Aus der Klinik und Poliklinik für Psychiatrie und Psychotherapie

Klinik der Universität München

Vorstand: Prof. Dr. Peter Falkai

Clinical data patterns supporting differential diagnosis of recent onset of depression and clinical high risk for psychosis

Differentialdiagnostische Einordnung von Patientinnen und Patienten in die depressive Ersterkrankung oder das klinische Hochrisikostadium für Psychosen mithilfe maschineller Mustererkennung

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von Antonia Wosgien aus
Bremen
2023

Mit Genehmigung der Medizinischen Fakultät der
Universität München

Berichterstatter:	Prof. Dr. Nikolaos Koutsouleris
Mitberichterstatter:	Prof. Dr. Guillaume Landry Prof. Dr. Dr. Susanne Lucae

Mitbetreuung durch den promovierten Mitarbeiter:	Dr. Lana Marija Kambeitz-Ilankovic

Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	26.01.2023

Zusammenfassung

Depressive Störungen sowie das Frühstadium einer Psychose beginnen typischerweise im späten Jugend- und frühen Erwachsenenalter. Die Depression zählt zu den psychischen Erkrankungen mit der höchsten Prävalenz und geht ebenso wie das Prodromalstadium der Psychose mit Beeinträchtigungen in mehreren Bereichen wie Kognition, Funktionalität und zwischenmenschliche Beziehungen einher. Beide Zustände weisen typische Kernmerkmale auf, werden jedoch insgesamt als sehr heterogen beschrieben. Besonders in frühen Stadien mentaler Erkrankungen sind Symptommuster weniger stabil als diffus. Hinzu kommt, dass die Rate an depressiven Erkrankungen bei Patientinnen und Patienten mit unterschwelligen psychotischen Symptomen mit bis zu 43% hoch ist. Diese Eigenschaften erschweren die differentialdiagnostische Einordnung.

Das grundlegende Ziel dieser Studie war es, mit Hilfe maschineller Lernverfahren ein charakteristisches Merkmalsmuster für beide Krankheitsbilder zu erkennen und die Differentialdiagnose zu erleichtern. Wir führten insgesamt vier Analysen durch, um 1) unter Einbezug unseres gesamten Datensatzes die wichtigsten Merkmale für die Klassifikation zu erfassen, 2) die einzelnen klinischen Domänen, in die sich unser Datensatz unterteilen lässt, separat für die Klassifikation zu nutzen und diese hinsichtlich der Sensitivität, Spezifität und balancierten Genauigkeit (engl. *balanced accuracy, BAC*) zu vergleichen, 3) die Generalisierbarkeit unseres Modells zu prüfen, indem wir einzelne Studienzentren ausließen und diese für die Validierung unseres Modells heranzogen.

Wir wandten maschinelles Lernen auf Daten von insgesamt 246 Patientinnen und Patienten an. Hierunter wurden 128 in die Gruppe der *recent-onset depression (ROD)* eingeordnet, während 118 die Kriterien des *clinical high risk (CHR)* für Psychose erfüllten. Rekrutiert wurden unsere Patientinnen und Patienten im Rahmen eines größeren Forschungsprojektes, der Personalized Prognostic Tools for Early Psychosis Management (PRONIA; https://www.pronia.eu/) Studie. Ziel der PRONIA-Studie ist es, Computermodelle zu entwickeln, welche die Früherkennung von psychotischen Erkrankungen verbessern und zudem die Prognose verschiedener Erkrankungen sowie das individuelle Therapieansprechen vorhersagen können. Um zu bewerten mit welcher Gewichtung verschiedene klinische Domänen zur Klassifikation beitragen,

zogen wir psychopathologische Symptome, Neurokognition, Persönlichkeitseigenschaften, das Funktionsniveau, Resilienz und autobiographische Erlebnisse für die Analysen heran.

In einem Kreuzvalidierungsverfahren konnte eine *support vector machine* zwischen Patientinnen und Patienten mit depressiven Störungen und solchen im Prodromalstadium der Psychose mit einer *BAC* von 74.3% (Sensitivität: 72.9%, Spezifität: 75.8%) unterscheiden. Die wichtigsten Domänen waren hierbei die Persönlichkeit sowie psychopathologische Symptome, welche sich über unsere Analysen hinweg als stabil erwiesen. Es fanden sich unter den wichtigsten Merkmalen vor allem Elemente aus drei Fragebögen: dem *Wisconsin Schizotypy Scales*, dem *NEO Five-Factor Inventory* und dem *Positive And Negative Syndrome Scale*. Betrachtet man jede Domäne separat, erzielten psychopathologische Symptome die beste Klassifikationsleistung mit einer *BAC* von 71.0% (Sensitivität: 71.8%, Spezifität: 70.2%). Anhand von Persönlichkeitsmerkmalen konnten die beiden Gruppen mit einer *BAC* von 64.2% (Sensitivität: 57.6%, Spezifität: 67.2%*)* unterschieden werden. Ein *leave-one-site-out* Kreuzvalidierungsverfahren erzielte eine *BAC* von 69.0% (Sensitivität: 66.9%, Spezifität: 71.1%).

Die gute Klassifikationsleistung zeigt, dass bei der Unterscheidung beider Gruppen ein generalisierbares, diskriminatives Muster existiert. Im Hinblick auf die zukünftige Forschung könnten die wichtigsten Merkmale unserer Analyse dazu beitragen, Patientinnen und Patienten mit einer depressiven Störung sowie solche im Prodromalstadium der Psychose effektiver zu identifizieren. Dies könnte die Grundlage bilden für ein neues Beurteilungsverfahren zur Unterstützung der Differentialdiagnose *CHR* vs. *ROD*.

Abstract

Major depressive disorder (MDD) and clinical high risk (CHR) for psychosis typically both commence in late adolescence and early adulthood. MDD as one of the most widely spread psychiatric disorders as well as CHR state are associated with impairments in several domains such as neurocognition and functioning. Although MDD and CHR state are differentiated through a number of core symptoms, their clinical manifestation is heterogeneous with rather diffuse and unstable symptom patterns in early stages of mental ill health. Additionally, comorbidity in CHR individuals is high and mood disorders can be diagnosed in up to 43% of patients, which makes differential diagnosis difficult.

The general aim of this study was to differentiate CHR from MDD patients using multivariate pattern analysis (MVPA) based on a behavioral, multi-domain dataset in order to investigate discriminative patterns for making differential diagnosis. Following three more specific aims, we conducted a set of four different analyses to 1) evaluate the most predictive feature patterns and their generalizability in an overall clinical classification, 2) analyze each of our five clinical domains separately in terms of their classification performance and compare sensitivity, specificity and balanced accuracy (BAC) and finally, 3) to test the geographic generalizability of the predictive patterns by leaving out single study centers and iteratively using them as a validation sample.

We applied machine learning to data from 246 patients, from which 118 were identified as CHR subjects and 128 with met criteria for recent-onset depression (ROD). Our patients were recruited for the Personalized Prognostic Tools for Early Psychosis Management (PRONIA; https://www.pronia.eu/) study, a research project funded by the European Union. PRONIA aims to develop a reliable prognostic system in order to predict the individualized risk of patients to develop affective and non-affective psychoses, contributing to a better understanding of early stages in mental ill health. Our feature set included data from five different clinical domains: psychopathological symptoms, neurocognition, personality traits, functioning, resilience and autobiographical experiences.

A support vector machine analysis, wrapped in a nested cross-validation framework, was able to differentiate between CHR and MDD with a significant BAC of 74.3% (sensitivity: 72.9%, specificity: 75.8%*)*. The most stable features across both overall clinical analyses were part of the personality and psychopathological symptoms domain. The most relevant questionnaires were the Positive and Negative Syndrome Scale (PANSS) with items assessing hallucinatory behavior, tension, conceptual disorganization and depression; the Wisconsin Schizotypy Scale (WSS) with items assessing magical ideation, perceptual aberration, physical and social anhedonia; the NEO Five-Factor Inventory (NEO-FFI) with items asking for neuroticism, openness and conscientiousness accordingly. The mono-clinical classifiers revealed that psychopathological symptoms could differentiate between CHR and ROD best with a BAC of 71.0% (sensitivity: 71.8%, specificity: 70.2%). The personality classifier yielded a BAC of 64.2% (sensitivity: 57.6%, specificity: 67.2%). Leave-one-site-out cross-validation (LOSOCV) estimating the multisite generalizability achieved an overall BAC of 69.0% (sensitivity: 66.9%, specificity: 71.1%).

The high classification performance confirms the presence of a generalizable, discriminative pattern able to differentiate between the two groups. With respect to future research, the most relevant feature subsets of our analysis might help to identify CHR and MDD more effectively and might even constitute the basis of a brief clinical assessment supporting differential diagnosis.

Acknowledgements

First of all, a huge “Thank you!” and my sincerest gratitude to both of my supervisors, Dr. Lana Marija Kambeitz-Ilankovic and Prof. Dr. Nikolaos Koutsouleris.

My thesis would not have been possible without the guidance and encouragement of Lana. She was so patient with me and always found the time to give me the feedback I needed- despite all the other things she had going on. I have the utmost respect for her dedication, her energy levels, and how committed she was to the team. Thanks to her, I have definitely gained a better understanding in the scientific and clinical field.

I would like to thank my supervisor Nikos for giving me the opportunity to join this outstanding research group and providing invaluable input and constructive feedback.

“Thank you” to the whole PRONIA-team for being the wonderful people they are, guiding me through the clinical process and providing help whenever necessary. A special thanks goes to Johanna Weiske for her continuous support on this work, sharing her knowledge with me and simply for her incredibly positive attitude.

Finally, I am profoundly grateful to my parents, my sister Kathi and to my friends. Nora, Ly, Eva and Moritz: your endless funny comments and encouraging words, while you yourselves were working on a thesis at the same time, always lifted me up whenever there was some kind of stumbling stone in front of me. All of you made this time more memorable.

Thank you for all your support, your patience and your love.

Contents

List of Abbreviations

ANOVA	Analysis of variance
APS	Attenuated psychotic symptoms
ARMS	At-risk mental state
ARMS-E	Early at-risk mental state
ARMS-L	Late at-risk mental state
BAC	Balanced accuracy
BDI-II	Beck Depression Inventory – Second Edition
BLIPS	Brief limited intermittent psychotic symptoms
CHR	Clinical high risk
CHARMS	Clinical high at-risk mental state
COGDIS	Cognitive disturbances scale
CTQ	Childhood Trauma Questionnaire
CV	Cross-validation
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FFM	Five-Factor Model
FROGS	Functional Remission of General Schizophrenia
GAF	Global Assessment of Functioning
GRS	Genetic risk and deterioration psychosis-risk syndrome
HC	Healthy control
LOSOCV	Leave-one-site-out cross-validation
M	Mean
MATRICS	Measurement and Treatment Research to Improve Cognition in Schizophrenia
MCCB	MATRICS Consensus Cognitive Battery
MDD	Major depressive disorder
MVPA	Multivariate pattern analysis
NEO-FFI	NEO Five-Factor Inventory
PANSS	Positive and Negative Syndrome Scale
PRONIA	Personalized prognostic tools for early psychosis managing

ROD	Recent-onset depression
ROP	Recent-onset psychosis
ROCF	Rey-Osterrieth Complex Figure
RSA	Resilience Scale for Adults
SD	Standard deviation
SPD	Schizotypal personality disorder
SPI-A	Schizophrenia Proneness Instrument - Adult Version
SVM	Support vector machine
UHR	Ultra-high risk
WSS	Wisconsin Schizotypy Scales

List of Tables

List of Figures

1 Introduction

1.1 Clinical high risk for psychosis

The manifestation of psychosis does not occur abruptly but is preceded by a prodromal phase, also referred to as clinical high-risk (CHR) state, which is often hard to identify as it is characterized by a high prevalence of comorbid psychiatric diagnoses (Addington et al., 2017; Fusar-Poli et al., 2013; Fusar-Poli et al., 2014). Several studies described non-psychotic comorbidity in nearly 80% of CHR patients, with depression and anxiety disorders being reported as the most common diagnoses (Addington et al., 2017; Fusar-Poli et al., 2014; Hui et al., 2013; Salokangas et al., 2012; Vargas et al., 2019). More specifically, up to 43% of CHR individuals suffer from a depressive disorder (Addington et al., 2017). The distress of affective symptoms such as depression or anxiety are described as one of the main subjective reasons for CHR individuals to seek help and might overshadow existing attenuated psychotic symptoms (Addington et al., 2017; Falkenberg et al., 2015; Shah et al., 2017).

CHR for psychosis can be divided into subgroups displaying different at-risk mental states (ARMS): The characteristics of early ARMS (ARMS-E) comprise basic symptoms that consist of subtle changes in formal thought structure and cognitive functioning (COGDIS) (McGlashan et al., 2010; Schultze-Lutter et al., 2012). Attenuated psychotic symptoms (APS) such as delusional mood, overvalued beliefs or disorganized communication as well as brief limited intermittent psychotic symptoms (BLIPS) are criteria of late ARMS (ARMS-L) which is also known as ultra-high risk (UHR) for psychosis (Klosterkötter, 2016; McGlashan et al., 2010; Schultze-Lutter et al., 2012). Another category of risk symptoms combines either familial risk factors, a diagnosed schizotypal personality disorder (SPD) or both, each with a marked decline in psychosocial functioning; this genetic risk and deterioration psychosis-risk syndrome (GRS) has been reported as criteria of both ARMS-E and UHR (Keshavan et al., 2011; Klosterkötter, 2016).

The subjective burden of these prodromal symptoms might be predated and even increased by affective disorders such as depression (Addington et al., 2017; Falkenberg et al., 2015; Shah et al., 2017). Co-occurring affective disorders are likely to impact ongoing psychopathology in

CHR patients such as suicidality or disorganized behavior and may predict poorer long-term outcomes (Fusar-Poli et al., 2014; Hui et al., 2013; Rutigliano et al., 2016).
Especially in early stages of mental ill health, symptom patterns can change quickly, are still comparatively broad and not very specific (McGorry, Hartmann, Spooner, & Nelson, 2018). In combination with the high rate of comorbidities, differential diagnosis becomes significantly more difficult (Falkenberg et al., 2015; Koutsouleris et al., 2018; McGorry et al., 2018; Wigman et al., 2012).

1.2 Major depressive disorder

Affective dysregulation such as depressive mood or states of anxiety has been reported to be strongly associated with subthreshold psychotic symptoms as they occur in prodromal psychosis (Falkenberg et al., 2015; Wigman et al., 2012). Affective dysregulation and slight perceptual distortions often occur together and both represent characteristic changes in early stages of mental ill health (Fusar-Poli et al., 2017; McGorry et al., 2018). Several studies suggest that the respective symptom patterns directly affect each other and even raise the idea of a shared vulnerability (Hanssen et al., 2003; Lin et al., 2015; Verdoux et al., 1999; Wigman et al., 2012). Recent research demonstrated that the onset of first episode psychosis can emerge out of risk states not captured by UHR criteria (Lee et al., 2018; McGorry et al., 2018) and that psychotic disorders may arise from other established mental disorders such as major depressive disorder (MDD; Fusar-Poli et al., 2017). Koutsouleris et al. (2020) observed that a notable proportion (17.4%) of patients with recent-onset depression (ROD) developed CHR symptoms during the follow-up period; 1.8 % of these cases eventually transited to full intensity manifestation of psychosis.
MDD is a serious, recurrent and wildly distributed psychiatric disease linked to impairments in areas of social, functional and interpersonal abilities as well as somatic morbidity and mortality (Bromet et al., 2011; McClintock et al., 2010). A comorbid attenuated psychotic pathology might worsen these impairments and additionally predict poorer outcome compared to depressive patients without psychotic symptoms (Perlis et al., 2011; Wigman et al., 2012). In a clinical assessment it can be difficult to distinguish a depression accompanied by attenuated psychotic symptoms from an emerging psychotic disorder with depressive and anxiety symptoms (Fusar-Poli et al., 2014).

Several studies investigated other aspects of MDD such as age of onset (Park et al., 2014; Zisook et al., 2004, 2007). Previous research indicates that there is a correlation between age of onset at first major depressive episode and clinical symptoms. An early age might have a negative impact on the course of illness with more frequent depressive episodes, a greater number of suicide attempts, more severe negative symptoms and more psychopathological symptoms associated with Axis I comorbidity; early recognition of MDD is therefore important in terms of early intervention and potential prevention of a more severe course of illness (Park et al., 2014; Zisook et al., 2004).

1.3 Neurocognition in CHR and MDD

Neurocognitive impairment is known as a significant and clinically relevant characteristic of both CHR for psychosis and MDD (Fusar-Poli et al., 2012; Rock et al., 2014). Numerous studies separately investigating cognitive performance in MDD and CHR subjects indicate that the two groups show overlapping deficits in several cognitive domains (Paolo Fusar-Poli et al., 2013; Rock et al., 2014; Zakzanis et al., 1999). Affective symptoms as they occur in depressive disorders are known to significantly affect cognitive abilities (Rock et al., 2014; Weightman et al., 2014). A meta-analysis examining cognitive performance in MDD patients revealed deficits particularly in the domains of executive functioning, memory and attention (Rock et al., 2014). Several studies investigated the negative impact of MDD on the neurocognitive domains of executive functioning, sustained attention and processing speed and described a correlation between symptom severity and cognitive impairment (McClintock et al., 2010). However, cognitive deficits partially seem to persist during remission, indicating that they occur separately from episodes of low mood in depression (Rock et al., 2014).

Comparatively, a meta-analysis on neurocognitive impairment in the CHR state suggests that CHR subjects share deficits in the domains of executive functioning, memory and attention with additional impairments in general intelligence, verbal fluency and social cognition (Fusar-Poli et al., 2012). Koutsouleris et al. (2012) investigated neurocognitive performance in different ARMS subjects versus healthy controls using machine learning techniques and found deficits in verbal learning/ memory as characteristic of ARMS-E whereas verbal IQ and executive functioning seem to be mainly impaired in ARMS-L. Though it might remain unclear whether any deterioration that occurs is specific to the prodromal period (Niendam et al., 2007), several

studies raise the idea of a subsequent deterioration of cognitive performance, starting from a state with subtle basic symptoms to full-blown psychosis (Frommann et al., 2011; Pukrop et al., 2006).

1.4 Personality traits in CHR and MDD

"Personality is the characteristic manner in which one thinks, feels, behaves and relates to others" (Widiger, 2011). Several studies investigated the influence of personality and psychopathology on one another and found that the two dimensions interact with each other in a bidirectional way (Widiger, 2011). Psychiatric disorders such as severe depressive episodes or psychotic experiences might fundamentally alter personality (Van Os & Jones, 2001; Widiger, 2011). Similarly, different personality traits such as high neuroticism may be associated with an increased vulnerability to develop a mental disorder (Dinzeo & Docherty, 2007; Van Os & Jones, 2001; Widiger, 2011).

Few studies investigated personality traits in CHR subjects with the aim to identify a characteristic profile and found that individuals symptomatically at risk of psychosis show comparatively high neuroticism and low extraversion scores (Schultze-Lutter et al., 2015). The high prevalence of depression in CHR patients which plays an important role regarding changes in neurocognition, social cognition and functioning might additionally contribute to an altered personality in patients at CHR for psychosis (Addington et al., 2017; Fusar-Poli et al., 2013; Schultze-Lutter et al., 2015).

For MDD, literature on personality is large and current evidence suggests that there is a correlation between depression and personality traits such as neuroticism, extraversion and conscientiousness (Klein et al., 2011). Some personality traits, neuroticism in particular, might even contribute to the onset, course and treatment response of depression (Klein et al., 2011). However, high neuroticism and low extraversion indicate a lower personality-related resilience, are associated with clinical disorders in general and might consequently reflect a general psychopathological status in MDD and CHR states (Ruhrmann et al., 2010; Schultze-Lutter et al., 2015; Widiger, 2011).

Looking beyond the FFM at different personality disorders, numerous studies have been describing both CHR and MDD as closely linked to SPD (Debbane et al., 2015; Emsley et al., 1999; Flückiger et al., 2016; Fonseca-Pedrero et al., 2011; Kwapil et al., 2008; Lewandowski

et al., 2006; Lysaker et al., 1995). In previous research, SPD has been alternatively linked to personality disorders or schizophrenia spectrum disorders and has additionally been described as a prodromal phase of schizophrenia (American Psychiatric Association, 2013; Cheli et al., 2019; Kwapil et al., 2008). The psychopathological pattern of SPD can be divided into positive, negative and disorganized symptoms: positive symptoms include ideas of reference, magical thinking and suspiciousness; negative symptoms on the other hand comprise anhedonia and reduced social interactions; disorganized symptoms involve disorganized thinking and impulsive nonconforming behavior (Cheli et al., 2019; Chemerinski et al., 2013; Mason, 1995). UHR criteria as well as mood disorders such as MDD have been reported to be closely linked to positive schizotypy (Debbane et al., 2015; Emsley et al., 1999; Lewandowski et al., 2006).

1.5 Functional disability in CHR and MDD

Profound social and occupational impairments are widely reported to be a relevant concern in the early development of psychosis as well as in patients diagnosed with MDD (Carrión et al., 2013; Fusar-Poli et al., 2013). These deficits reduce independence, affect interpersonal relationships, lower productivity, limit educational attainment and overall decrease quality of life (Bechdolf et al., 2005; Carrión et al., 2013).

