung_L-Tyroxin	0,0
Aufnahme_Campral	0,0	Entlassung_Thyronaiod	0,0
Aufnahme_Atosil	0,0	Entlassung_Thiamazol	0,0
Aufnahme_Milnaneurax	0,0	Entlassung_Lasea	0,0
Aufnahme_Chlorprothixen	0,0	Entlassung_Paroxat	0,0
Aufnahme_Modafinil	0,0	Entlassung_Trimineurin	0,0
Aufnahme_Melperon	0,0	Entlassung_Campral	0,0
Aufnahme_Zyprexa	0,0	Entlassung_Atosil	0,0
Aufnahme_Aripiprazol	0,0	Entlassung_Milnaneurax	0,0
Aufnahme_Quilonum	0,0	Entlassung_Chlorprothixen	0,0
Aufnahme_Anafranil	0,0	Entlassung_Modafinil	0,0
Aufnahme_Amineurin	0,0	Entlassung_Melperon	0,0
Entlassung_Zyprexa	0,0	F60.6	0,0

Entlassung_Aripiprazol	0,0	F31.0	0,0
Entlassung_Quilonum	0,0	F31.3	0,0
Entlassung_Anafranil	0,0	F41.0	0,0
Entlassung_Amineurin	0,0	F70.1	0,0
Entlassung_Risperidon	0,0	F25.1	0,0
Entlassung_Methylphenidat	0,0	F70.0	0,0
Entlassung_Sedonium	0,0	F50.1	0,0
Entlassung_Xeplion	0,0	F45.40	0,0
Entlassung_Tianeurax	0,0	F19.1	0,0
Entlassung_Reagila	0,0	F33.3	0,0
Entlassung_Fluvoxamin	0,0	F50.2	0,0
F31.4	0,0	F60.9	0,0
F32.1	0,0	F20	0,0
F32.2	0,0	F23.1	0,0
F33.1	0,0	F60.2	0,0
F33.2	0,0	F48.1	0,0
F43.1	0,0	F94.8	0,0
F41.9	0,0	F21	0,0
F60.5	0,0	F14.2	0,0
F15	0,0	F31.8	0,0
F90.0	0,0	F31.1	0,0
F40.0	0,0	F60.7	0,0
F10.2	0,0	F41.1	0,0
F61	0,0	F44.81	0,0
F63.0	0,0	F44.2	0,0
F43.0	0,0	F79.1	0,0
F40.1	0,0	F41.2	0,0

F12.2	0,0	F42.1	0,0
F12.1	0,0	F32.9	0,0
F10.0	0,0	F31.6	0,0
F10.1	0,0	F23.0	0,0
F22.0	0,0	F44.9	0,0
F42.2	0,0	F20.1	0,0
F90.1	0,0	F20.8	0,0
F13.1	0,0	F20.0	0,0
F32.0	0,0	B15_24	0,0
F60.30	0,0	C51_58	0,0
F19.2	0,0	E10_E14	0,0
F50.8	0,0	E00_E90	0,0
F17.2	0,0	E66	0,0
F17.1	0,0	I44_49	0,0
F55.9	0,0	I60_69	0,0
F63.8	0,0	I70_79	0,0
F32.3	0,0	J43_44	0,0
F14.1	0,0	J45_J46	0,0
F60.8	0,0	S00_T98	0,0
F43.2	0,0	V01_Y98	0,0
X60_84	0,0	Suizidalität	3,8
Compliance	0,6	Suizidversuche	18,5
Verwahrlosung	2,5	Borderlinesymptome	11,5
Alter bei erster psychischer Behandlung	2,5	Depressive Stimmung	7,6
Jahr erster psychischer Behandlung	2,5	Selbstmordgedanken	26,8

Wachheit	1,9	Krankheitseinsicht	14,6
Bewusstsein	3,2	Depressionssymptome	0,0
Orientierung	6,4	GAF bei der Aufnahme	2,5
CGI bei der Aufnahme	2,5	GAF bei der Entlassung	16,6
CGI bei der Entlassung	16,6		

**Anmerkung.** BDI: *Beck-Depressions-Inventar*; BSL-23: Borderline Symptom Liste 23; CGI: *Clinical Global Impression*; GAF: *Global Assessment of Functioning*; Alle nach „Aufnahme“ oder „Entlassung“ aufgeführten Inhalte stellen Medikamentennamen oder ihre Inhaltsstoffe dar. Weitere Abkürzungen kennzeichnen ICD10-Codes.

## Danksagung

An erster Stelle bedanke ich mich bei PD Dr. Susanne Neufang für die kompetente und intensive Betreuung. Die zugewandte und hilfsbereite Art, die gewissenhafte Durchsicht der Arbeit und lösungsorientierte Begleitung haben die Umsetzung dieser Arbeit deutlich erleichtert und mich maßgeblich unterstützt.

Ein großer Dank geht an die Klinik und Poliklinik für Psychiatrie und Psychotherapie aus dem LVR-Klinikum Düsseldorf für die Bereitstellung der klinischen- und MRT-Daten.

Außerdem geht mein herzlicher Dank an die Institutsleitung Frau Prof. Eva Meisenzahl-Lechner und Herr Dr. Milenko Kujovic für die fachkundigen Anregungen im Hinblick auf klinisch bedeutsame Aspekte, welche die Arbeit bereichert haben.

Nicht zuletzt bedanke ich mich bei meinen geschätzten Freunden im privaten und universitären Umfeld. Eure Unterstützung hat mir enorm viel Kraft und Motivation gegeben. Besonders in schwierigen Situationen diente sie für mich als Anker, um an meinen Zielen festzuhalten.

Ich bedanke mich zutiefst bei meiner Familie für die bedingungslose Begleitung, Unterstützung und Liebe während meiner kompletten Karrierelaufbahn. Insbesondere geht hierbei ein großer Dank an meine Schwester Nurefsan Demirbuga für das Korrekturlesen der Arbeit. Mein besonderer Dank gilt meinen Eltern, die mich seit dem Kindesalter in jedem Lebensbereich, allen voran in der Bildung, mit Verzicht auf eigenen Bedürfnissen förderten.