Social and role impairments in CHR states are described as stable over time and independent from positive symptoms (Cornblatt et al., 2012). Functional deficits in combination with clinical and sociodemographic risk factors in the CHR state have been suggested to precede poor clinical outcomes and may be a critical predictor of long-term disability (Cornblatt et al., 2012; Fusar-Poli et al., 2010; Salokangas et al., 2014). A poor social functioning prognosis was reported to be associated with an increased prevalence of comorbid disorders at follow-up, MDD in particular (Addington et al., 2017; Koutsouleris et al., 2018; Wigman et al., 2012).

Similarly, early stages of depressive disorders are reported to be associated with a persistent functional decline and an overall reduced quality of life (Bora, Harrison, Yücel, & Pantelis, 2013). MDD and CHR individuals share these impairments in early stages of mental ill health which supports the idea of similar neurobiological changes in recent research (Koutsouleris et al., 2018).

1.6 The relevance of environmental factors for CHR and MDD

Childhood adversities and traumatic experiences were repeatedly identified as transdiagnostic risk factors for mental disorders. Previous studies have reported a link between different forms of childhood maltreatment such as sexual or emotional abuse, and mental disorders in adulthood (Bailer et al., 2014; Gibb et al., 2003, 2007; Haidl et al., 2020; Popovic et al., 2020). Emotional abuse in particular was strongly related to depressive disorders whereas sexual abuse may appear as a non-specific risk for psychopathology in general (Bailer et al., 2014; Gibb et al., 2003, 2007; Popovic et al., 2020). Some researchers focused on traumatic experiences in patients with UHR for psychosis and suggested that childhood maltreatment has a strong negative impact on the course and outcome of the illness (Şahin et al., 2013; Thompson et al., 2014). Sexual abuse experiences in particular might be associated with higher transition rates from UHR to psychosis (Thompson et al., 2014).

Another aspect regarding childhood trauma experiences is the importance of resilience. The term "resilience" can be understood as a protective factor as it describes the individual's ability to react resiliently in the aftermath of adversity (Wagnild & Young, 1993). Schulz and colleagues (2014) found that low resilience scores are associated with a higher risk for MDD. Similarly, researchers found that patients with CHR for psychosis show lower levels of resilience and that higher levels of resilience may result in lower negative symptoms, less depressive and anxiety symptoms as well as higher levels of role functioning (Marulanda & Addington, 2016).

1.7 Multivariate pattern analysis

1.7.1 Theory

Over the past few years, there has been a growing interest in machine learning methods in psychiatric research as it provides certain strengths in terms of differential diagnosis and predicting the course of disorders from different patient populations (Dwyer et al., 2018; Kambeitz et al., 2015; Klöppel et al., 2012; Orrù et al., 2012). The major aim of multivariate pattern analysis (MVPA) is to extract regularities in data and develop algorithms which can be used to predict outcome measures such as a particular psychiatric diagnosis (Wolfers et al., 2015).

These algorithms are called classifiers that learn the relationship between features (the data set, e.g. neurocognitive data) and the outcome measure (e.g. group membership: CHR or ROD) in a supervised way (Wolfers et al., 2015).
A supervised learning algorithm analyzes a training set with predefined labels and produces a function which is then validated on a separated, unseen test set to investigate its performance. Conversely, the classical interferential approach tries to fit data in already existing models and aims to detect statistically significant group differences. (Dwyer et al., 2018; Dwyer & McGuire, 2016) In recent years, these methods have come under increasing criticism as weaknesses have become apparent, for example, with regard to reproducibility (Dwyer et al., 2018). In this regard, MVPA methods have been successfully employed as alternative statistical tools as they provide optimal methods for classifying individuals on a single-subject level, operate with empirical estimation and additionally support an improved generalizability (Dwyer et al., 2018; Koutsouleris et al., 2012; Orrù et al., 2012). In contrast to univariate statistics, "[MVPA methods are able to analyze] complex multivariate relationships related to high-dimensional data" (Dwyer et al., 2018). This constitutes an important characteristic regarding the fact that mental health disorders present heterogeneous and rather diffuse symptom patterns including constant interactions within and among various domains such as neurocognition or environmental aspects (Borsboom & Cramer, 2013; McGorry et al., 2018). With respect to these characteristics, machine learning techniques could be highly useful in clinical practice (Dwyer et al., 2018; Koutsouleris et al., 2012; Orrù et al., 2012).

Support vector machine One of the most widely and effectively used classification algorithms is the support vector machine (SVM) (Orrù et al., 2012). SVM projects the data points into a high dimensional space where the classification function builds a decision boundary, a *hyperplane*, that separates these data points into groups (Orrù et al., 2012; Wolfers et al., 2015). This hyperplane is calculated from the data points, the so-called *support vectors*, that lie closest to the separating surface and thus constitute the hardest points to classify (Davatzikos et al., 2005). As an infinite number of hyperplanes exist, the algorithm tries to find the hyperplane that separates these support vectors best by iteratively maximizing the distance between them. (Davatzikos et al., 2005; Orrù et al., 2012) Figure 1 illustrates a binary classification based on SVM decision.

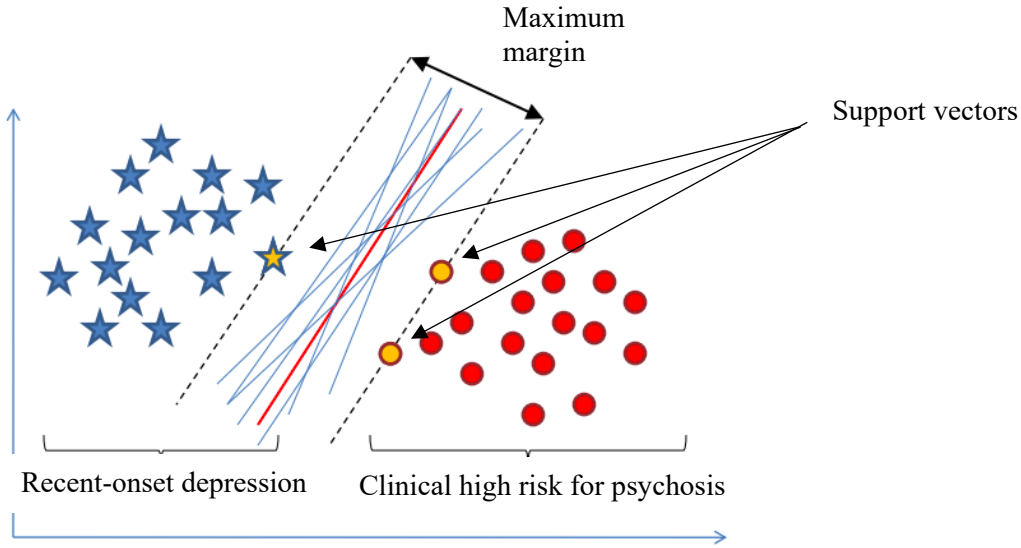

Figure 1: Simplified representation of a binary classification based on SVM decision scores. The algorithm builds a classification boundary between two groups basing on support vectors. The optimal separating hyperplane is the one which maximizes the margin between the support vectors. Figure from Statnikov, Hardin, Guyon & Aliferis (2009), slightly modified.

1.7.2 Application in psychiatric research

In recent years, MVPA methods, SVM in particular, have been increasingly applied in neuropsychiatry based on, for example, clinical assessment scales (Koutsouleris et al., 2012, 2016) or neuroimaging data (Kambeitz et al., 2017). Multivariate approaches have aimed to identify biomarkers, a biological measure for psychiatric diseases such as neurocognitive test data, which can be used for early recognition, treatment planning or prediction of disease progression (Dwyer et al., 2018; Orrù et al., 2012; Zarogianni et al., 2013).

A recent study found that "[v]ia thoroughly cross-validated machine learning methods, […] social functioning impairment can be correctly predicted in up to 83% of patients in CHR states and 70% of patients with ROD […]", based on functional level assessments, gray matter volume images and a combined feature set (Koutsouleris et al., 2018). The latter machine learning approach with combined neuroimaging and clinical models outperformed expert prognostication. Another recent study investigated the prognostic performance of multimodal machine learning methods with respect to psychosis transition of CHR and ROD patients and aimed to evaluate to what extent an integration of human and algorithmic prognostication might be valuable in clinical practice. Koutsouleris et al. (2020) showed that human prognostic abilities can be improved by integrating ratings performed by clinicians and algorithmic pattern recognition to margins that likely justify clinical implementation.

In a critical overview, Zarogianni et al. (2013) discussed the main difficulties of the potential translation of machine learning approaches into clinical practice and revealed important framework conditions. They suggested that a large sample with a certain heterogeneity, demonstrating a range of clinical manifestations, needs to be investigated in order to avoid overfitting and poor generalizability. Furthermore, the subjects need to be followed in a longitudinal manner with regard to a better reliability of biomarkers for psychiatric diseases (Orrù et al., 2012).
To our knowledge, MVPA methods have not been used to differentiate CHR from ROD patients until today.

1.8 Aims of this study

The general aim of this study was to differentiate CHR from ROD patients using MVPA based on a behavioral, multi-domain feature set. Our hypothesis was that there is a discriminative pattern which contributes to a better understanding regarding differential diagnosis of ROD and CHR state. To investigate our hypothesis, we performed a clear set of four multivariate analyses following three more specific aims:

1) First, this study examined the most predictive patterns of the CHR vs. ROD classification in a comprehensive clinical feature set using a sparse algorithm. In order to better separate out the top performing variables, we conducted a second analysis with an additional feature selection method. We expected this preprocessing step to potentially enhance classification performance.
2) We decided to analyze each of the five clinical data domains separately to compare sensitivity, specificity and balanced accuracy (BAC). In this regard, it was also interesting to see which features seem to be the most important within each domain and to compare them with the results of our overall clinical analyses.
3) Finally, we aimed to assess the degree to which our clinical model from the initial overall analysis generalizes across different geographic sites. We iteratively left single study centers out, trained models on the remaining sites and then applied the predictions to the left-out site.

2 Material and Methods

2.1 Participants

2.1.1 Study sample

The participants in this study were recruited as part of a European research project, the Personalized Prognostic Tools for Early Psychosis Management (PRONIA; https://www.pronia.eu/) study which collects multi-modal data from healthy control (HC) participants and young patients who meet the criteria for ROD, CHR and recent-onset psychosis (ROP) at baseline testing and several follow-up time points. In a multisite approach, the PRONIA study aims to develop reliable and generalizable prognostic tools which potentially increase prognostic certainty in terms of individualized future disease development.

In the current study, we used the clinical data of patients who meet the criteria for CHR and ROD. All participants provided their written informed consent prior to inclusion in the PRONIA study. A total of 246 patients, 128 with ROD and 118 with CHR for psychosis, were recruited at seven different PRONIA sites in Europe: the Departments of Psychiatry of the Ludwig-Maximilian-University (LMU) in Munich, Bavaria, Germany; University of Cologne (UKK) in Cologne, North-Rhineland Westphalia, Germany; University of Turku (Turku) in Turku, Finland; University of Basel (UBS) in Basel, Switzerland; the Institute of Mental Health at University of Birmingham (BHAM) in Birmingham, England; University of Udine (Udine) in Udine, Italy; and University of Milan (Milan) in Milan, Italy between February 2014 and June 2016. All patients underwent extensive assessments including clinical interviews composed of questionnaires such as the Schizophrenia Proneness Instrument – Adult version (Schultze-Lutter et al., 2007), self-rating questionnaires concerning personality traits and environmental risk factors as well as neurocognitive testing. For a full list of assessments conducted within PRONIA, please refer to Table A.1 in the Appendix. Demographic and clinical data of the study population is shown in Table 2.1.

Demographic and clinical variables	whole group	CHR	ROD
n	246	118	128
Mean age [years] (SD)	24 (5.7)	23 (5.0)	25 (6.0)
Sex (male) [%]	49.6	45.5	43.4
CHR [%]	48.0	100	-
Current treatment with			
Typical antipsychotics [%]	4.9	3.4	6.3
Atypical antipsychotics [%]	17.9	20.3	15.6
Mood stabilizer [%]	69.1	60.2	77.3
Participants per site, No. (%)	**whole group**	**CHR**	**ROD**
Munich	81 (33)	38 (32)	43 (34)
Basel	32 (13)	16 (14)	16 (12)
Milan	14 (6)	7 (6)	7 (5)
Cologne	44 (18)	19 (16)	25 (20)
Birmingham	27 (11)	13 (11)	14 (11)
Turku	23 (9)	13 (11)	10 (8)
Udine	25 (10)	12 (10)	13 (10)

Table 2.1: Sample characteristics and sample sizes of the seven European sites.

2.1.2 Inclusion and exclusion criteria

Inclusion criteria General criteria applying for both groups, CHR and ROD, were (i) age between 15 and 40 years, (ii) language skills sufficient for participation and (iii) sufficient capacity to consent.

For inclusion in the CHR group the patients had to fulfill the criteria for one of the following risk states: (i) Cognitive Disturbances (COGDIS) in the SPI-A, (ii) Attenuated Positive Symptom Psychosis-Risk Syndrome (APS), (iii) Brief Intermittent Psychotic Symptom Psychosis-Risk Syndrome, and / or (iv) Genetic Risk and Deterioration Psychosis-Risk Syndrome (GRS). The distribution of CHR individuals among the different subgroups can be seen in Table 2.2.

Patients were included as ROD if (i) criteria were fulfilled for MDD in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV-TR, American Psychiatric Association, 2000) (ii) within the past three months, (iii) it is the first MDD episode and (iv) the onset of depression was during the last 24 months.

Number of CHR among the different subgroups, No. (%)	whole group	COGDIS	APS	BLIPS	GRS	APS and GRS
n	118	31	72	3	9	3

Table 2.2: Distribution of the CHR subjects among the different subgroups. *Note:* COGDIS= Cognitive Disturbances; APS= Attenuated Positive Symptom Psychosis-Risk Syndrome; BLIPS: Brief Intermittent Psychotic Symptom Psychosis-Risk Syndrome; GRS= Genetic Risk and Deterioration Psychosis-Risk Syndrome

Exclusion criteria Exclusion criteria applying for all groups of the PRONIA study were (i) IQ below 70, (ii) hearing not sufficient for neurocognitive testing, (iii) current or past head trauma with loss of consciousness longer than 5 minutes, (iv) current or past known neurological disorder of the brain, (v) current of past known somatic disorder potentially affecting the structure or functioning of the brain, (vi) current or past alcohol dependency, (vii) current polytoxicomania (poly-dependency) or polytoxicomania (poly-dependency) within the past six months, and (viii) medical reasons that made MRI impossible.

In the CHR and ROD group the following specific exclusion criteria were applied: (i) antipsychotic medication for more than 30 days (cumulative number of days) at or above minimum dosage of the '1st episode psychosis' range of Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e.V. (DGPPN) S3 Guidelines, or (ii) any intake of antipsychotic medication (i.e., independent of duration of intake) within the past 3 months before psychopathological baseline assessments (including self-ratings and screening assessments) at or above minimum dosage of the '1st episode psychosis' range of DGPPN S3 Guidelines.

2.2 Clinical questionnaires

In our study, we used several behavioral assessments which comprised 276 features in total. This includes self-rating questionnaires to which the Wisconsin Schizotypy Scales (WSS), the NEO Five-Factor Inventory (NEO-FFI), the Beck's Depression Inventory (BDI-II), the Resilience Scale for Adults (RSA) and the Childhood Trauma Scale (CTQ) belong. Three Questionnaires have to be evaluated by a clinician, namely: the Functional Remission of General Schizophrenia (FROGS), the Global Assessment of Functioning (GAF) and the Positive and Negative Symptom Scale (PANSS). In addition to the clinical questionnaires, we administered a neurocognitive test battery to our subjects.

PANSS The standardized and validated PANSS measures the severity of psychotic symptoms (Kay et al., 1987). This questionnaire comprises 30 items, seven items refer to the excess of normal functions, the positive symptoms, such as hallucinations or disorganized thinking. Another seven items refer to the possible loss of functions in schizophrenia, the negative symptoms, characterized by cognitive and social impairment or altered affect. The last 16 items constitute a general psychopathology scale focusing on symptoms such as anxiety, guilt feelings or tension. Each question can be rated on a scale ranging from absent (one) to severe (seven) in an ascending order with more severe symptoms representing a higher score. (Kay et al., 1987) For a complete list of PANSS items, please refer to Table A.4 in the Appendix.

GAF The GAF is a measurement of the DSM-IV-TR that attempts to quantify the patient's overall level of functioning at different time points. The scale encompasses three major areas: psychological, social and occupational functioning which can be rated on a scale from one to 100. High scores display superior functioning, while low scores denote significant impairment. The focus lies hereby on the patient's subjectively perceived highest functioning: currently, in the past month, in the past year and highest functioning in lifetime. (Goldman et al., 1992; Hall, 1995)

Beside the overall estimate of its three dimensions, a split version of the GAF has been developed; the psychological dimension comprises the GAF-S (symptoms) construct, while the GAF-D (functional disability and impairment) construct is composed of the social and occupational dimensions (Pedersen et al., 2007; Pedersen & Karterud, 2012). In our study we included the GAF Disability and Impairment rating for past month, past year and lifetime.

FROGS The FROGS was constructed to evaluate remission in schizophrenic patients and aims to assess the functional level based on different domains. It consists of five domains (daily life, activities, relationships, quality of adaption, health and treatments) evaluating a total of 19 items (Llorca et al., 2009). Each item can be rated on a scale between one ("do not do") and five ("do perfectly"); the lowest score represents the worst level, while the highest is indicative of ideal functional recovery (Lançon et al., 2012).

BDI- II The BDI-II, created by Aaron T. Beck, is a widely used self-rating instrument to measure the severity of depression in clinical and research areas. The questionnaire retains 21

items with four options under each item, ranging from absent (zero) to severe (three). The different categories in the BDI-II cover questions about mood, behavior, activation and cognitive level. (Beck et al., 1996)

NEO-FFI The NEO-FFI is a well standardized measure composed of five broad personality trait domains, based on the model developed by Costa and McCrae (1992): these Big Five dimensions are commonly labeled neuroticism, extraversion, openness, agreeableness and conscientiousness. Each domain consists of twelve items which can be rated on a five-point scale ranging from "strongly disagree" to "strongly agree", resulting in 60 items for the whole questionnaire (Costa & McCrae, 1992). A complete list of NEO-FFI items can be found in Table A.4 in the Appendix.

WSS The WSS were developed in the 1970s and 1980s by Chapman and colleagues; meanwhile, the scales have become a frequently used tool for improved assessment of positive and negative schizotypy (Winterstein et al., 2011). The questionnaire comprises four scales: magical ideation and perceptual aberration (positive symptoms dimension) as well as social anhedonia and physical anhedonia (negative symptoms dimension). Magical ideation describes belief in magical influences, connections, and causalities that contradict culturally established ways of thinking. Perceptual aberration refers to perceptual and bodily distortions as they often occur in schizophrenic patients. The anhedonia scales were designed to assess negative schizotypy with physical anhedonia referring to reduced pleasure in sensory experiences and social anhedonia referring to deficits in social interactions and the desire and pleasure deriving from it. (Eckblad & Chapman, 1983; Kwapil et al., 2008)
Each scale originally consisted of 30 to 61 items but was then shortened in order to refine the scales with stronger items and make them more time efficient. The short form was found to reveal the effects that are also identified by the longer scales with a strong internal consistency. (Gross et al., 2015; Winterstein et al., 2011) In our study, we used this short form which comprises 15 items for each scale thus 60 items in total. For a complete list of WSS items, please refer to Table A.5 in the Appendix.

CTQ The CTQ is an originally 70-item screening inventory to assess self-reported abuse and neglect in childhood and adolescence before the age of 18. The questionnaire comprises five subscales: emotional, physical and sexual abuse as well as emotional and physical neglect. The

items consist of more objectively phrased items while others call for a more subjective evaluation. Each question can be rated on a five-point scale ranging from “never true” (zero) to “very often true” (four). (Bernstein et al., 1997)
In the PRONIA-study, a brief version of the CTQ, consisting of 28 items was used. It comprises the five subscales from the original questionnaire, each represented by five items. To address the risk of underreporting of maltreatment, a three-item minimization/denial validity scale was included (Bernstein et al., 1997, 2003). In several studies, the short version of the CTQ proved to be a valid and reliable self-rating instrument to retrospectively assess childhood maltreatment (Klinitzke et al., 2012; Wingenfeld et al., 2010).

RSA The RSA is a multifactorial scale based on large literature on resilience research, assessing protective factors across six domains: perception of self, planned future, social competence, structured style, family cohesion and social resources. It comprises 33 items in total, with seven options under each item ranging from one to seven. Higher scores on the RSA seem to imply a significant differentiation between healthy controls and psychiatric patients, a more well-adjusted big five personality profile, higher tolerance to pain and a less negative impact of stress on mental health. (Hilbig et al., 2015; Hjemdal et al., 2011)

2.3 Neurocognitive testing

In this study, a cross-domain neurocognitive test battery was administered to all subjects assessing IQ, processing speed, working memory, verbal and visual memory, social cognition as well as executive functions. We have based our subset of tests specifically on the recommended measurements of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative, which established the widely accepted MATRICS Consensus Cognitive Battery (MCCB) (Nuechterlein et al., 2008). The acquired neurocognitive data comprises 13 standardized tests from which 22 variables were computed, as shown more specifically in Table 2.3.

Cognitive Domain	Variables
Verbal IQ	
Wortschatz-Test (WT, WAIS-III)	1. Standard score
Fluid intelligence	
Matrizen-Test (MZ, WAIS-III)	2. Standard score
Processing speed	
Trail making Test, part A (TMT-A)	3. Time to completion (s)
Digit Symbol Test (DST, BACS)	4. Raw score correct
Working memory	
Subject Ordered Pointing task (SOPT)	
- 6 elements	5. Error score
- 8 elements	6. Error score
- 10 elements	7. Error score
Auditory Digit Span (forward) (FDS)	8. Maximum digits string length correct
Auditory Digit Span (backward) (BDS)	9. Maximum digits string length correct
Memory/Learning	
Rey Auditory Verbal Learning Test (RAVLT)	
- Immediate (RAVLT-IR)	10. Sum of raw score correct in trials 1-5
- Delayed (RAVLT-DR)	11. Raw score correct
- After interference	12. Raw score correct
Rey-Osterrieth Complex Figure (ROCF)	
- T0	13. Sum of correct elements 14. Time to completion
- Immediate recall	15. Sum of correct elements 16. Time to completion
- Delayed recall	17. Sum of correct elements 18. Time to completion
Executive functions	
Trail Making Test, part B (TMT-B)	19. Time to completion (s)
Verbal fluency (phonetic) (PVF)	20. Sum of correct responses
Verbal fluency (semantic) (SVF)	21. Sum of correct responses
Social cognition	
Diagnostic Analysis of Nonverbal Accuracy (DANVA)	22. Sum of correct responses

Table 2.3: Neurocognitive test battery. Cognitive domains were defined according to Schultze-Lutter et al. (2007) and Jaeggi, Buschkuehl, Jonides & Perrig (2008). See also Koutsouleris et al. (2012)

2.4 Data analyses

2.4.1 Sociodemographic and clinical data analysis

We performed chi-square tests to test for group differences in categorial variables, i.e. gender and site. An independent-samples t-test was conducted to compare the mean age of CHR subjects to the mean age of ROD subjects.

We additionally ran independent-samples t-tests to evaluate group differences in symptom severity based on the different questionnaires and the neurocognitive test battery we used for the classification. Regarding neurocognition, we chose the Rey-Osterrieth Complex Figure (ROCF) to represent the neurocognitive domain when investigating significant differences between CHR and ROD as well as potential site effects.

To consider any site effects with respect to the behavioral assessments within each group, we conducted one-way ANOVAs for CHR and ROD separately. Within the groups, a one-way ANOVA was conducted for each single assessment.

Significance levels were defined at $p < .05$; the Bonferroni-Holms procedure was used to correct for multiple comparisons. The analyses were run in MATLAB r2015a (The MathWorks, Natick, MA, USA, www.mathworks.com) on Linux.

2.4.2 Multivariate pattern analysis

MVPA was performed using the pattern recognition tool NeuroMiner (Koutsouleris & PRONIA WP2-Team; http://www.pronia.eu/the-project/work-plan/wp2-surrogate-marker/) to implement a fully automated machine learning pipeline.
The SVM algorithm first approached a subsample of the data with known labels and several features during the training phase in which the program learned decision rules from these features to classify CHR versus ROD on a single-subject level. Subsequently, during the testing phase, the learned classification rule was applied to the remaining unlabeled subsample and each subject was grouped in either one of the two groups, again on a single-subject level. (Dwyer, 2017)

2.4.2.1 Classification framework

Repeated double cross-validation In order to validate the classifier performance and to separate training from testing set, the pipeline was embedded into a repeated double cross-validation (CV) framework (Filzmoser et al., 2009).
In CV, the data set is split into k non-overlapping subgroups of which iteratively k-1 partitions are used to train the classifier (training data) and the left-out one to evaluate the classifier's generalizability (test data) (Varma & Simon, 2006).
In double CV, the training and test sets are split into an inner (CV1) and an outer (CV2) cycle, each divided into k folds. In CV1, the classification parameters are optimized by training on k-1 folds and testing on the left-out training set. All models originating from this inner loop are combined into an ensemble predictor by averaging the decision scores of the CV1 base learners. The ensemble classifier is then applied to the outer cycle to predict group membership of the unseen test subjects. The application of the derived model from CV1 to the left-out test data in CV2 avoids information leakage between training and test data and prevents overfitting. The predictions of unseen test subjects in the outer CV cycle were averaged to an ensemble decision value that was then aggregated for each test subject in the outer training partitions in which the subject was not involved. In majority voting, the decision scores that each model produces for the respective subject is then averaged to obtain the test subject's final group membership. (Koutsouleris et al., 2012; Varma & Simon, 2006)
In repeated double CV, participants are additionally randomly permuted across folds and the CV cycle is repeated for each of these permutations.

To summarize, repeated double CV (Filzmoser et al., 2009) creates an inner cycle, where pre-processing and machine learning parameters are computed, while the generalization error is estimated from the outer cycle. This procedure is repeated for each permutation of subjects within their groups and ensures a strict separation between training and test data, a high generalizability of the results and avoids overfitting. Figure 2.1 shows this process schematically.

In our analyses, we used a 10x10 repeated double CV where 10 permutations were performed with 10 folds each so that the inner cycle (CV1) presented 10 x 10 = 100 different training and validation sets for each CV2 test sample. On the outer cycle the same number of permutations and folds were generated, summing to a total of 100 x 100 = 10000 different training and test sets. This ensured a high variability in the data and thus a high degree of generalizability.

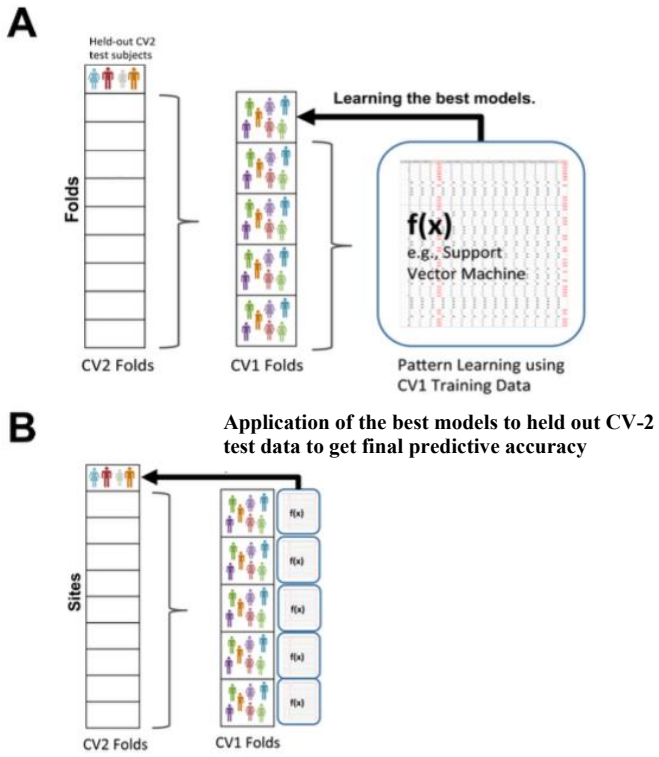

Figure 2.1: Repeated double cross-validation. Figure retrieved from the NeuroMiner Manual (Dwyer, 2017).

2.4.2.2 Preprocessing within the inner CV cycle

Analysis 1: Overall clinical classifier without an additional feature selection method

Prior to training a classifier, we preprocessed our data in the following steps: First, we standardized the data by subtracting the mean from the feature's absolute value and dividing it by the standard deviation. As a result, the mean of all variables was changed to zero and the standard deviation to one.

Then we imputed missing values using the k-nearest neighbor (k-NN) algorithm. This method "[…] uses observations in the neighborhood to impute missing values", requires a selection of the number of nearest neighbors and a distance metric (Tutz & Ramzan, 2015). In our analyses, we chose seven neighbors and Euclidean as the distance metric. K-NN takes a weighted average of the variable of the seven closest neighbors and uses it as the imputation estimate (Liu & Gopalakrishnan, 2017). Missing values could be found in 35.0% of our subjects which led to a total of 543 missing values in our feature set. We did not correct for age, gender, label or site in our analyses.

Analysis 2: Soft feature selection

In a following analysis, we added another step to the preprocessing pipeline: a soft feature selection method. The features were weighted in the training folds in terms of relevance to the label, based on Pearson's correlation coefficient, and then ranked from most to least correlated (Guyon & Elisseeff, 2003). Using feature selection methods such as filters allows for the unbiased selection of the most predictive features from a large feature set. By removing redundant, noisy or irrelevant features, classification accuracy can be increased and the probability of overfitting for noisy data is reduced. When using data with features of many different domains it is likely that those features provide different amounts of information and are consequently more or less useful in terms of predictive accuracy. (Chandrashekar & Sahin, 2014; Shardlow, 2016) In soft feature selection, the relevance measure is defined independently from the learning algorithm and focuses on intrinsic characteristics of the features based on univariate statistics instead of classifier performance (Guyon & Elisseeff, 2003; Kojadinovic & Wottka, 2000; Zhu et al., 2007). Hard feature selection methods such as wrappers on the other hand measure the utility of features based on classifier performance and can restrict the feature pool to a certain percentage aiming for an optimal model (Dwyer, 2017; Kohavi & John, 1997). We chose soft feature selection instead of hard feature selection to select the top performing variables as we already used a sparse algorithm (linear L1-regularized L2-loss support vector classification) and do not have as many features as when analyzing large data sets such as imaging data; we intended to achieve a balance between keeping as many variables as possible and potentially increasing accuracy by focusing on the most relevant features.

Analysis 3: Mono-clinical classifiers

The feature set comprises five different clinical domains: psychopathological symptoms, neurocognition, personality, functional level as well as the domain describing environmental factors and autobiographical experiences. We decided to build five mono-clinical classifiers, each representing one single domain. The same preprocessing pipeline and CV scheme as in Analysis 1 was used for each of the five classifiers.

Analysis 4: Multisite generalizability

We then implemented a leave-one-site-out cross-validation (LOSOCV) approach to assess the degree to which our model generalizes across different geographic sites. Generalizability can broadly be defined as the ability of an algorithm to perform accurately on previously unseen data (Dwyer et al., 2018).
In our study, the participants were recruited across seven different PRONIA sites in Europe which are described in more detail above and presented in Table 2.1. Each of the seven study sites was iteratively held back as a validation sample, while data from the remaining six sites entered the inner CV cycle where cases were again iteratively assigned to training and test data to identify the most predictive parameter combinations; an ensemble classifier is then applied to the outer cycle with data of the left-out site to predict group membership of the unseen test subjects (Dwyer et al., 2018; König et al., 2007; Koutsouleris et al., 2016). The same preprocessing pipeline as in Analysis 1 was used for this approach.

In each of our four analyses, the preprocessed training set entered a linear L1-regularized L2-loss support vector classification implemented in the LIBLINEAR library (Fan et al., 2008; https://www.csie.ntu.edu.tw/~cjlin/liblinear/) that determined the ideal between-group boundary. L1-regularization, also called Lasso regularization, uses the sum of absolute values of weights assigned to the features as a penalty term and adds it to the error of the hypothesis (in our case group membership). As with L1-regularization, less important features may be forced to be exactly zero, the most important features to our classification problem were revealed. (Fan et al., 2008)

2.4.2.3 Evaluation of the classifiers' performance

The performance of our classifiers were evaluated in means of Sensitivity, Specificity, Balanced Accuracy (BAC), False Positive Rate (FPR), Positive/Negative Predictive Value (PPV, NPV), Diagnostic Odds Ratio (DOR) and Number Needed to Diagnose (NND). Descriptions of these metrics can be found in Table 2.4.
We focused mainly on BAC which is defined as an average accuracy obtained for each label that avoids the problem of potential imbalances between groups (Brodersen et al., 2010; Wolfers et al., 2015).

In our fourth analysis, the generalization performance of our model was measured by comparing the accuracy achieved with the training models in the inner circle and the accuracy achieved for the validation in the outer circle. Serving as an overall estimate of the between-site generalizability, an average accuracy was obtained.
Statistical significance was assessed through permutation testing and defined as a *P* value less than .05 (Golland & Fischl, 2003; Koutsouleris et al., 2016).

Measure	**Description**
Sensitivity	The proportion of affected cases with a positive test result (true positives) in reference to all affected cases
Specificity	The proportion of nonaffected cases with a negative test result (true negatives) in reference to all nonaffected cases
Accuracy	The fraction of correctly predicted cases in reference to all cases
Balanced Accuracy	The accuracy in terms of true positive and negative cases balanced by the sample size of each positive and negative group; used to optimize groups with unbalanced sample sizes
False Positive Rate	The proportion of all the cases that are nonaffected which will be identified as affected
Positive predictive value	The probability that cases with positive test results are actually positive
Negative predictive value	The probability that cases with a negative test result are actually negative
Diagnostic odds ratio	A ratio of the probability that the test is positive in subjects who are positive for the condition relative to that for negative results
Youden's index	The addition of sensitivity and specificity minus 1 with a range between -1 (no discrimination) and 1 (perfect discrimination)
Number needed to diagnose	The inverse of Youden's index (1/Y); the number of subjects who need to be examined in order to correctly identify one affected subject

Table 2.4: Performance metrics used to interpret results and optimize predictions. This table was adapted from Dwyer et al., (2018), slightly modified (Larner, 2018).

Visualizing the ten most discriminative features In NeuroMiner, visualization refers to the representation of the model weights for each feature (Dwyer, 2017). The weighting of these features does not apprise of their statistical significance but informs about the relevance for the binary classification (Jankowski & Usowicz, 2011). We sorted the features for the classification CHR vs. ROD by CV-ratio. The CV-ratio refers to the mean weight of a feature divided by the standard error (Dwyer, 2017). We selected ten features with the highest CV-ratio for inspection.

3 Results

3.1 Sociodemographic and clinical data

No significant differences in gender were found when comparing CHR and ROD (χ^2 (245, *N=246*) *=.401, p=.587)*. Regarding interactions between label and site, no significant differences were found either (χ^2 (6, *N=246*) *=1.19, p =.977)*. The two groups differed significantly in terms of age, with a significantly younger mean age among CHR subjects *(T = 2.827, df=244, p =.005,* see Table 3.1).

CHR individuals scored significantly higher on the PANSS positive (*T* = 9.052, *p <.001*), the positive factor of the WSS comprising the perceptual aberration (*T* = 4.323, *p <.001*) and magical ideation scale (*T* = 5.901, *p <.001*). ROD individuals showed significantly higher scores on the FROGS (*T* = -2.262, *p =.02*) and the ROCF (sum of correct elements at T0: *T* = 2.155, *p =.03*). The results can be found in Table 3.1 and Figure 3.1.

Within the CHR group, the scores of the GAF differed significantly between the seven sites (highest score in past year: *F* = 5.63, *p* <.001; highest score in past month: *F* = 6.06, *p* <.001). For the BDI-II, the FROGS and the ROCF, significant differences were found, however, these differences did not survive multiple comparisons correction using the Bonferroni-Holms procedure. Within the ROD group, all scales of the PANSS differed significantly between the sites (PANSS total sum: *F* = 5.95, *p* < .001; PANSS positive sum: *F* = 3.68, *p* = .002; PANSS negative sum: *F* = 6.18, *p* <.001; PANSS general sum: *F* = 3.90, *p* = .001) as well as the conscientiousness domain of the NEO-FFI (*F* = 5.64; *p* <.001) and the GAF with highest lifetime score (*F* = 5.70, *p* <.001) and highest score in past month (*F* = 4.59, *p* <.001). Differences found for the BDI-II, the social anhedonia scale of the WSS, the GAF (highest lifetime score) and the ROCF were no longer significant after multiple comparisons correction. For an overview of the assessment scores for each site, please refer to Table 3.2 and Table 3.3. The results of our multiple comparison analyses can be found in Table A.5 and Table A.6 in the Appendix.

	CHR	ROD	T/χ^2	*P* Value
Sample sizes				
Total No.	118	128		
Participants per site, No. (%)				
Munich	38 (32)	43 (34)		
Basel	16 (14)	16 (12)		
Milan	7 (6)	7 (5)	$\chi^2 = 1.191$	.977
Cologne	19 (16)	25 (20)		
Birmingham	13 (11)	14 (11)		
Turku	13 (11)	10 (8)		
Udine	12 (10)	13 (10)		
Sociodemographic Variables				
Age: mean [years](SD)	23 (5.0)	25 (6.0)	$t = 3.712$	.005*
Gender: male/female (%)	54/64 (46/54)	55/73 (43/57)	$\chi^2 = .401$	.587
Psychopathological symptoms, mean (SD)				
PANSS total	50.8 (13.1)	47.6 (10.9)	$t = 2.138$	.034*
PANSS positive sum	10.3 (3.0)	7.7 (1.2)	$t = 9.204$	<.001*
PANSS negative sum	12.6 (5.8)	12.5 (5.0)	$t = 0.112$	.911
PANSS general sum	27.9 (6.9)	27.4 (6.7)	$t = 0.622$	.535
BDI-II	25.2 (12.4)	26.4 (13.6)	$t = -0.730$	.466
Personality, mean (SD)				
WSS - total	21.3 (5.8)	18.0 (4.7)	$t = 4.742$	<.001*
- magical ideation	3.3 (2.3)	1.8 (1.6)	$t = 5.901$	<.001*
- perceptual aberration	1.9 (2.4)	0.7 (1.8)	$t = 4.323$	<.001*
- physical anhedonia	9.1 (1.9)	8.6 (2.3)	$t = 1.840$	.067
- social anhedonia	7.0 (2.2)	6.7 (1.8)	$t = 1.047$	.296
NEO-FFI - neuroticism	39.5 (3.5)	39.7 (3.6)	$t = -0.256$	.798
- extraversion	45.7 (3.8)	46.0 (4.8)	$t = -0.589$	.557
- openness	46.5 (5.1)	46.9 (4.1)	$t = -0.651$	.516
- agreeableness	38.0 (4.2)	37.4 (4.0)	$t = 1.074$	.284
- conscientiousness	61.3 (5.8)	60.7 (5.6)	$t = 0.686$	.493
Functioning, mean (SD)				
GAF Disability, highest lifetime score	79.7 (7.8)	82.0 (7.2)	$t = -2.406$	.017*
GAF Disability, highest score in past year	68.9 (12.5)	71.7 (13.4)	$t = -1.711$	.088
GAF Disability, highest score in past month	56.2 (14.1)	56.7 (14.7)	$t = -0.303$	.762
FROGS	65.4 (14.1)	69.5 (14.5)	$t = -2.262$	.025*
Environment, mean (SD)				
CTQ	33.4 (6.1)	33.9 (6.9)	$t = -0.551$	.582
RSA	137.6 (28.4)	144.3 (25.7)	$t = -1.861$	.064
Neurocognition, mean (SD)				
ROCF, sum of correct elements				
- T0	34.3 (2.6)	33.9 (2.8)	$t = 0.906$	.366
- immediate recall	22.4 (6.8)	23.8 (5.8)	$t = -1.651$	.100
- delayed recall	21.9 (6.6)	23.6 (5.8)	$t = 2.155$	.032*
ROCF, time to completion				
-T0	279.7 (152.0)	246.9 (112.5)	$t = 1.891$	.060
- immediate recall	199.3 (104.6)	190.9 (100.5)	$t = 0.626$	.532
- delayed recall	133.0 (82.4)	127.9 (82.3)	$t = 0.472$	.637

Table 3.1: Analysis of sociodemographic variables and behavioral assessments. *Note:* *significant at $p < .05$

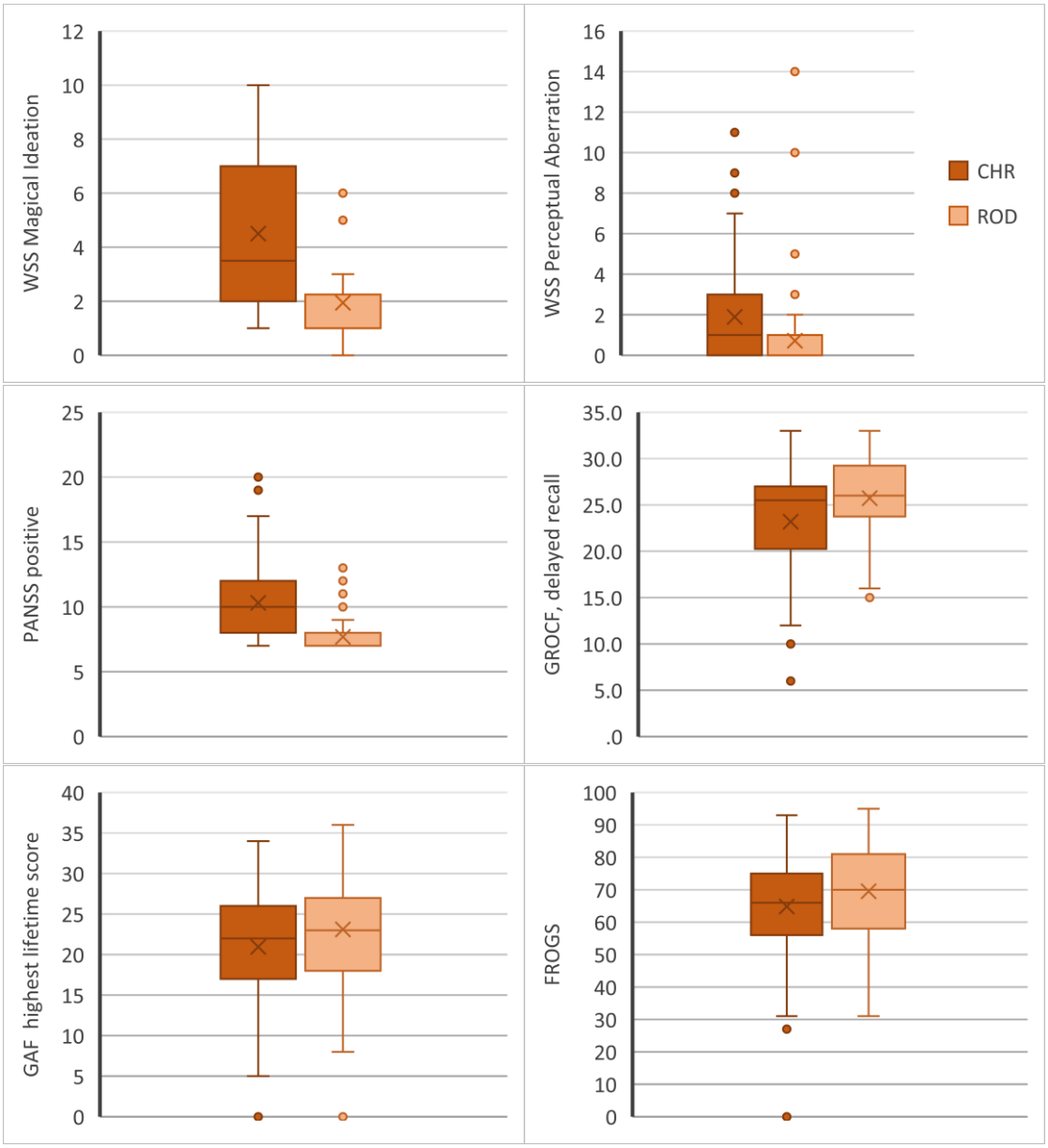

Figure 3.1: Box plots presenting the symptom severity in individuals with CHR vs. ROD. Shown are the questionnaires and one task of the neurocognitive assessment in which interactions with label were found. Higher scores in PANSS positive, magical ideation and perceptual aberration define a more severe state whereas higher scores in FROGS and ROCF describe a better functioning.

	LMU	UBS	Milan	UKK	BHAM	Turku	Udine	*F*	*P* Value
Psychopathological symptoms, mean (SD)									
PANSS total	52.7 (12.8)	45.3 (10.2)	57.6 (14.4)	55.2 (14.7)	50.2 (15.4)	47.7 (12.0)	45.9 (9.7)	1.74	.118
- positive sum	10.4 (2.5)	10.3 (3.4)	11.6 (2.8)	11.3 (3.4)	10.6 (3.5)	9.3 (3.2)	8.6 (1.5)	1.54	.172
- negative sum	14.0 (6.2)	10.4 (4.5)	15.1 (6.8)	13.6 (7.3)	10.5 (5.7)	11.0 (4.2)	11.8 (3.0)	1.59	.157
- general sum	28.2 (6.7)	24.5 (5.2)	30.9 (8.4)	30.3 (6.3)	29.1 (8.1)	27.4 (7.2)	25.5 (7.1)	1.62	.147
BDI-II	31.6 (10.6)	18.4 (10.5)	24.0 (10.7)	26.4 (13.8)	26.7 (12.3)	21.3 (10.4)	19.9 (10.6)	4.46	<.001*[1]
Personality, mean (SD)									
WSS total	22.5 (7.0)	20.1 (4.0)	19.1 (6.1)	23.1 (5.8)	20.8 (2.9)	19.3 (6.7)	19.5 (2.9)	1.26	.282
- magical ideation	3.7 (2.8)	3.3 (2.4)	2.3 (1.8)	3.6 (1.9)	3.0 (1.8)	3.8 (2.7)	3.2 (1.0)	1.00	.430
- perceptual aberration	2.5 (2.7)	1.7 (2.7)	1.0 (1.5)	2.8 (3.0)	1.2 (1.4)	1.3 (1.1)	0.8 (1.4)	1.82	.101
- physical anhedonia	8.9 (1.9)	8.9 (1.4)	8.7 (2.8)	9.3 (1.5)	9.6 (1.7)	8.6 (2.6)	9.7 (1.8)	0.72	.637
- social anhedonia	7.4 (2.3)	6.9 (2.1)	7.1 (1.3)	7.2 (1.9)	6.9 (1.6)	6.1 (3.1)	6.8 (1.9)	0.58	.745
NEO-FFI									
- neuroticism	40.6 (3.1)	39.2 (4.5)	40.6 (4.2)	39.3 (2.9)	38.9 (3.4)	37.3 (2.7)	39.1 (3.6)	1.71	.125
- extraversion	45.8 (3.5)	44.8 (3.7)	44.0 (2.2)	45.2 (3.5)	47.5 (4.3)	44.3 (4.5)	47.8 (3.8)	1.75	.116
- openness	46.8 (4.8)	47.4 (3.3)	45.3 (4.5)	44.4 (6.4)	44.4 (4.5)	46.5 (6.1)	50.1 (3.6)	2.11	.058
- agreeableness	57.3 (3.4)	56.6 (6.1)	54.7 (2.4)	56.8 (4.3)	57.3 (4.5)	53.5 (3.5)	56.9 (3.7)	1.55	.170
- conscientiousness	62.5 (5.6)	63.3 (5.0)	59.7 (5.4)	59.6 (6.5)	61.8 (4.7)	57.8 (4.2)	61.4 (7.7)	1.75	.116
Functioning, mean (SD)									
GAF Disability, highest lifetime score	81.1 (6.3)	81.4 (5.5)	76.9 (12.5)	74.0 (9.9)	84.1 (4.3)	81.4 (5.3)	76.8 (8.7)	3.67	.002*[1]
GAF Disability, highest score in past year	71.0 (10.9)	71.1 (11.2)	50.7 (11.4)	62.9 (13.6)	76.2 (10.4)	66.9 (8.4)	73.4 (11.3)	5.63	<.001*
GAF Disability, highest score in past month	51.0 (12.7)	57.6 (11.5)	50.7 (11.4)	52.5 (11.4)	73.8 (12.5)	56.5 (11.8)	60.1 (16.6)	6.06	<.001*
FROGS	58.8 (13.4)	66.4 (11.8)	59.9 (16.8)	63.3 (14.6)	72.9 (12.7)	75.0 (11.6)	73.7 (9.1)	4.57	<.001*[1]
Environment, mean (SD)									
CTQ	33.2 (4.9)	35.6 (4.6)	31.0 (5.5)	34.9 (7.2)	34.4 (4.1)	33.1 (6.7)	29.7 (9.6)	1.44	.205
RSA	132.6 (27.4)	146.0 (32.2)	143.7 (38.4)	140.8 (28.6)	143.5 (34.4)	142.3 (24.1)	123.8 (15.0)	0.98	.442
Neurocognition, mean (SD)									
ROCF, sum of correct elements									
- T0	33.1 (3.1)	34.8 (2.0)	32.7 (2.0)	35.3 (1.2)	35.5 (1.0)	35.6 (1.3)	34.1 (3.4)	3.96	.001*[1]
- immediate recall	21.9 (6.0)	22.1 (9.0)	21.4 (4.6)	22.6 (6.2)	22.3 (7.1)	27.0 (2.8)	20.5 (9.5)	1.16	.331
- delayed recall	20.6 (6.0)	22.5 (7.8)	20.7 (3.8)	23.3 (5.4)	21.5 (7.7)	26.3 (3.7)	20.0 (9.1)	1.55	.170
ROCF, time to completion									
- T0	291.5 (151.5)	303.5 (254.8)	251.1 (77.9)	324.7 (116.1)	233.2 (136.7)	251.0 (141.7)	226.5 (77.1)	0.88	.513
- immediate recall	208.8 (109.2)	214.6 (128.6)	215.7 (126.8)	207.3 (75.0)	123.9 (63.5)	230.6 (133.8)	173.2 (64.2)	1.51	.183
- delayed recall	139.5 (87.9)	153.6 (111.8)	154.5 (96.1)	123.8 (55.0)	94.4 (40.1)	159.5 (109.9)	102.3 (28.2)	1.20	.312

Table 3.2: Results of the behavioral assessments of CHR subjects for Munich (LMU), Basel (UBS), Milan (Milan), Cologne (UKK), Birmingham (BHAM), Turku (Turku) and Udine (Udine) separately; a one-way ANOVA was conducted for each assessment tool to test for significant differences between the sites. *Note:* significant at $p < .05$, [1]no longer significant after correcting for multiple comparisons using the Bonferroni-Holms procedure.

	LMU	UBS	Milan	UKK	BHAM	Turku	Udine	*F*	*P* Value
Psychopathological symptoms, mean (SD)									
PANSS total	48.5 (8.7)	58.5 (13.7)	44.9 (8.8)	48.6 (11.7)	40.5 (6.1)	40.8 (4.8)	43.2 (9.9)	5.95	<.001*
- positive sum	7.3 (0.8)	8.9 (2.1)	7.4 (0.8)	7.7 (1.4)	7.8 (0.9)	7.5 (0.8)	7.5 (0.7)	3.68	.002*
- negative sum	13.6 (4.6)	16.6 (4.8)	9.0 (3.3)	13.1 (5.4)	8.6 (3.0)	9.8 (2.1)	10.8 (4.7)	6.18	<.001*
- general sum	27.6 (5.7)	33.1 (8.7)	28.4 (8.0)	27.8 (7.0)	24.1 (3.7)	23.5 (3.4)	24.9 (5.6)	3.90	.001*
BDI-II	30.5 (12.4)	28.4 (11.4)	17.0 (16.9)	25.9 (14.6)	32.1(10.7)	16.9 (9.2)	18.0 (13.4)	3.76	.002*[1]
Personality, mean (SD)									
WSS total	17.6 (4.2)	19.8 (7.3)	20.1 (4.7)	17.2 (5.0)	19.9 (3.1)	16.9 (4.5)	15.6 (2.5)	1.65	.139
- magical ideation	1.6 (1.3)	2.6 (2.6)	2.7 (1.7)	1.8 (1.4)	2.2 (1.4)	1.7 (1.2)	1.2 (0.7)	1.68	.131
- perceptual aberration	0.6 (2.2)	1.3 (2.6)	2.0 (2.3)	0.6 (1.3)	0.3 (0.8)	0.7 (1.1)	0.4 (0.9)	1.12	.354
- physical anhedonia	8.6 (2.4)	8.9 (1.7)	9.1 (1.7)	8.4 (2.9)	9.3 (1.9)	7.8 (2.2)	8.0 (2.0)	0.72	.637
- social anhedonia	6.8 (1.5)	6.9 (2.5)	6.3 (2.4)	6.4 (1.6)	8.2 (1.5)	6.6 (1.1)	5.8 (1.6)	2.56	.023*[1]
NEO-FFI									
- neuroticism	40.1 (4.3)	39.0 (2.4)	42.2 (2.4)	39.1 (4.0)	39.5 (2.1)	38.9 (4.7)	39.5 (2.5)	0.82	.557
- extraversion	46.6 (4.5)	46.2 (5.2)	48.1 (4.6)	45.0 (5.2)	46.1 (5.6)	42.4 (3.8)	47.0 (3.2)	1.63	.144
- openness	47.8 (4.7)	47.8 (3.2)	48.4 (4.6)	46.6 (4.5)	46.0 (4.3)	44.3 (5.0)	46.0 (3.8)	1.29	.266
- agreeableness	56.5 (5.0)	56.5 (4.0)	60.3 (5.4)	56.8 (4.4)	56.0 (4.7)	53.0 (5.4)	56.7 (3.7)	1.73	.121
- conscientiousness	62.4 (4.7)	62.0 (3.4)	60.1 (5.1)	61.6 (5.8)	60.4 (5.5)	52.4 (4.7)	59.5 (6.5)	5.64	<.001*
Functioning, mean (SD)									
GAF Disability, highest lifetime score	80.0 (6.2)	88.7 (6.6)	80.1 (9.0)	79.4 (6.9)	85.6 (5.5)	85.3 (4.6)	79.6 (7.3)	5.70	<.001*
GAF Disability, highest score in past year	70.1 (10.8)	74.9 (23.3)	73.4 (9.8)	68.6 (10.9)	73.4 (12.0)	77.5 (8.7)	71.8 (15.5)	0.83	.551
GAF Disability, highest score in past month	52.0 (12.0)	49.6 (21.8)	68.4 (11.5)	56.9 (11.0)	69.2 (11.8)	57.9 (9.3)	60.0 (15.3)	4.59	<.001*
FROGS	64.9 (14.0)	64.5 (18.0)	79.0 (10.6)	68.0 (11.8)	75.9 (10.7)	79.7 (11.5)	74.4 (15.8)	3.47	.003*[1]
Environment, mean (SD)									
CTQ	34.6 (5.7)	35.0 (7.4)	31.3 (4.2)	33.7 (8.6)	35.3 (8.3)	32.0 (5.5)	31.6 (8.3)	0.67	0.672
RSA	138.6 (22.2)	154.4 (35.6)	157.5 (27.6)	142.7 (20.2)	144.8 (20.6)	141.4 (21.3)	149.4 (36.4)	1.12	0.353
Neurocognition, mean (SD)									
ROCF, sum of correct elements									
- T0	32.5 (3.9)	33.4 (2.4)	35.0 (1.8)	34.7 (1.3)	35.0 (1.4)	35.7 (0.7)	35.0 (1.0)	4.33	<.001*[1]
- immediate recall	22.3 (5.0)	21.3 (6.9)	24.1 (4.4)	25.3 (4.8)	25.3 (7.9)	26.4 (4.3)	25.3 (6.1)	1.97	.076
- delayed recall	21.9 (5.3)	21.7 (6.9)	22.6 (5.5)	25.4 (5.1)	25.6 (6.3)	26.5 (5.1)	24.7 (5.6)	2.17	.050
ROCF, time to completion									
- T0	255.6 (93.7)	296.7 (183.4)	223.3 (57.8)	234.1 (95.5)	205.4 (141.6)	252.9 (88.2)	229.4 (86.7)	1.00	.431
- immediate recall	180.8 (79.4)	248.2 (165.8)	138.3 (62.8)	201.9 (97.6)	132.3 (44.6)	233.6 (105.0)	183.9 (75.1)	2.43	.030*[1]
- delayed recall	117.5 (52.7)	188.8 (181.8)	90.0 (41.3)	129.3 (54.3)	102.2 (33.7)	143.6 (55.3)	118.7 (38.1)	2.22	.046*[1]

Table 3.3: Results of the behavioral assessments of ROD subjects for Munich (LMU), Basel (UBS), Milan (Milan), Cologne (UKK), Birmingham (BHAM), Turku (Turku) and Udine (Udine) separately; a one-way ANOVA was conducted for each assessment tool to test for significant differences between the sites. *Note:* significant at $p < .05$, [1]no longer significant after correcting for multiple comparisons using the Bonferroni-Holms procedure.

3.2 Classification results

Analysis 1: Overall clinical classifier without an additional feature selection method

For this classification algorithm, a BAC of 73.5% was achieved, with 85 correctly identified CHR and 96 correctly classified ROD patients (sensitivity: 72.0%, specificity: 75.0%, *p=.015*). Detailed statistics of the classification model are reported in Table 3.4.
The ten most discriminative features for the classification of CHR vs. ROD patients were: WSS 04, WSS 25, WSS 24, NEO-FFI 28, WSS 20, NEO-FFI 41, WSS 36, WSS 10, FROGS 06 and BDI-II 15 (see Figure 3.2 and Table 3.5).

Eight of these items belong to the personality domain and are part of either WSS or NEO-FFI. More precisely, six items belong to the WSS and two to the NEO-FFI. Assigning the WSS items to the different scales, three of them belong to the physical anhedonia scale (WSS 04, WSS 20, WSS 10) and one item each to the magical ideation scale (WSS 24), the perceptual aberration scale (WSS 25) and the social anhedonia scale (WSS 36). If one considers the single WSS items as representative for the scales, CHR subjects scored particularly high on the magical ideation (WSS 24; CHR: *M=0.25, SD= 0.44,* ROD*: M=0.05, SD=0.21; T=0.546, p<.001*) and the physical anhedonia scale (WSS 10, CHR: *M=0.86, SD=0.35,* ROD: *M=0.73, SD=0.44; T=-2.522, p=.012*) whereas ROD subjects scored high on the perceptual aberration (WSS 25, CHR: *M=0.09*, *SD=0.86,* ROD: *M= 0.05, SD= 0.73; T=-1.158, p=.248*), the social anhedonia (WSS 36, CHR: *M=0.75, SD=0.43,* ROD: *M=0.88, SD=0.33; T=2.502, p=.013*) and the physical anhedonia scale (WSS 04, CHR: *M=0.80, SD=0.40,* ROD: *M=0.81, SD=0.39; T=0.28, p=.778*; WSS 20, CHR: *M=0.22, SD=0.42*, ROD: *M=0.25, SD=0.43; T=0.586, p=.586*).
The NEO-FFI as part of the personality domain occurred beside the WSS among the most relevant features. NEO-FFI 28 is part of the openness domain, NEO-FFI 41 belongs to the neuroticism domain. ROD individuals scored high on both items (NEO 28, CHR: *M=3.22, SD=1.25,* ROD: *M=3.20, SD=1.14; T=-0.114, p=.909*; NEO-FFI 41, CHR: *M=3.30, SD=1.14,* ROD: *M=3.50, SD=1.05; T=1.436, p=.152*).

Beside the personality domain, the psychopathological symptoms domain also appeared among the ten most informative items. CHR subjects scored high on one item of the BDI-II (BDI-II 15, CHR: *M=1.33, SD=0.86, ROD: M=1.35, SD=0.78; T=0.202, p=.840*).

Finally, among the ten most discriminative items, one item of the FROGS occurred and revealed high scores in the CHR group: FROGS 06 (CHR: *M=3.14, SD=1.37*; ROD: *M=3.12, SD=1.35; T=-0.106, p=.916*).

Algorithm	Preprocessing Pipeline	TP	TN	FP	FN	Sens [%]	Spec [%]	BAC [%]	FPR [%]	PPV [%]	NPV [%]	DOR	NND
LIBLINEAR L1-regularized L2-loss SVC	Standardization Imputation of missing values	85	96	32	33	72.0	75.0	73.5	25.0	72.6	74.4	8.3	2.1

Table 3.4: Cross-validation performance for the classification using a sparse algorithm. In preprocessing, the data was standardized by subtracting the mean from the feature's absolute value and dividing it by the standard deviation; missing values were imputed using the k-NN algorithm. *Note:* SCV = support vector classification.

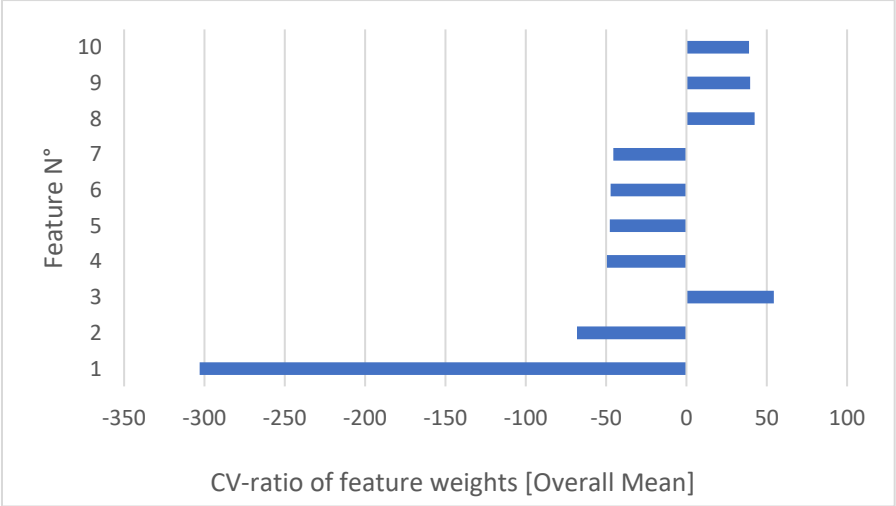

Figure 3.2: Feature weights of the ten most discriminative items for the classification CHR vs. ROD. The higher the value of the item, the more important for differentiating the two groups. Positive values are associated with higher scores in our first group (CHR) whereas negative values show higher scores in our second group (ROD). Feature number corresponding to Table 3.5.

N°	Item name	Description	Domain
1	WSS 04	I have often found walks to be relaxing and enjoyable.	Physical Anhedonia
2	WSS 25	Sometimes I have had a passing thought that some part of my body was rotting away.	Perceptual Aberration
3	WSS 24	I have sometimes felt that strangers were reading my mind.	Magical Ideation
4	NEO-FFI 28	I often try new and foreign foods.	Openness
5	WSS 20	The beauty of sunsets is greatly overrated.	Physical Anhedonia
6	NEO-FFI 41	Too often, when things go wrong, I get discouraged and feel like giving up.	Neuroticism
7	WSS 36	Knowing that I have friends who care about me gives me a sense of security.	Social Anhedonia
8	WSS 10	The sound of the rain falling on the roof has made me feel snug and secure.	Physical Anhedonia
9	FROGS 06	Participation in activities (e.g. sports, reading), organization of spare time	Negative symptom dimension (personal activities)
10	BDI-II 15	Loss of energy	Depressive symptoms

Table 3.5: The ten most discriminative features for the classification CHR vs. ROD. Shown are the contents of the features and the domain they are generally assessing. N°: Feature number corresponding to Figure 3.2.

Analysis 2: Soft feature selection

After implementing a soft feature selection method, a BAC of 74.3% was achieved, with 86 correctly identified CHR and 97 correctly classified ROD patients (sensitivity: 72.9%, specificity: 75.8%, *p=.027*). Detailed statistics can be found in Table 3.6; Figure 3.4 additionally shows the decision scores for every single participant in this analysis.

When visualizing the data and looking at the CV-ratio of feature weights, the ten most discriminative items were: PANSS P3, PANSS G4, ROCF (delayed recall, sum of correct elements), WSS 24, PANSS P2, PANSS G6, FROGS 11, NEO-FFI 43, ROCF (time to completion, T0) and NEO-FFI 50 (see Figure 3.3 and Table 3.7).

The personality domain is now represented by three items: WSS 24 (magical ideation), NEO-FFI 43 (openness) and NEO-FFI 50 (conscientiousness). CHR subjects scored continually high on WSS 24. Regarding the NEO-FFI, ROD subjects scored high on NEO-FFI 50 (CHR: *M=3.13, SD=1.13,* ROD: *M=3.55, SD=0.92; T=3.219, p=.002*) as part of the conscientiousness domain, whereas CHR achieved a higher score on NEO-FFI 43 (CHR: *M=3.14, SD=1.31*, ROD: *M=2.73, SD=1.21; T=-2.437, p=.016*) as part of the openness dimension.

In this analysis, the psychopathological variables were stronger represented among the ten most informative features. The CHR subjects scored particularly high on the positive symptoms scale of the PANSS (P3=Hallucinatory behavior, CHR: *M=1.58, SD=0.97*, ROD: *M=1.05, SD=0.23; T=-5.891, p<.001;* P2=Conceptual disorganization, CHR: *M=1.42, SD=0.83*, ROD: *M=1.09, SD=0.33; T=-4.142, p<.001*) whereas ROD subjects scored high on the general items PANSS G4 (CHR: *M=1.72, SD=1.01,* ROD: *M=2.01, SD=1.17; T=2.072, p=.039*) which assesses tension, and PANSS G6 (CHR: *M=3.54, SD=1.44,* ROD: *M=4.34, SD=1.3; T=3.972, p<.001*) which assesses depressive tendencies.

With respect to the neurocognitive domain, the ROCF (sum of correct elements for delayed recall, time to completion at timepoint T0) appears to be a relevant test for the classification CHR vs. ROD. In general, ROD subjects performed noticeably better compared to CHR subjects. Regarding the delayed recall, the sum of correct responses was significantly higher in the ROD group (CHR: *M=21.92, SD=6.55,* ROD: *M=23.63, SD=5.78; T=2.155, p=.032*) and the CHR subjects needed more time to complete the recognition task (CHR: *M=279.74, SD=151.98*, ROD: *M=246.89, SD=112.47; T=-1.891, p=.060*).

Beside personality traits, psychopathological symptoms and neurocognition, one item of the functional level domain showed up. FROGS 11, a criterion determining the *Relationships* factor, revealed high scores in the ROD group (CHR: *M=2.97, SD=1.37,* ROD: *M=3.40, SD=1.37; T=2.431, p=.016)*.

Algorithm	Preprocessing Pipeline	TP	TN	FP	FN	Sens [%]	Spec [%]	BAC [%]	FPR [%]	PPV [%]	NPV [%]	DOR	NND
LIBLINEAR L1-regularized L2-loss SVC	Standardization Imputation of missing values Soft feature selection	86	97	31	32	72.9	75.8	74.3	24.2	73.5	75.2	9.1	2.1

Table 3.6: Cross-validation performance after implementing a feature selection method to the preprocessing pipeline. The data was standardized by subtracting the mean from the feature's absolute value and dividing it by the standard deviation; missing values were imputed using the k-NN algorithm. *Note:* SCV = support vector classification.

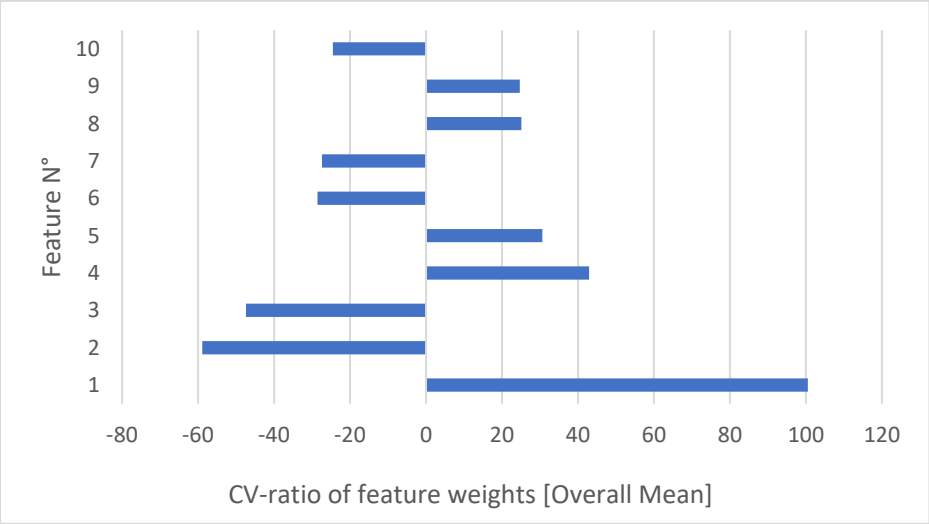

Figure 3.3: Feature weights of the ten most discriminative items in the classification CHR vs. ROD after implementing a feature selection method. The higher the value of the item, the more important for differentiating the two groups. Positive values are associated with higher scores in our first group (CHR) whereas negative values show higher scores in our second group (ROD). Feature number corresponding to Table 3.7.

Feature N°	Item name	Description	Domain
1	PANSS P3	Hallucinatory behavior	Positive symptoms
2	PANSS G4	Tension	General symptoms
3	ROCF- delayed recall	Neurocognition task	Visual memory
4	WSS 24	I have sometimes felt that strangers were reading my mind.	Magical Ideation
5	PANSS P2	Conceptual disorganization	Positive symptoms
6	PANSS G6	Depression	General symptoms
7	FROGS 11	Relationships necessary to maintain harmonious inclusion in society	Negative symptom dimension (Social functioning)
8	NEO-FFI 43	Sometimes when I am reading poetry or looking at a work of art, I feel a chill or wave of excitement.	Openness
9	ROCF- time to completion at T0	Neurocognition task	Visual memory
10	NEO-FFI 50	I am a productive person who always gets the job done.	Conscientiousness

Table 3.7: The ten most discriminative features for the classification CHR vs. ROD after implementing a soft feature selection method. Shown are the contents of the features and the domain they are generally assessing. N°: Feature number corresponding to Figure 3.3.

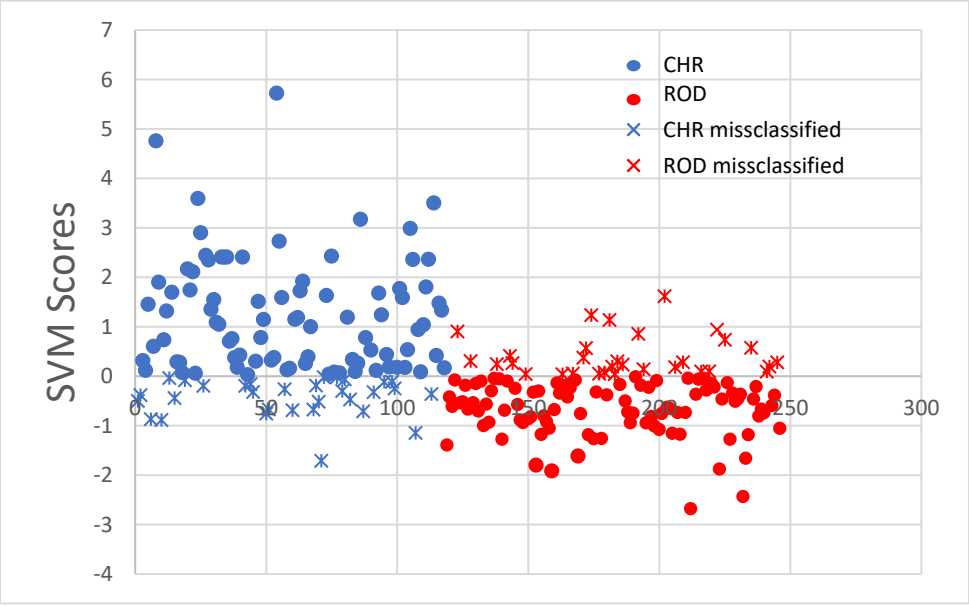

Figure 3.4: SVM classification with soft feature selection. Shown are the SVM scores for every single participant. The hyperplane determines which participant is classified correctly.

Analysis 3: Mono-clinical classifiers

The highest BAC of 71.0% (sensitivity: 71.8, specificity: 70.2%) was achieved in the psychopathological symptoms domain (BDI-II, PANSS), whereas the worst BAC of 47.0% (sensitivity: 44.4%, specificity: 49.6%) was achieved when classifying CHR and ROD based on environmental factors (CTQ, RSA) only (see Table 3.8).

We additionally wanted to evaluate which features seem to be the most important within each domain, particularly regarding personality traits, as the WSS seemed to be quite relevant for the classification CHR vs. ROD in the first analysis. With the personality classifier, items from the NEO-FFI and the WSS were more equally distributed compared to the first analysis. The two most relevant items included the WSS 24 and the NEO-FFI 41 which occurred in our first analysis as well.

Domains	TP	TN	FP	FN	Sens [%]	Spec [%]	BAC [%]	FPR [%]	PPV [%]	NPV [%]	DOR	NND
Personality (WSS, NEO-FFI)	68	86	42	50	**57.6**	**67.2**	**62.4**	32.8	61.8	63.2	3.1	4.0
Psychopathological symptoms (PANSS, BDI-II)	84	85	36	33	**71.8**	**70.2**	**71.0**	29.8	70.0	72.0	5.8	2.4
Functional level (GAF, FROGS)	57	58	70	61	**48.3**	**45.3**	**46.8**	54.7	44.9	48.7	0.8	15.7
Neurocognition	62	72	56	54	**53.4**	**56.2**	**54.8**	43.8	52.5	57.1	1.5	10.3
Environment (CTQ, RSA)	52	63	64	65	**44.4**	**49.6**	**47.0**	50.4	44.8	49.2	0.8	16.8

Table 3.8: Classification performance of the five clinical domains. *Note:* PANSS: Positive and Negative Syndrome Scale; BDI-II: Beck Depression Inventory-II; NEO-FFI: NEO Five-Factor Inventory; WSS: Wisconsin Schizotypy Scales (short form); GAF: Global Assessment of Functioning; FROGS: Functional Remission in General Schizophrenia; CTQ; Childhood Trauma Questionnaire, RSA: Resilience Scale for Adults.

Analysis 4: Multisite generalizability

To investigate multisite generalizability, we then applied a LOSOCV and yielded a significant overall BAC of 69.0% (sensitivity: 66.9%, specificity: 71.1%, *p=.007,* see Table 3.9). Compared to our first analysis which used the same preprocessing pipeline, there was a BAC loss of 4.5%, a loss of 5.1% for sensitivity and a loss of 3.9% of specificity.

Looking at the results of every single cycle in which one site was held out and used as validation

sample, it becomes apparent that the classification generalizability differs noticeably depending on the site the models are tested on. The highest BAC of 81.4% was achieved when testing the ensemble classifier of the inner cycle on the data of the unseen subjects of Cologne in the outer cycle. Comparatively, subjects from Milan or Udine are more likely to be misclassified with low accuracies of 50.0% and 59.0% respectively. The results for each individual cycle of the LOSOCV can be seen in Table 3.10.

Algorithm	Preprocessing Pipeline	TP	TN	FP	FN	Sens [%]	Spec [%]	BAC [%]	FPR [%]	PPV [%]	NPV [%]	DOR	NND
LIBLINEAR L1-regularized L2-loss SVC	Standardization Imputation of missing values	79	91	37	39	66.9	71.1	69.0	28.9	68.1	70.0	5.4	2.6

Table 3.9: Leave-one-site-out cross-validation performance. For the classification we used a sparse algorithm. In preprocessing, we standardized the data by subtracting the mean from the feature's absolute value and dividing it by the standard deviation; we imputed missing values using the k-NN algorithm. *Note*: SCV = support vector classification.

Site left out:	LMU	UBS	Milan	UKK	BHAM	Turku	Udine	Overall
Training BAC [%]	71.9	71.8	75.3	71.6	72.3	72.6	72.5	69.0%
Testing BAC [%]	68.6	71.9	50.0	81.4	63.5	75.8	59.0	
Sensitivity [%]	66.7	66.7	42.8	84.2	61.5	77.7	58.3	66.9
Specificity [%]	71.1	76.4	57.1	80.0	64.3	78.6	61.5	71.1

Table 3.10: Leave-one-site-out cross-validation: Results for each cycle. In each the cycle of the LOSOCV one site was held out as a testing fold while the data of the remaining six sites was used for training SVM models.

4 Discussion

Within this dissertation, we aimed to better understand how CHR and ROD differ across several clinical domains and how these domains contribute to a multivariate classification of the two groups. To achieve this, we performed a set of four analyses following three more specific aims: First, we wanted to evaluate the most predictive feature patterns in an overall clinical classification. When training a classifier with pooled information from all five mono-clinical domains as feature set, we achieved a BAC of 73.5% which means that CHR and ROD were well separable. Among the ten most predictive features, the majority was from the personality domain (WSS and NEO-FFI items). This indicates that personality seems to be important in terms of differential diagnosis and that there might be distinctive personality traits differentiating the two groups. This finding was supported by a second analysis we conducted to better separate out the informative variables by adding a feature selection method to the preprocessing pipeline. This classifier correctly discriminated patients with CHR for psychosis from ROD individuals with a cross-validated BAC of 74.4 % and personality remained an important domain.

Within our second aim, we decided to analyze each of the five clinical domains separately to compare sensitivity, specificity and BAC. The mono-clinical classifiers revealed that psychopathological symptoms could differentiate between CHR and ROD with a BAC of 71.0%, while the personality domain yielded a BAC of only 62.4%. This demonstrates that the overall pattern is predictive, not single items which is an important characteristic of MVPA.

Finally, we tested the geographic generalizability of the predictive pattern across seven European PRONIA sites. We performed a LOSOCV and achieved a BAC of 69.0%. The classification CHR vs. ROD seems, therefore, to be reasonably stable considering geographical and cultural aspects as well as different sample sizes.

4.2 The most predictive feature patterns in the overall clinical analyses

Within the first aim of this study, we looked at the ten most discriminative features of both of our overall clinical analyses and found that the personality domain might be of particular importance for differential diagnosis. In our first analysis, six of the most relevant features are part of the WSS, two of them belong to the NEO-FFI. Therefore, the WSS seemed to be the most informative regarding classification of CHR vs. ROD. This constitutes an important finding since psychopathological symptoms assessed by PANSS were also part of the feature set.
When we added a soft feature selection method to our preprocessing pipeline, the ranking and weighting of the variables changed almost completely and only one item remained amongst the ten most discriminative features: "*I have sometimes felt that strangers were reading my mind"* which is part of the WSS. Despite the fact that the overall weighting of the feature set changed, personality traits still occurred among the most relevant items. This shows that differences in certain personality characteristics seem to be in fact stable and important regarding differential diagnosis.
All four scales of the WSS occurred among the most relevant features: magical ideation, perceptual aberration, social anhedonia and physical anhedonia. Flückiger and colleagues (2016) investigated the psychosis-predictive value of the WSS and its association with CHR states. They found that only physical anhedonia was associated with CHR for psychosis. However, in their analyses, they used the total scores for positive responses to each of the scales, thus fully representing the scales. Several other studies described SPD as a prodromal phase of schizophrenia and referred to the CHR criteria, UHR in particular, as closely linked to the positive symptoms of SPD (Debbane et al., 2015; Flückiger et al., 2016; Fonseca-Pedrero et al., 2011). The high scores for CHR on the magical ideation scale support these findings.
ROD individuals, on the other hand, scored high on an item of the perceptual aberration scale which belongs besides magical ideation to the positive factors of the WSS (Kwapil et al., 2008). Previous studies reported not only UHR for psychosis, but also mood disorders as significantly associated with positive schizotypy (Emsley et al., 1999; Lewandowski et al., 2006; Lysaker et al., 1995). Young adults identified by the perceptual aberration scale showed distinctly high rates of mood disorders such as depression (Kwapil et al., 2008). Another aspect is that one of the most common comorbid disorders of depression is the somatization disorder, which is a cluster of symptoms lasting over at least two years without any known overt medical cause. This results in additional symptoms in depressive patients such as hyperalgesia, muscular tension and somatic pain (Anderson et al., 2014; Shimodera et al., 2012).

Physical anhedonia is a negative symptom of SPD; it can be described as an absence of pleasure from physical or sensory experiences and is reported to be related to both, CHR and depressive symptoms (Jhung et al., 2016; Shankman et al., 2010). Research focusing on the association between depression and physical anhedonia described this state as a stable factor, as linked to severity of depressive symptoms and as potentially affecting social functioning in patients (Shankman et al., 2010). ROD individuals are reported to be less involved in personal activities including e.g. sports, reading or different kinds of hobbies (American Psychiatric Association, 2013; Shankman et al., 2010) which is presumably correlated with a decreased feeling of pleasure regarding physical or sensory experiences. The high scores of depressive individuals on two items of the physical anhedonia scale and the lower scores on the corresponding FROGS item (*Participation in activities*) are in line with previous findings.

Besides the WSS, the NEO-FFI also assesses personality traits and is represented among the most relevant items in both of our overall clinical analyses. The corresponding FFM dimensions that occurred are neuroticism, openness and conscientiousness.
Previous studies on personality in CHR states and MDD suggested that there is no characteristic profile for the respective diagnosis. Several studies mainly reported higher levels of neuroticism and lower levels of extraversion in CHR states and MDD (Klein et al., 2011; Schultze-Lutter et al., 2015). This presumably reflects a general psychopathological status as it indicates a lower personality-related resilience and is, therefore, associated with clinical disorders in general (Ruhrmann et al., 2010; Schultze-Lutter et al., 2015; Widiger, 2011). Nevertheless, the two groups might still differ regarding the degree to which a certain dimension is altered. For example, in our analyses, ROD individuals scored higher on ”*Too often, when things go wrong, I get discouraged and feel like giving up*“ which is part of the neuroticism dimension. Neuroticism has been described as a fundamentally important aspect regarding depressive disorders as it is known as a strong risk factor for the lifetime prevalence of MDD, and might predict onset and symptom severity of the disease (Hirschfeld et al., 1989; Xia et al., 2011). This correlation can also be seen vice versa as experiencing depressive symptoms yield in higher levels of neuroticism (Reich et al., 1987). However, single items can hardly be seen as diagnostically conclusive.

In our first analysis, eight of the ten most relevant features belong to the personality domain; in our second analysis on the other hand, psychopathological symptoms seemed to be equally relevant for the classification. Four of the ten most discriminative features belong to the

PANSS. The fact that psychopathological features differ between the two groups and that they can be used to discriminate between CHR and ROD constitutes a rather probable outcome; psychopathological symptoms such as depressive mood or hallucinatory behavior are among the defining criteria of the respective diagnosis (American Psychiatric Association, 2013; McGlashan et al., 2010). These items are therefore already intuitively more clearly assignable to one of the groups compared to personality traits such as high neuroticism or openness. This might raise the question if using the sparse algorithm plus a feature selection method may have removed too many features, while leaving only the ones that are most likely.
CHR subjects scored high on positive symptoms (conceptual disorganization and hallucinatory behavior) whereas ROD subjects received a higher sum on general symptoms (tension and depression). Disorganized communication and perceptual abnormalities/ hallucinations are PANSS items as well as UHR criteria for psychosis (Klosterkötter et al., 2016). This is in line with our CHR subjects describing these symptoms more often and as more severe compared to our ROD subjects. General symptoms on the other hand, dysphoric mood in particular, are an integral element of the definition of depression as already mentioned above (American Psychiatric Association, 2013). The higher scores on the general symptom tension could be explained by the fact that insomnia is one of the core symptoms of MDD and markedly effects emotional abilities; sleep disturbances have been consistently reported to be associated with subjective reports of irritability and emotional volatility (Goldstein & Walker, 2014; Krause et al., 2017). Another possible explanation would be the frequently occurring comorbidities of depressive disorders such as anxiety disorders (American Psychiatric Association, 2013).

With respect to the lower scores in CHR individuals on the FROGS item addressing social functioning, literature reveals that social factors are substantially involved in the pathogenesis and consequences of depression as well as of CHR for psychosis - In the prodromal period of psychosis, social impairment has been described as one of the key features preceding the onset of frank psychosis (Fusar-Poli et al., 2013; Jang et al., 2011). Some researchers even suggested that the degree of social deficits in UHR individuals does not differ from patients in the early stages of psychotic illness or even the more established disease (Addington et al., 2008). While social impairment is a core deficit of both CHR and ROD, our study suggests that patients in CHR states may have greater difficulties coping with the attenuated symptoms affecting social bonds (e.g. unstable ideas of reference or an increased emotional reactivity in response to social interactions) which possibly contributes to a decline in social functioning (Bechdolf et al., 2005; Schultze-Lutter et al., 2007).

In addition to personality, psychopathological symptoms and functional level, the neurocognitive domain is represented by two features. Previous research regarding cognitive domains in CHR individuals has reported several areas of neurocognitive deficits, including visuospatial ability and visual memory, assessed by the ROCF test (Fusar-Poli et al., 2012; Koutsouleris et al., 2012; Shin et al., 2006). Our study supports previous findings, as CHR individuals achieved lower scores on correct elements of the ROCF and needed more time to complete the task.

4.4 Predictive accuracy of the mono-clinical classifiers

We split the feature set in five domains comprising psychopathological symptoms, personality traits, neurocognitive performance, functional level as well as environmental and autobiographical aspects. Within the second aim of our study, we wanted to investigate how each of the clinical data domains performed by itself when differentiating CHR and ROD. The mono-clinical classifiers revealed that psychopathological symptoms showed the best performance with a BAC of 71.0%. This constitutes another interesting finding since we covered the psychopathological symptoms domain with PANSS and BDI-II only. Comparatively, the personality domain performed considerably worse than the overall feature set with an accuracy of only 62.4%. This supports the idea that other characteristics besides personality traits are important for the classification and that a good prediction performance is based on an overall pattern rather than a small combination of features.

As personality traits seemed to be relevant for the classification CHR vs. ROD in our overall clinical analyses, we additionally wanted to evaluate which items seem to be the most important within this particular domain. We found that the two most relevant items were WSS 24 and NEO-FFI 41 which occurred in our first analysis as well. This supports our finding that magical ideation, which also occurred in our second analysis, and neuroticism seem to be stable and important elements for the classification.

4.5 Multisite generalizability of the model

In order to test the multisite generalizability of our model, we performed a LOSOCV. The idea behind this analysis is that similar accuracies to our first two overall clinical analyses can be achieved if a model is trained on six of the seven sites and then tested on the held-out site. We yielded a high overall BAC of 69.0% which shows that our linear SVM model generalized well across seven sites distributed across Europe.

In this analysis, we additionally gathered information on the heterogeneity of the data between sites as we iteratively used each site as a validation sample. Accuracies for each individual site ranged from 50.0% in Milan to 81.4% in Cologne. An explanation that first comes to mind is the phenomenon of overfitting, in which the statistical models only reflect the peculiarities of the samples they are trained on (Kernbach & Staartjes, 2022). Another reason might be differences in the process of recruiting and testing the participants. However, in this case, a more obvious explanation would be the unequal sample sizes of the respective sites. The sample in Milan consists of only seven CHR and seven ROD subjects whereas the sample of Munich includes almost six times as many (38 CHR and 43 ROD subjects).

Despite the differences regarding each individual cycle, the overall high BAC of our LOSOCV approach suggests, that the classification CHR vs. ROD appears to be reasonably stable considering geographical and cultural aspects as well as different sample sizes.

4.6 Limitations and future directions

Finally, some limitations have to be considered when interpreting our results. In our univariate analyses, CHR and ROD subjects differed significantly in terms of age, with a significantly younger mean age among CHR subjects. We decided, however, not to correct for age in our analyses; several studies examining different aspects of both CHR and ROD, including ours, display a significantly younger mean age in CHR individuals (Haidl et al., 2020; Koutsouleris et al., 2020, 2015; Schultze-Lutter et al., 2007). Previous research indicates that most first psychotic episodes occur before age 24 (Jongsma et al., 2018; Kirkbride et al., 2012) implying that prodromal symptoms emerge at an even younger age. Depressive disorders also typically commence in late adolescence and early adulthood, however, show a wider distribution; a meaningful proportion of lifetime MDD starts in middle or even late adulthood (Kessler & Bromet,

2013). Thus, for our analyses, we considered age less of a nuisance variable but as a factor containing illness-related variance.

When investigating potential site effects with respect to the behavioral assessments within each group separately, we found a few significant differences between the seven sites. The sum scores of several questionnaires differed significantly between the sites. One reason for this could be the interrater reliability, which displays the independency of the results from the individual rater or, in other words, is a measure for objectivity. Another reason might be geographical or cultural aspects affecting the results, which is also taken into account by the LOSOCV we conducted to investigate the multisite generalizability. The different sites yielded accuracies ranging from 50.0% in Milan to 81.4% in Cologne. Even if considering these findings as limitations, it undeniably reflects daily clinical practice particularly regarding cultural diversity and different clinicians affecting objectivity to a certain degree.

Looking at the ten most discriminative features in both of our overall clinical analyses, we found that they changed almost completely after adding another preprocessing step to our pipeline. This raises the topic of algorithm stability. In this context, stability refers to the impact of small changes in the training set on the output of the system (Bousquet & Elisseeff, 2001). Machine learning algorithms are very sensitive to changes in preprocessing classification settings which highlights the importance of well-informed decision making in machine learning approaches (Bousquet & Elisseeff, 2001; Dwyer et al., 2018).
Several studies investigating stability of learning algorithms displayed differences depending on the specific type of classification algorithm (Bousquet & Elisseeff, 2001; Huan Xu et al., 2012). Huan Xu et al. (2012) focused on sparse algorithms such as L1-regularization, which we used in our analyses. The researchers defined sparse algorithms as not stable and stated that sparsity and stability might even contradict each other. In order to check our model for geographic generalizability and stability, we conducted a LOSOCV and achieved a high BAC (69.0%) which indicates that our model is reasonably stable and generalizes well. Also indicative of stability is the fact that personality traits remained quite relevant features across all our overall clinical analyses when focusing on domains rather than single items.

Classifying CHR and ROD in our analyses, we achieved a high BAC of 74.3% with the use of eight different questionnaires and an assessment of neurocognitive performance. When we looked at the ten most discriminative features specifically, items of WSS, NEO-FFI, PANSS,

FROGS and the neurocognitive test battery occurred. The question arose if it would be possible to build a new tool around those variables which would help to identify and differentiate ROD and CHR individuals even better and more effectively. Focusing on building a brief assessment tool based on the ten most relevant features, there are some limitations to consider. First of all, a model based on only ten features might be too condensed and restricted; it might be better to opt for the top 10% or even 20% of features.

Another limitation, which goes in a similar direction, is that we can hardly assume those single items to be fully representative for the respective questionnaires. The way these questionnaires work is that they generate sum scores which are related to different levels of impairment. The fact that mostly ROD individuals experience for example something outside their bodies as part of their body does not necessarily mean they generally score high on the perceptual aberration scale. This approach would be too limited and unstable. Regarding this, we plan to conduct a more thorough approach in the future and implement all items of the most relevant questionnaires, namely WSS, NEO-FFI, PANSS, FROGS as well as the neurocognitive assessment. Our first step would be to build a classifier for each questionnaire and the neurocognitive test battery separately. In a second step, the most relevant variables of each questionnaire would be identified, and different weights will be assigned to them. Aiming for the main goal to build a tool separating CHR and ROD in an improved and more effective way, we intend for the algorithm to additionally take the different item weights into account when classifying the two groups. External validation on an entirely different study sample would also be a milestone to achieve in the future.

If one imagines differential diagnosis in a clinical setting, one could argue that it is particularly difficult to distinguish between ROD patients and ARMS-E individuals who meet only two COGDIS criteria. Most individuals in CHR states would also meet the criteria for a comorbid mental disorder, depression in particular (Addington et al., 2017; Fusar-Poli et al., 2017). This CHR subgroup might show depressive symptoms and additionally suffer from slight disturbances in expressive speech and sudden loss of train of thought several times a week (COGDIS; Schultze-Lutter et al., 2012). Dividing the CHR subjects in subgroups of different symptom severity might counteract the effects of heterogeneity in terms of symptom severity and improve classification of CHR vs. ROD.

Furthermore, instead of focusing on only two different groups, future studies could follow a broader identification approach. In terms of CHR for psychosis, several research groups recently highlighted the importance of a more dynamic and transdiagnostic strategy (Fusar-Poli

et al., 2017; Lee et al., 2018; McGorry et al., 2018). McGorry et al. (2018) investigated the ARMS for psychosis and challenged different aspects of the current concept. They focused on a transdiagnostic risk state and suggested a pluripotent model including several subthreshold states besides attenuated psychotic symptoms as well as any disorder outcome such as depressive, bipolar or personality disorder.

Based on this adapted "Clinical High At Risk Mental State" (CHARMS) our anticipated tool could be extended to other subthreshold states such as subthreshold bipolar states and borderline personality features of reduced range. The first step would be to compare each of those other "at-risk" states to our CHR group within our five domains at baseline. In a second step, the most relevant features would be used to build a classifier in the same manner as we plan to do for differential diagnosis of CHR for psychosis and ROD.

Instead of looking at baseline data only, repeated assessments would be needed to investigate the stability of those symptom patterns as well as how some symptoms might attract others. This potentially guides the identification of characteristic dynamics in the development of particular disorders (McGorry et al., 2018; Nelson et al., 2017).

4.7 Conclusion

In summary, we were able to distinguish CHR from ROD individuals with a high BAC of 74.3% using clinical measures: self-rating questionnaires (WSS, NEO-FFI, BDI-II, CTQ, RSA), observer-rated scales (PANSS, GAF, FROGS) and a neurocognitive test battery. In the overall clinical analyses, mainly items from the personality and psychopathological symptoms domain comprised the most predictive features, with WSS, NEO-FFI and PANSS as the most relevant questionnaires. Comparatively, the mono-clinical classifiers revealed that psychopathological symptoms, which we covered with PANSS and BDI-II only, showed the best performance with a BAC of 71.0%; the personality domain achieved a BAC of 62.4%. Within our aim to test the multisite generalizability of our predictive pattern, we conducted a LOSOCV and yielded a BAC of 69.0%. Thus, our model seemed to be considerably stable and generalizable across seven PRONIA sites in Europe. The characterizing items for each, CHR and ROD separately, were validated by previous research.

Considering the fact that the two groups can be difficult to differentiate depending on the present symptoms, a brief clinical assessment tool separating CHR from ROD more effectively constitutes a promising approach. Future studies could implement other subthreshold states of mental disorders and compare them to CHR for psychosis to gain a better understanding of early stages of mental ill health in general. In this regard, our study might provide further insights.

5 Appendix

Screening Assessments
1. General Data
2. Reasons for Referral
3. Somatic state and health history
4. CHR Assessment Tool I
4.1. Schizophrenia Proneness Instrument - Adult Version Cognitive Disturbance (SPI-A COGDIS)
4.2. Structured Interview for Prodromal Syndromes, P Items (SIPS P)
4.3. CAARMS Items (CAARMS)
4.4. Ultra-High Risk Criteria (Schizotypy, Genetic Risk)
4.5. Global Assessment of Functioning (GAF)
5. Clinical High Risk Criteria
5.1. Clinical High Risk Criteria (CHR Criteria)
5.2. SIPS CHR Intake Criteria (SIPS Intake Criteria)
5.3. CAARMS CHR Criteria (CAARMS CHR Criteria)
6. Treatment Documentation
7. Structured Clinical Interview for DSM IV 1 Screening (SCID-1 Screening)
8. Structured Clinical Interview for DSM IV 1 Summary (SCID-1 Summary)
9. Inclusion/Exclusion Criteria (IC/EC)
Baseline Assessments
10. Demographic and Biographic Data
11. Premorbid Adjustment Scale (PAS)
12. CHR Assessment Tool II
12.1. Schizophrenia Proneness Instrument Adult Version (SPI-A)
12.2. Structured Interview for Prodromal Syndromes N-G Items (SIPS N-G)
13. Positive and Negative Syndrome Scale (PANSS)
14. Scale for the Assessment of Negative Symptoms (SANS)
15. Chart of Life Events (CoLE)
16. Functional Remission in General Schizophrenia (FROGS)
17. Global Functioning Social / Role (GF S/R)
18. Prognostic Evaluation
Self-rated Assessments
1. WHO Quality Of Life - Short Version (WHOQOL-BREF)
2. Multidimensional Scale of Perceived Social Support (MSPSS)
3. Resilience Scale for Adults (RSA)
4. Coping Inventory for Stressful Situations (CISS-24)
5. Social Phobia Inventory (SPIN)
6. Beck Depression Inventory II (BDI-II)
7. Edinburgh Handedness Inventory Short Version (EHI-SV)
8. Level of Expressed Emotion Scale (LEE)
9. Wisconsin Schizotypy Scales - Short Form (WSS)
10. The Everyday Discrimination Scale T0 (EDS T0)
11. Bullying Scale T0 (BS T0)
12. Childhood Trauma Questionnaire (CTQ)
13. NEO Five-Factor Inventory (NEO-FFI)

Table A. 1: PRONIA complete list of observer and self-rated questionnaires. The observer-rated assessments are split up into two sessions of screening and baseline, while participants of the study are given the self-rated assessments to complete themselves.

Magical Ideation
1. I have felt that there were messages for me in the way things were arranged, like in a store window.
8. I have occasionally had the silly feeing that a TV radio broadcaster knew I was listening to him.
11. I have noticed sounds on my records that are not there at other times.
14. I have had the momentary feeling that someone's place has been taken by a look-a-like.
17. At times I perform certain little rituals to ward off negative influences.
24. I have sometimes felt that strangers were reading my mind.
27. If reincarnation were true, it would explain some unusual experiences I have had.
30. I have sometimes had the passing thought that strangers are in love with me.
33. The hand motions that strangers make seem to influence me at times.
40. I have sometimes been fearful of stepping on sidewalk cracks.
43. Numbers like 13 and 7 have no special powers.
46. I have had the momentary feeling that I might not be human.
49. I think I could learn to read others' minds if I wanted to.
53. Horoscopes are right too often for it to be a coincidence.
60. I have worried that people on other planets may be influencing what happens on Earth.
Perceptual Aberration
2. Occasionally it has seemed as if my body had taken on the appearance of another person's body.
5. I have sometimes felt confused as to whether my body was really my own.
12. I have sometimes had the feeling that my body is decaying inside.
15. Sometimes I have felt that I could not distinguish my body from other objects around me.
19. I have felt that something outside my body was a part of my body.
22. Sometimes I have had feelings that I am united with an object near me.
25. Sometimes I have had a passing thought that some part of my body was rotting away.
32. I have sometimes felt that some part of my body no longer belongs to me.
35. I can remember when it seemed as though one of my limbs took on an unusual shape.
38. I sometimes have to touch myself to make sure I'm still there.
41. I have sometimes had the feeling that one of my arms or legs is disconnected from the rest of my body.
48. I have had the momentary feeling that my body has become misshapen.
51. Sometimes I feel like everything around me is tilting.
56. Parts of my body occasionally seem dead or unreal.
59. At times I have wondered if my body was really my own.
Social Anhedonia
3. Having close friends is not as important as many people say.
6. I never had really close friends in high school.
9. I prefer watching television to going out with other people.
16. Just being with friends can make me feel really good.
18. I'm much too independent to really get involved with other people.
21. I prefer hobbies and leisure activities that do not involve other people.
28. I don't feel very close to my friends.
31. People who try to get to know me better usually give up after a while.
36. Knowing that I have friends who care about me gives me a sense of security.
39. People are usually better off if they stay aloof from emotional involvements with most others.
42. If given the choice, I would much rather be with others than be alone.
45. Although there are things that I enjoy doing by myself, I usually seem to have more fun when I do things with other people.
52. I feel pleased and gratified as I learn more and more about the emotional life of my friends.
54. When things are going really good for my close friends, it makes me feel good too.
57. Making new friends isn't worth the energy it takes.

(continued)

(continued)

Physical Anhedonia
4. I have often found walks to be relaxing and enjoyable.
7. A brisk walk has sometimes made me feel good all over.
10. The sound of the rain falling on the roof has made me feel snug and secure.
13. After a busy day, a slow walk has often felt relaxing.
20. The beauty of sunsets is greatly overrated.
23. The sound of rustling leaves has never much pleased me.
26. It has often felt good to massage my muscles when they are tired or sore.
29. Flowers aren't as beautiful as many people claim.
34. I like playing with and petting soft little kittens or puppies.
37. I don't understand why people enjoy looking at the stars at night.
44. When I'm feeling a little sad, singing has often made me feel happier.
47. Beautiful scenery has been a great delight to me.
50. The first winter snowfall has often looked pretty to me.
55. A good soap lather when I'm bathing has sometimes soothed and refreshed me.
58. Standing on a high place and looking out over the view is very exciting.

Table A. 2: The Wisconsin Schizotypy Scales. The WSS is used to assess positive and negative schizotypy in clinical and non-clinical samples. The questionnaire comprises four scales: magical ideation and perceptual aberration (positive factor) as well as social anhedonia and physical anhedonia (negative factor). In our study, we used the short form which comprises 15 items for each scale thus 60 items in total which can be answered with either "true" or "false". (Eckblad & Chapman, 1983; Gross et al., 2015)

Neuroticism
1. I am not a worrier.
6. I often feel inferior to others.
11. When I'm under a great deal of stress, sometimes I feel like I'm going to pieces.
16. I rarely feel lonely and blue.
21. I often feel tense and jittery.
26. Sometimes I feel completely worthless.
31. I rarely feel fearful or anxious.
36. I often get angry at the way people treat me.
41. Too often when things go wrong, I get discouraged and feel like giving up.
46. I am seldom sad or depressed.
51. I often feel helpless and want someone else to solve my problems.
56. At times I have been so ashamed I just wanted to hide.
Extraversion
2. I like to have a lot of people around me.
7. I laugh easily.
12. I don't consider myself especially 'light-hearted'.
17. I really enjoy talking to people.
22. I like to be where the action is.
27. I usually prefer to do things alone.
32. I often feel as if I'm bursting with energy.
37. I am a cheerful high-spirited person.
42. I am not a cheerful optimist.
47. My life is fast-paced.
52. I am a very active person.
57. I would rather go my own way than be a leader of others.
Openness
3. I don't like to waste time daydreaming.
8. Once I find the right way to do something, I stick to it.
13. I am intrigued by the patterns I find in art and nature.
18. I believe letting students hear controversial speakers can only confuse and mislead them.
23. Poetry has little or no effect on me.
28. I often try new and foreign foods.
33. I seldom notice the moods or feelings that different environments produce.
38. I believe we should look to our religious authorities for decisions on moral issues.
43. Sometimes when reading poetry or looking at a work of art, I feel a chill or wave of excitement.
48. I have little interest in speculating on the nature of the universe or the human condition.
53. I have a lot of intellectual curiosity.
58. I often enjoy playing with theory and abstract ideas.
Agreeableness
4. I try to be courteous to everyone I meet.
9. I often get into arguments with my family and co-workers.
14. Some people think I am selfish and egotistical.
19. I would rather co-operate with others than compete with them.
24. I tend to be cynical and skeptical of others' intentions.
29. I believe that most people will take advantage of you if you let them.
34. Most people I know like me.
39. Some people think of me as cold and calculating.
44. I am hard-headed and tough-minded in my attitudes.
49. I generally try to be thoughtful and considerate.
54. If I don't like people, I let them know it.
59. If necessary I am willing to manipulate people to get what I want.

(continued)

(continued)

Conscientiousness
5. I keep my belongings clean and neat.
10. I am pretty good about pacing myself so as to get things done on time.
15. I am not a very methodical person.
20. I try to perform all the tasks assigned to me conscientiously.
25. I have a clear set of goals and work towards them in an orderly fashion.
30. I waste a lot of time before settling down to work.
35. I work hard to accomplish my goals.
40. When I make a commitment I can always be counted on to follow through.
45. Sometimes I'm not as dependable or reliable as I should be.
50. I am a productive person who always gets the job done.
55. I never seem to be able to get organized.
60. I strive for excellence in everything I do.

Table A. 3: The NEO Five-Factor Inventory. The NEO-FFI is a well standardized measure composed of five broad personality trait domains based on the model developed by Costa and McCrae (1992). These Big Five dimensions are commonly labeled neuroticism, extraversion, openness, agreeableness and conscientiousness. Each domain consists of twelve items which can be rated on a five-point scale ranging from “strongly disagree” to “strongly agree”. (Costa & McCrae, 1992)

Positive Symptoms
P1 Delusions
P2 Conceptual disorganization
P3 Hallucinatory behavior
P4 Excitement
P5 Grandiosity
P6 Suspiciousness/persecution
P7 Hostility
Negative Symptoms
N1 Blunted affect
N2 Emotional withdrawal
N3 Poor Rapport
N4 Passive/apathetic social withdrawal
N5 Difficulty in abstract thinking
N6 Lack of spontaneity and flow of conversation
N7 Stereotyped thinking
General Symptoms
G1 Somatic concern
G2 Anxiety
G3 Guilt feelings
G4 Tension
G5 Mannerism and posturing
G6 Depression
G7 Motor retardation
G8 Uncooperativeness
G9 Unusual thought content
G10 Disorientation
G11 Poor attention
G12 Lack of judgement and insight
G13 Disturbance of volition
G14 Poor impulse control
G15 Preoccupation
G16 Active social avoidance

Table A. 4: Positive and Negative Syndrome Scale symptoms**.** The PANSS measures the severity of psychotic symptoms (Kay et al., 1987). Seven items refer to the positive symptoms such as hallucinations or disorganized thinking. Another seven items refer to the negative symptoms characterized by cognitive and social impairment or altered affect. The last sixteen items constitute a general psychopathology scale focusing on symptoms such as anxiety, guilt feelings or tension. Each question can be rated on a scale ranging from absent (one) to severe (seven) in an ascending order with more severe symptoms representing a higher score. (Kay et al., 1987)

Behavioral assessment tool	Significant difference between: Site, mean (SD)		Site, mean (SD)		Uncorrected *P*-value	Corrected *P*-value
BDI-II	LMU	31.6 (10.6)	UBS	18.4 (10.5)	.003	.068
	LMU	31.6 (10.6)	Udine	19.9 (10.6)	.038	.720
	LMU	31.6 (10.6)	Milan	24.0 (10.7)	.019	.388
GAF highest lifetime	LMU	81.1 (6.3)	UKK	74.0 (9.9)	.014	.272
score	UKK	74.0 (9.9)	BHAM	84.1 (4.3)	.004	.085
GAF highest score past	LMU	71.0 (10.9)	Milan	50.7 (11.4)	.001	.010*
year	UKK	62.9 (13.6)	BHAM	76.2 (10.4)	.021	.358
	UBS	71.1 (11.2)	Milan	50.7 (11.4)	.002	.038*
	Turku	66.9 (8.4)	Milan	50.7 (11.4)	.040	.638
	Udine	73.4 (11.3)	Milan	50.7 (11.4)	.001	.016*
	BHAM	76.2 (10.4)	Milan	50.7 (11.4)	<.001	.002*
GAF highest score past	LMU	52.0 (12.0)	BHAM	73.8 (12.5)	<.001	<.001*
month	UKK	52.5 (11.4)	BHAM	73.8 (12.5)	<.001	.003*
	UBS	57.6 (11.5)	BHAM	73.8 (12.5)	.014	.233
	Turku	56.5 (11.8)	BHAM	73.8 (12.5)	.012	.208
	Milan	50.7 (11.4)	BHAM	73.8 (12.5)	.003	.055
FROGS	LMU	58.8 (13.4)	Turku	75.0 (11.6)	.005	.104
	LMU	58.8 (13.4)	Udine	73.7 (9.1)	.015	.271
	LMU	58.8 (13.4)	BHAM	72.9 (12.7)	.017	.322
ROCF sum of correct	LMU	33.1 (3.1)	UKK	35.3 (1.2)	.022	.470
Elements, T0	LMU	33.1 (3.1)	Turku	35.6 (1.3)	.025	.506
	LMU	33.1 (3.1)	BHAM	35.5 (1.0)	.035	.666

Table A. 5: Results of the multiple comparison analyses within the CHR group using the Bonferroni-Holms procedure. The one-way ANOVA showed that significant differences between the sites regarding several assessment scores exist. Shown are all assessment tools that differed significantly between the seven sites within the CHR group, together with the respective uncorrected and corrected *P* values. Significance was defined as $p < .05$, *Tests that survive the Bonferroni-Holms procedure to correct for multiple comparisons.

Behavioral assessment tool	Significant difference between:				Uncorrected *P* value	Corrected *P* value
	Site, mean (SD)		Site, mean (SD)			
PANSS total sum	LMU	48.5 (8.7)	UBS	58.5 (13.7)	.009	.167
	UKK	48.6 (11.7)	UBS	58.5 (13.7)	.029	.484
	Udine	43.2 (9.9)	UBS	58.5 (13.7)	.001	.011*
	Turku	40.8 (4.8)	UBS	58.5 (13.7)	<.001	.003
	BHAM	40.5 (6.1)	UBS	58.5 (13.7)	<.001	<.001*
	Milan	44.9 (8.8)	UBS	58.5 (13.7)	.035	.565
PANSS positive sum	LMU	7.3 (0.8)	UBS	8.9 (2.1)	<.001	.002*
	UKK	7.7 (1.4)	UBS	8.9 (2.1)	.034	.670
	Udine	7.5 (0.7)	UBS	8.9 (2.1)	.036	.685
PANSS negative sum	LMU	13.6 (4.6)	BHAM	8.6 (3.0)	.005	.091
	UKK	13.1 (5.4)	BHAM	8.6 (3.0)	.041	.660
	Udine	10.8 (4.7)	UBS	16.6 (4.8)	.009	.151
	Turku	9.8 (2.1)	UBS	16.6 (4.8)	.003	.062
	BHAM	8.6 (3.0)	UBS	16.6 (4.8)	<.001	.001*
	Milan	9.0 (3.3)	UBS	16.6 (4.8)	.003	.064
PANSS general sum	LMU	27.6 (5.7)	UBS	33.1 (8.7)	.044	.796
	Udine	24.9 (5.6)	UBS	33.1 (8.7)	.008	.152
	Turku	23.5 (3.4)	UBS	33.1 (8.7)	.003	.057
	BHAM	24.1 (3.7)	UBS	33.1 (8.7)	.002	.035*
BDI-II	LMU	30.5 (12.4)	Udine	18.0 (13.4)	.039	.810
	LMU	30.5 (12.4)	Turku	16.9 (9.2)	.045	.907
WSS Social Anhedonia	UKK	6.4 (1.6)	BHAM	8.2 (1.5)	.031	.625
	Udine	5.8 (1.6)	BHAM	8.2 (1.5)	.011	.235
NEO-FFI	LMU	62.4 (4.7)	Turku	52.4 (4.7)	<.001	<.001
Conscientiousness	UKK	61.6 (5.8)	Turku	52.4 (4.7)	<.001	.002
	Udine	59.5 (6.5)	Turku	52.4 (4.7)	.021	.362
	UBS	62.0 (3.4)	Turku	52.4 (4.7)	<.001	.004
	BHAM	60.4 (5.5)	Turku	52.4 (4.7)	.005	.082
	Milan	60.1 (5.1)	Turku	52.4 (4.7)	.039	.630
GAF, highest lifetime	LMU	80.0 (6.2)	UBS	88.7 (6.6)	<.001	.002*
score	UKK	79.4 (6.9)	UBS	88.7 (6.6)	<.001	.004*
	Udine	79.6 (7.3)	UBS	88.7 (6.6)	.003	.065
GAF, highest score past	LMU	52.0 (12.0)	BHAM	69.2 (11.8)	.001	.016*
month	LMU	52.0 (12.0)	Milan	68.4 (11.5)	.047	.842
	UBS	49.6 (21.8)	BHAM	69.2 (11.8)	.002	.031*
	UBS	49.6 (21.8)	Milan	68.4 (11.5)	.036	.686
FROGS	LMU	64.9 (14.0)	Turku	79.7 (11.5)	.034	.718
ROCF sum of correct	LMU	32.5 (3.9)	UKK	34.7 (1.3)	.013	.279
elements at T0	LMU	32.5 (3.9)	Udine	35.0 (1.0)	.043	.772
	LMU	32.5 (3.9)	Turku	35.7 (0.7)	.019	.384
	LMU	32.5 (3.9)	BHAM	35.0 (1.4)	.034	.637
ROCF time to completion, immediate recall	UBS	248.2 (165.8)	BHAM	132.3 (44.6)	.036	.747
ROCF time to completion, delayed recall	LMU	117.5 (52.7)	UBS	188.8 (181.8)	.045	.951

Table A. 6: Results of the multiple comparison analyses within the ROD group using the Bonferroni-Holms procedure. The one-way ANOVA showed that significant differences between the sites regarding several behavioral assessment scores exist. Shown are all assessment tools that differed significantly between the seven sites within the ROD group, together with the respective uncorrected and corrected P-values. Significance was defined as $p < .05$. *Tests that survive the Bonferroni-Holms procedure to correct for multiple comparisons.

6 References

Addington, J., Penn, D., Woods, S. W., Addington, D., & Perkins, D. O. (2008). Social functioning in individuals at clinical high risk for psychosis. *Schizophrenia Research*, *99*(1–3), 119–124. https://doi.org/10.1016/j.schres.2007.10.001

Addington, J., Piskulic, D., Liu, L., Lockwood, J., Cadenhead, K. S., Cannon, T. D., … Woods, S. W. (2017). Comorbid diagnoses for youth at clinical high risk of psychosis. *Schizophrenia Research*, *190*, 90–95. https://doi.org/10.1016/j.schres.2017.03.043

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders : DSM-5.* American Psychiatric Association.

Anderson, G., Berk, M., & Maes, M. (2014). Biological phenotypes underpin the physio-somatic symptoms of somatization, depression, and chronic fatigue syndrome. *Acta Psychiatrica Scandinavica*, *129*(2), 83–97. https://doi.org/10.1111/acps.12182

Bailer, J., Witthöft, M., Wagner, H., Mier, D., Diener, C., & Rist, F. (2014). Childhood maltreatment is associated with depression but not with hypochondriasis in later life. *Journal of Psychosomatic Research*, *77*(2), 104–108. https://doi.org/10.1016/j.jpsychores.2014.06.004

Bechdolf, A., Pukrop, R., Köhn, D., Tschinkel, S., Veith, V., Schultze-Lutter, F., … Klosterkötter, J. (2005). Subjective quality of life in subjects at risk for a first episode of psychosis: A comparison with first episode schizophrenia patients and healthy controls. *Schizophrenia Research*, *79*(1), 137–143. https://doi.org/10.1016/j.schres.2005.06.008

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Bernstein, D. P., Ahluvalia, T., Pogge, D., & Handelsman, L. (1997). Validity of the Childhood Trauma Questionnaire in an Adolescent Psychiatric Population. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(3), 340–348. https://doi.org/10.1097/00004583-199703000-00012

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., … Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect*, *27*(2), 169–190. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12615092

Bora, E., Harrison, B. J., Yücel, M., & Pantelis, C. (2013, October). Cognitive impairment in euthymic major depressive disorder: A meta-analysis. *Psychological Medicine*. Psychol Med. https://doi.org/10.1017/S0033291712002085

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bousquet, O., & Elisseeff, A. (2001). Algorithmic stability and generalization performance. In *Advances in Neural Information Processing Systems*. Retrieved from http://papers.nips.cc/paper/1854-algorithmic-stability-and-generalization-performance.pdf

Brodersen, K. H., Soon Ong, C., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. https://doi.org/10.1109/ICPR.2010.764

Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., de Girolamo, G., … Kessler, R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine*, *9*(1), 90. https://doi.org/10.1186/1741-7015-9-90

Carrión, R. E., McLaughlin, D., Goldberg, T. E., Auther, A. M., Olsen, R. H., Olvet, D. M.,

… Cornblatt, B. A. (2013). Prediction of functional outcome in individuals at clinical high risk for psychosis. *JAMA Psychiatry*, *70*(11), 1133–1142. https://doi.org/10.1001/jamapsychiatry.2013.1909

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods q. *Computers and Electrical Engineering*, *40*, 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Cheli, S., Lysaker, P. H., & Dimaggio, G. (2019). Metacognitively oriented psychotherapy for schizotypal personality disorder: A two-case series. *Personality and Mental Health*, *13*(3), 155–167. https://doi.org/10.1002/pmh.1447

Chemerinski, E., Triebwasser, J., Roussos, P., & Siever, L. J. (2013). Schizotypal personality disorder. *Journal of Personality Disorders*, *27*(5), 652–679. https://doi.org/10.1521/PEDI_2012_26_053

Cornblatt, B. A., Carrión, R. E., Addington, J., Seidman, L., Walker, E. F., Cannon, T. D., … Lencz, T. (2012). Risk factors for psychosis: impaired social and role functioning. *Schizophrenia Bulletin*, *38*(6), 1247–1257. https://doi.org/10.1093/schbul/sbr136

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., … Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *NeuroImage*, *28*(3), 663–668. https://doi.org/10.1016/j.neuroimage.2005.08.009

Debbane, M., Eliez, S., Badoud, D., Conus, P., Fluckiger, R., & Schultze-Lutter, F. (2015). Developing Psychosis and Its Risk States Through the Lens of Schizotypy. *Schizophrenia Bulletin*, *41*(suppl 2), S396–S407. https://doi.org/10.1093/schbul/sbu176

Dinzeo, T. J., & Docherty, N. M. (2007). Normal personality characteristics in schizophrenia: a review of the literature involving the FFM. *The Journal of Nervous and Mental Disease*, *195*(5), 421–429. https://doi.org/10.1097/01.nmd.0000253795.69089.ec

Dwyer, D. (2017). *NeuroMiner Manual PRONIA, Personalized Prognostic Tools for*. Retrieved from https://www.pronia.eu/fileadmin/websites/pronia/Neurominer/Neurominer_Manual_1.2.pdf

Dwyer, D., 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

Dwyer, D., & McGuire, P. (2016). Structural and Functional MRI in the Prediction of Psychosis (Vol. 181, pp. 83–94). Karger Publishers. https://doi.org/10.1159/000440916

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*(2), 215–225. https://doi.org/10.1037/0022-006X.51.2.215

Emsley, R. A., Oosthuizen, P. P., Joubert, A. F., Roberts, M. C., & Stein, D. J. (1999). Depressive and Anxiety Symptoms in Patients With Schizophrenia and Schizophreniform Disorder. *The Journal of Clinical Psychiatry*, *60*(11), 747–751. https://doi.org/10.4088/JCP.v60n1105

Falkenberg, I., Valmaggia, L., Byrnes, M., Frascarelli, M., Jones, C., Rocchetti, M., … Fusar-Poli, P. (2015). Why are help-seeking subjects at ultra-high risk for psychosis help-seeking? *Psychiatry Research*, *228*(3), 808–815. https://doi.org/10.1016/j.psychres.2015.05.018

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). *LIBLINEAR: A Library for Large Linear Classification*. *Journal of Machine Learning Research* (Vol. 9). Retrieved from http://www.csie.ntu.edu.tw/

Filzmoser, P., Liebmann, B., & Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics*, *23*(4), 160–171. https://doi.org/10.1002/cem.1225

Flückiger, R., Ruhrmann, S., Debbané, M., Michel, C., Hubl, D., Schimmelmann, B. G., … Schultze-Lutter, F. (2016). Psychosis-predictive value of self-reported schizotypy in a clinical high-risk sample. *Journal of Abnormal Psychology*, *125*(7), 923–932. https://doi.org/10.1037/abn0000192

Fonseca-Pedrero, E., Paino, M., Lemos-Giráldez, S., & Muñiz, J. (2011). Schizotypal traits and depressive symptoms in nonclinical adolescents. *Comprehensive Psychiatry*, *52*(3), 293–300. https://doi.org/10.1016/J.COMPPSYCH.2010.07.001

Frommann, I., Pukrop, R., Brinkmeyer, J., Bechdolf, A., Ruhrmann, S., Berning, J., … Wagner, M. (2011). Neuropsychological profiles in different at-risk states of psychosis: Executive control impairment in the early - And additional memory dysfunction in the late - Prodromal state. *Schizophrenia Bulletin*, *37*(4), 861–873. https://doi.org/10.1093/schbul/sbp155

Fusar-Poli, P., Byrne, M., Valmaggia, L., Day, F., Tabraham, P., Johns, L., … OASIS Team. (2010). Social dysfunction predicts two years clinical outcome in people at ultra high risk for psychosis. *Journal of Psychiatric Research*, *44*(5), 294–301. https://doi.org/10.1016/j.jpsychires.2009.08.016

Fusar-Poli, Paolo, Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., … Yung, A. (2013). The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry*, *70*(1), 107–120. https://doi.org/10.1001/jamapsychiatry.2013.269

Fusar-Poli, Paolo, Deste, G., Smieskova, R., Barlati, S., Yung, A. R., Howes, O., … Borgwardt, S. (2012). Cognitive Functioning in Prodromal Psychosis. *Archives of General Psychiatry*, *69*(6), 562–571. https://doi.org/10.1001/archgenpsychiatry.2011.1592

Fusar-Poli, Paolo, Nelson, B., Valmaggia, L., Yung, A. R., & McGuire, P. K. (2014). Comorbid Depressive and Anxiety Disorders in 509 Individuals With an At-Risk Mental State: Impact on Psychopathology and Transition to Psychosis. *Schizophrenia Bulletin*, *40*(1), 120. https://doi.org/10.1093/SCHBUL/SBS136

Fusar-Poli, Paolo, Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T., & McGuire, P. (2017). Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*, *74*(5), 493–500. https://doi.org/10.1001/jamapsychiatry.2017.0284

Gibb, B. E., Butler, A. C., & Beck, J. S. (2003). Childhood abuse, depression, and anxiety in adult psychiatric outpatients. *Depression and Anxiety*, *17*(4), 226–228. https://doi.org/10.1002/da.10111

Gibb, B. E., Chelminski, I., & Zimmerman, M. (2007). Childhood emotional, physical, and sexual abuse, and diagnoses of depressive and anxiety disorders in adult psychiatric outpatients. *Depression and Anxiety*, *24*(4), 256–263. https://doi.org/10.1002/da.20238

Goldman, H. H., Skodol, A. E., & Lave, T. R. (1992). Revising axis V for DSM-IV: A review of measures of social functioning. *American Journal of Psychiatry*, *149*(9), 1148–1156. https://doi.org/10.1176/ajp.149.9.1148

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual Review of Clinical Psychology*, *10*, 679–708. https://doi.org/10.1146/annurev-clinpsy-032813-153716

Golland, P., & Fischl, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2732*, 330–341. https://doi.org/10.1007/978-3-540-45087-0_28

Gross, G. M., Silvia, P. J., Barrantes-Vidal, N., & Kwapil, T. R. (2015). The dimensional

structure of short forms of the Wisconsin Schizotypy Scales. *Schizophrenia Research*, *166*(1–3), 80–85. https://doi.org/10.1016/j.schres.2015.05.016

Guyon, I., & Elisseeff, A. (2003). *An Introduction to Variable and Feature Selection André Elisseeff*. *Journal of Machine Learning Research* (Vol. 3). Retrieved from http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

Haidl, T. K., Schneider, N., Dickmann, K., Ruhrmann, S., Kaiser, N., Rosen, M., … Piccin, S. (2020). Validation of the Bullying Scale for Adults - Results of the PRONIA-study. *Journal of Psychiatric Research*, *129*, 88–97. https://doi.org/10.1016/j.jpsychires.2020.04.004

Hall, R. C. W. (1995). Global Assessment of Functioning: A Modified Scale. *Psychosomatics*, *36*(3), 267–275. https://doi.org/10.1016/S0033-3182(95)71666-8

Hanssen, M., Peeters, F., Krabbendam, L., Radstake, S., Verdoux, H., & van Os, J. (2003). How psychotic are individuals with non-psychotic disorders? *Social Psychiatry and Psychiatric Epidemiology*, *38*(3), 149–154. https://doi.org/10.1007/s00127-003-0622-7

Hilbig, J., Viliūnienė, R., Friborg, O., Pakalniškienė, V., & Danilevičiūtė, V. (2015). Resilience in a reborn nation: Validation of the Lithuanian Resilience Scale for Adults (RSA). *Comprehensive Psychiatry*, *60*, 126–133. https://doi.org/10.1016/j.comppsych.2015.02.003

Hirschfeld, R. M., Klerman, G. L., Lavori, P., Keller, M. B., Griffith, P., & Coryell, W. (1989). Premorbid personality assessments of first onset of major depression. *Archives of General Psychiatry*, *46*(4), 345–350. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2649038

Hjemdal, O., Friborg, O., Braun, S., Kempenaers, C., Linkowski, P., & Fossion, P. (2011). The Resilience Scale for Adults: Construct Validity and Measurement in a Belgian Sample. *International Journal of Testing*, *11*(1), 53–70. https://doi.org/10.1080/15305058.2010.508570

Huan Xu, Caramanis, C., & Mannor, S. (2012). Sparse Algorithms Are Not Stable: A No-Free-Lunch Theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1), 187–193. https://doi.org/10.1109/TPAMI.2011.177

Hui, C., Morcillo, C., Russo, D. A., Stochl, J., Shelley, G. F., Painter, M., … Perez, J. (2013). Psychiatric morbidity, functioning and quality of life in young people at clinical high risk for psychosis. *Schizophrenia Research*, *148*(1–3), 175–180. https://doi.org/10.1016/J.SCHRES.2013.05.026

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jang, J. H., Shin, N. Y., Shim, G., Park, H. Y., Kim, E., Jang, G.-E., … Kwon, J. S. (2011). Longitudinal Patterns of Social Functioning and Conversion to Psychosis in Subjects at Ultra-High Risk. *Australian & New Zealand Journal of Psychiatry*, *45*(9), 763–770. https://doi.org/10.3109/00048674.2011.595684

Jankowski, N., & Usowicz, K. (2011). Analysis of feature weighting methods based on feature ranking methods for classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7063 LNCS, pp. 238–247). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24958-7_28

Jhung, K., Park, J. Y., Song, Y. Y., Kang, J. I., Lee, E., & An, S. K. (2016). Experiential pleasure deficits in the prodrome: A study of emotional experiences in individuals at ultra-high risk for psychosis and recent-onset schizophrenia. *Comprehensive Psychiatry*, *68*, 209–216. https://doi.org/10.1016/j.comppsych.2016.04.021

Jongsma, H. E., Gayer-Anderson, C., Lasalvia, A., Quattrone, D., Mulè, A., Szöke, A., …

Cristofalo, D. (2018). Treated incidence of psychotic disorders in the multinational EU-GEI study. *JAMA Psychiatry*, *75*(1), 36–46. https://doi.org/10.1001/jamapsychiatry.2017.3554

Kambeitz, J., Cabral, C., Sacchet, M. D., Gotlib, I. H., Zahn, R., Serpa, M. H., … Koutsouleris, N. (2017). Detecting Neuroimaging Biomarkers for Depression: A Meta-analysis of Multivariate Pattern Recognition Studies. *Biological Psychiatry*, *82*(5), 330–338. https://doi.org/10.1016/j.biopsych.2016.10.028

Kambeitz, J., Kambeitz-Ilankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., … Koutsouleris, N. (2015). Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *40*(7), 1742–1751. https://doi.org/10.1038/npp.2015.22

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261–276. https://doi.org/10.1093/schbul/13.2.261

Kernbach, J. M., & Staartjes, V. E. (2022). Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. *Acta Neurochirurgica. Supplement*, *134*, 15–21. https://doi.org/10.1007/978-3-030-85292-4_3

Keshavan, M. S., DeLisi, L. E., & Seidman, L. J. (2011). Early and broadly defined psychosis risk mental states. *Schizophrenia Research*, *126*(1–3), 1–10. https://doi.org/10.1016/j.schres.2010.10.006

Kessler, R. C., & Bromet, E. J. (2013, March). The epidemiology of depression across cultures. *Annual Review of Public Health*. NIH Public Access. https://doi.org/10.1146/annurev-publhealth-031912-114409

Kirkbride, J. B., Errazuriz, A., Croudace, T. J., Morgan, C., Jackson, D., Boydell, J., … Jones, P. B. (2012, March 22). Incidence of schizophrenia and other psychoses in England, 1950-2009: A systematic review and meta-analyses. *PLoS ONE*. Public Library of Science. https://doi.org/10.1371/journal.pone.0031660

Klein, D. Np. and depression: explanatory models and review of the evidence., Kotov, R., & Bufferd, S. J. (2011). Personality and depression: explanatory models and review of the evidence. *Annual Review of Clinical Psychology*, *7*, 269–295. https://doi.org/10.1146/annurev-clinpsy-032210-104540

Klinitzke, G., Romppel, M., Häuser, W., Brähler, E., & Glaesmer, H. (2012). Die deutsche Version des Childhood Trauma Questionnaire (CTQ) – psychometrische Eigenschaften in einer bevölkerungsrepräsentativen Stichprobe. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, *62*(02), 47–51. https://doi.org/10.1055/s-0031-1295495

Klöppel, S., Abdulkadir, A., Jack, C. R., Koutsouleris, N., Mourão-Miranda, J., Vemuri, P., & Vemuri, P. (2012). Diagnostic neuroimaging across diseases. *NeuroImage*, *61*(2), 457–463. https://doi.org/10.1016/j.neuroimage.2011.11.002

Klosterkötter, J. (2016). Early Detection of Psychosis - State of the Art and Future Perspectives (Vol. 181, pp. 1–14). Karger Publishers. https://doi.org/10.1159/000440910

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/s0004-3702(97)00043-x

Kojadinovic, I., & Wottka, T. (2000). *Comparison between a filter and a wrapper approach to variable subset selection in regression problems*. *Proceedings of ESIT*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.1217&rep=rep1&type=pdf

König, I. R., Malley, J. D., Weimar, C., Diener, H.-C., Ziegler, A., & German Stroke Study Collaboration. (2007). Practical experiences on the necessity of external validation. *Statistics in Medicine*, *26*(30), 5499–5511. https://doi.org/10.1002/sim.3069

Koutsouleris, N., Davatzikos, C., Bottlender, R., Patschurek-Kliche, K., Scheuerecker, J.,

Decker, P., … Meisenzahl, E. M. (2012). Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophrenia Bulletin*, *38*(6), 1200–1215. https://doi.org/10.1093/schbul/sbr037

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., … Meisenzahl, E. (2020). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2020.3604

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., … 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. https://doi.org/10.1016/S2215-0366(16)30171-7

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., … PRONIA Consortium. (2018). Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2018.2165

Koutsouleris, N., Meisenzahl, E. M., Borgwardt, S., Riecher-Rössler, A., Frodl, T., Kambeitz, J., … Davatzikos, C. (2015). Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain : A Journal of Neurology*, *138*(Pt 7), 2059–2073. https://doi.org/10.1093/brain/awv111

Krause, A. J., Simon, E. Ben, Mander, B. A., Greer, S. M., Saletin, J. M., Goldstein-Piekarski, A. N., & Walker, M. P. (2017). The sleep-deprived human brain. *Nature Reviews. Neuroscience*, *18*(7), 404–418. https://doi.org/10.1038/nrn.2017.55

Kwapil, T. R., Barrantes-Vidal, N., & Silvia, P. J. (2008). The dimensional structure of the Wisconsin Schizotypy Scales: factor identification and construct validity. *Schizophrenia Bulletin*, *34*(3), 444–457. https://doi.org/10.1093/schbul/sbm098

Lançon, C., Baylé, F.-J., Llorca, P.-M., Rouillon, F., Caci, H., Lancrenon, S., & Gorwood, P. (2012). Time-stability of the "Functional Remission of General Schizophrenia" (FROGS) scale. *European Psychiatry*, *27*(6), 437–441. https://doi.org/10.1016/j.eurpsy.2011.02.014

Larner, A. J. (2018). Number Needed to Diagnose, Predict, or Misdiagnose: Useful Metrics for Non-Canonical Signs of Cognitive Status? *Dementia and Geriatric Cognitive Disorders Extra*, *8*(3), 321–327. https://doi.org/10.1159/000492783

Lee, T. Y., Lee, J., Kim, M., Choe, E., & Kwon, J. S. (2018). Can we predict psychosis outside the clinical high-risk state? A systematic review of non-psychotic risk syndromes for mental disorders. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbx173

Lewandowski, K. E., Barrantes-Vidal, N., Nelson-Gray, R. O., Clancy, C., Kepley, H. O., & Kwapil, T. R. (2006). Anxiety and depression symptoms in psychometrically identified schizotypy. *Schizophrenia Research*, *83*(2–3), 225–235. https://doi.org/10.1016/j.schres.2005.11.024

Lin, A., Wood, S. J., Nelson, B., Beavan, A., McGorry, P., & Yung, A. R. (2015). Outcomes of nontransitioned cases in a sample at ultra-high risk for psychosis. *American Journal of Psychiatry*, *172*(3), 249–258. https://doi.org/10.1176/appi.ajp.2014.13030418

Liu, Y., & Gopalakrishnan, V. (2017). An Overview and Evaluation of Recent Machine Learning Imputation Methods Using Cardiac Imaging Data. *Data*, *2*(1). https://doi.org/10.3390/data2010008

Llorca, P.-M., Lançon, C., Lancrenon, S., Bayle, F.-J., Caci, H., Rouillon, F., & Gorwood, P. (2009). The "Functional Remission of General Schizophrenia" (FROGS) scale: Development and validation of a new questionnaire. *Schizophrenia Research*, *113*(2–3), 218–225. https://doi.org/10.1016/J.SCHRES.2009.04.029

Lysaker, P. H., Bell, M. D., Bioty, S. M., & Zito, W. S. (1995). The frequency of associations between positive and negative symptoms and dysphoria in schizophrenia.

Comprehensive Psychiatry, *36*(2), 113–117. https://doi.org/10.1016/S0010-440X(95)90105-1

Marulanda, S., & Addington, J. (2016). Resilience in individuals at clinical high risk for psychosis. *Early Intervention in Psychiatry*, *10*(3), 212–219. https://doi.org/10.1111/eip.12174

Mason, O. (1995). A confirmatory factor analysis of the structure of schizotypy. *European Journal of Personality*, *9*(4), 271–281. https://doi.org/10.1002/per.2410090404

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis. *Neuropsychology*, *24*(1), 9–34. https://doi.org/10.1037/a0017336

McGlashan, T. H., Walsh, B., & Woods, S. (2010). *The psychosis-risk syndrome : handbook for diagnosis and follow-up*. Oxford University Press.

McGorry, P. D., Hartmann, J. A., Spooner, R., & Nelson, B. (2018). Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry. *World Psychiatry*, *17*(2), 133–142. https://doi.org/10.1002/wps.20514

Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017, May 1). Moving from static to dynamic models of the onset of mental disorder a review. *JAMA Psychiatry*. American Medical Association. https://doi.org/10.1001/jamapsychiatry.2017.0001

Niendam, T. A., Bearden, C. E., Zinberg, J., Johnson, J. K., O'Brien, M., & Cannon, T. D. (2007). The course of neurocognition and social functioning in individuals at ultra high risk for psychosis. *Schizophrenia Bulletin*, *33*(3), 772–781. https://doi.org/10.1093/schbul/sbm020

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., … Marder, S. R. (2008). The MATRICS consensus cognitive battery, part 1: Test selection, reliability, and validity. *American Journal of Psychiatry*, *165*(2), 203–213. https://doi.org/10.1176/appi.ajp.2007.07010042

Orrù, 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. https://doi.org/10.1016/j.neubiorev.2012.01.004

Park, S.-C., Hahn, S.-W., Hwang, T.-Y., Kim, J.-M., Jun, T.-Y., Lee, M.-S., … Park, Y. C. (2014). Does age at onset of first major depressive episode indicate the subtype of major depressive disorder?: the clinical research center for depression study. *Yonsei Medical Journal*, *55*(6), 1712–1720. https://doi.org/10.3349/ymj.2014.55.6.1712

Pedersen, G., Hagtvet, K. A., & Karterud, S. (2007). Generalizability studies of the Global Assessment of Functioning-Split version. *Comprehensive Psychiatry*, *48*(1), 88–94. https://doi.org/10.1016/j.comppsych.2006.03.008

Pedersen, G., & Karterud, S. (2012). The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. *Comprehensive Psychiatry*, *53*(3), 292–298. https://doi.org/10.1016/j.comppsych.2011.04.007

Perlis, R. H., Uher, R., Ostacher, M., Goldberg, J. F., Trivedi, M. H., Rush, A. J., & Fava, M. (2011). Association between bipolar spectrum features and treatment outcomes in outpatients with major depressive disorder. *Archives of General Psychiatry*, *68*(4), 351–360. https://doi.org/10.1001/archgenpsychiatry.2010.179

Popovic, D., Ruef, A., Dwyer, D. B., Hedderich, D., Antonucci, L. A., Kambeitz-Ilankovic, L., … Koutsouleris, N. (2020). O8.5. Signs Of Adversity - a novel machine learning approach to childhood trauma, brain structure and clinical profiles. *Schizophrenia Bulletin*, (46), 20–20. Retrieved from https://academic.oup.com/schizophreniabulletin/article/46/Supplement_1/S20/5838922

Pukrop, R., Schultze-Lutter, F., Ruhrmann, S., Brockhaus-Dumke, A., Tendolkar, I.,

Bechdolf, A., … Klosterkötter, J. (2006). Neurocognitive functioning in subjects at risk for a first episode of psychosis compared with first- and multiple-episode schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, *28*(8), 1388–1407. https://doi.org/10.1080/13803390500434425

Reich, J., Noyes, R., Hirschfeld, R., Coryell, W., & O'Gorman, T. (1987). State and personality in depressed and panic patients. *American Journal of Psychiatry*, *144*(2), 181–187. https://doi.org/10.1176/ajp.144.2.181

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. https://doi.org/10.1017/S0033291713002535

Ruhrmann, S., Schultze-Lutter, F., & Klosterkötter, J. (2010). Probably at-risk, but certainly ill — Advocating the introduction of a psychosis spectrum disorder in DSM-V. *Schizophrenia Research*, *120*(1–3), 23–37. https://doi.org/10.1016/j.schres.2010.03.015

Rutigliano, G., Valmaggia, L., Landi, P., Frascarelli, M., Cappucciati, M., Sear, V., … Fusar-Poli, P. (2016). Persistence or recurrence of non-psychotic comorbid mental disorders associated with 6-year poor functional outcomes in patients at ultra high risk for psychosis. *Journal of Affective Disorders*, *203*, 101–110. https://doi.org/10.1016/j.jad.2016.05.053

Şahin, S., Yüksel, Ç., Güler, J., Karadayı, G., Akturan, E., Göde, E., … Üçok, A. (2013). The history of childhood trauma among individuals with ultra high risk for psychosis is as common as among patients with first-episode schizophrenia. *Early Intervention in Psychiatry*, *7*(4), 414–420. https://doi.org/10.1111/eip.12022

Salokangas, R. K. R., Heinimaa, M., From, T., Löyttyniemi, E., Ilonen, T., Luutonen, S., … EPOS group. (2014). Short-term functional outcome and premorbid adjustment in clinical high-risk patients. Results of the EPOS project. *European Psychiatry*, *29*(6), 371–380. https://doi.org/10.1016/j.eurpsy.2013.10.003

Salokangas, R. K. R., Ruhrmann, S., von Reventlow, H. G., Heinimaa, M., Svirskis, T., From, T., … Klosterkötter, J. (2012). Axis I diagnoses and transition to psychosis in clinical high-risk patients EPOS project: Prospective follow-up of 245 clinical high-risk outpatients in four countries. *Schizophrenia Research*, *138*(2–3), 192–197. https://doi.org/10.1016/j.schres.2012.03.008

Schultze-Lutter, F., Klosterkötter, J., Nikolaides, A., & Ruhrmann, S. (2015). Personality dimensions in persons symptomatically at risk of psychosis: pronounced but lacking a characteristic profile. *Early Intervention in Psychiatry*, *9*(3), 242–247. https://doi.org/10.1111/eip.12147

Schultze-Lutter, F., Ruhrmann, S., Fusar-Poli, P., Bechdolf, A., Schimmelmann, B. G., & Klosterkötter, J. (2012). Basic symptoms and the prediction of first-episode psychosis. *Current Pharmaceutical Design*, *18*(4), 351–357. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22239566

Schultze-Lutter, F., Ruhrmann, S., Picker, H., Von Reventlow, H. G., Brockhaus-Dumke, A., & Klosterkötter, J. (2007). Basic symptoms in early psychotic and depressive disorders. *British Journal of Psychiatry*, *191*(SUPPL. 51). https://doi.org/10.1192/bjp.191.51.s31

Schultze-Lutter, F., Ruhrmann, S., Picker, H., Von Reventlow, H. G., Daumann, B., Brockhaus-Dumke, A., … Pukrop, R. (2007). Relationship between subjective and objective cognitive function in the early and late prodrome. *British Journal of Psychiatry*, *191*(S51), s43–s51. https://doi.org/10.1192/bjp.191.51.s43

Schulz, A., Becker, M., Van der Auwera, S., Barnow, S., Appel, K., Mahler, J., … Grabe, H. J. (2014). The impact of childhood trauma on depression: Does resilience matter? Population-based results from the Study of Health in Pomerania. *Journal of Psychosomatic Research*, *77*(2), 97–103. https://doi.org/10.1016/j.jpsychores.2014.06.008

Shah, J. L., Crawford, A., Mustafa, S. S., Iyer, S. N., Joober, R., & Malla, A. K. (2017). Is the clinical high-risk state a valid concept? Retrospective examination in a first-episode psychosis sample. *Psychiatric Services*, *68*(10), 1046–1052. https://doi.org/10.1176/appi.ps.201600304

Shankman, S. A., Nelson, B. D., Harrow, M., & Faull, R. (2010). Does physical anhedonia play a role in depression? A 20-year longitudinal study. *Journal of Affective Disorders*, *120*(1–3), 170–176. https://doi.org/10.1016/j.jad.2009.05.002

Shardlow, M. (2016). An Analysis of Feature Selection Techniques. *The University of Manchester*, 1–7. Retrieved from http://syllabus.cs.manchester.ac.uk/pgt/2018/COMP61011/goodProjects/Shardlow.pdf

Shimodera, S., Kawamura, A., & Furukawa, T. A. (2012). Physical pain associated with depression: results of a survey in Japanese patients and physicians. *Comprehensive Psychiatry*, *53*(6), 843–849. https://doi.org/10.1016/j.comppsych.2011.11.004

Shin, M.-S., Park, S.-Y., Park, S.-R., Seol, S.-H., & Kwon, J. S. (2006). Clinical and empirical applications of the Rey–Osterrieth Complex Figure Test. *Nature Protocols*, *1*(2), 892–899. https://doi.org/10.1038/nprot.2006.115

Statnikov, A., Hardin, D., Guyon, I., & Aliferis, C. F. (2009). *A Gentle Introduction to Support Vector Machines in Biomedicine*. Retrieved from https://med.nyu.edu/chibi/sites/default/files/chibi/Final.pdf

Thompson, A. D., Nelson, B., Yuen, H. P., Lin, A., Amminger, G. P., McGorry, P. D., … Yung, A. R. (2014). Sexual trauma increases the risk of developing psychosis in an ultra high-risk "prodromal" population. *Schizophrenia Bulletin*, *40*(3), 697–706. https://doi.org/10.1093/schbul/sbt032

Tutz, G., & Ramzan, S. (2015). Improved methods for the imputation of missing data by nearest neighbor methods. *Computational Statistics & Data Analysis*, *90*, 84–99. https://doi.org/10.1016/J.CSDA.2015.04.009

Van Os, J., & Jones, P. B. (2001). Neuroticism as a risk factor for schizophrenia. *Psychological Medicine*, *31*(6), 1129–1134. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11513380

Vargas, T., Ahmed, A. O., Strauss, G. P., Brandes, C. M., Walker, E. F., Buchanan, R. W., … Mittal, V. A. (2019). The latent structure of depressive symptoms across clinical high risk and chronic phases of psychotic illness. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0563-x

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*(1), 91. https://doi.org/10.1186/1471-2105-7-91

Verdoux, H., Van Os, J., Maurice-Tison, S., Gay, B., Salamon, R., & Bourgeois, M. L. (1999). Increased occurrence of depression in psychosis-prone subjects: A follow-up study in primary care settings. *Comprehensive Psychiatry*, *40*(6), 462–468. https://doi.org/10.1016/S0010-440X(99)90091-3

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement*, *1*(2), 165–178. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7850498

Weightman, M. J., Air, T. M., & Baune, B. T. (2014). A review of the role of social cognition in major depressive disorder. *Frontiers in Psychiatry*, *5*, 179. https://doi.org/10.3389/fpsyt.2014.00179

Widiger, T. (2011). Personality and psychopathology. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *10*(2), 103–106. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21633679

Wigman, J. T. W., van Nierop, M., Vollebergh, W. A. M., Lieb, R., Beesdo-Baum, K., Wittchen, H.-U., & van Os, J. (2012). Evidence that psychotic symptoms are prevalent in disorders of anxiety and depression, impacting on illness onset, risk, and severity--

implications for diagnosis and ultra-high risk research. *Schizophrenia Bulletin*, *38*(2), 247–257. https://doi.org/10.1093/schbul/sbr196

Wingenfeld, K., Spitzer, C., Mensebach, C., Grabe, H., Hill, A., Gast, U., … Driessen, M. (2010). Die deutsche Version des Childhood Trauma Questionnaire (CTQ): Erste Befunde zu den psychometrischen Kennwerten. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, *60*(11), 442–450. https://doi.org/10.1055/s-0030-1247564

Winterstein, B. P., Silvia, P. J., Kwapil, T. R., Kaufman, J. C., Reiter-Palmon, R., & Wigert, B. (2011). Brief assessment of schizotypy: Developing short forms of the Wisconsin Schizotypy Scales. *Personality and Individual Differences*, *51*(8), 920–924. https://doi.org/10.1016/j.paid.2011.07.027

Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., & Marquand, A. F. (2015). From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, *57*, 328–349. https://doi.org/10.1016/J.NEUBIOREV.2015.08.001

Xia, J., He, Q., Li, Y., Xie, D., Zhu, S., Chen, J., … Wang, X. (2011). The relationship between neuroticism, major depressive disorder and comorbid disorders in Chinese women. *Journal of Affective Disorders*, *135*(1–3), 100–105. https://doi.org/10.1016/j.jad.2011.06.053

Zakzanis, K. K., Leach, L., & Kaplan, E. (1999). *Neuropsychological differential diagnosis. - PsycNET*. Retrieved from https://psycnet.apa.org/record/1999-04317-000

Zarogianni, E., Moorhead, T. W. J., & 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. https://doi.org/10.1016/j.nicl.2013.09.003

Zhu, Z., Ong, Y.-S., & Dash, M. (2007). Wrapper–Filter Feature Selection Algorithm Using a Memetic Framework. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *37*(1), 70–76. https://doi.org/10.1109/TSMCB.2006.883267

Zisook, S., John Rush, A., Albala, A., Alpert, J., Balasubramani, G. K., Fava, M., … Wisniewski, S. (2004). Factors that differentiate early vs. later onset of major depression disorder. *Psychiatry Research*, *129*(2), 127–140. https://doi.org/10.1016/j.psychres.2004.07.004

Zisook, S., Lesser, I., Stewart, J. W., Wisniewski, S. R., Balasubramani, G. K., Fava, M., … Rush, A. J. (2007). Effect of Age at Onset on the Course of Major Depressive Disorder. *American Journal of Psychiatry*, *164*(10), 1539–1546. https://doi.org/10.1176/appi.ajp.2007.06101757

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Eidesstattliche Versicherung

Wosgien, Antonia

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

Clinical data patterns supporting differential diagnosis of recent onset of depression and clinical high risk for psychosis

Differentialdiagnostische Einordnung von Patientinnen und Patienten in die depressive Ersterkrankung oder das klinische Hochrisikostadium für Psychosen mithilfe maschineller Mustererkennung

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Berlin, den 21.07.22

Ort, Datum

Antonia Wosgien

Name der Doktorandin