THE UNIVERSITY
of EDINBURGH

Machine learning and brain imaging in psychosis

Eleni Zarogianni

PhD Degree

University of Edinburgh

September 2015

Declaration

I declare that the work presented in this thesis is my own and the contribution of others is clearly documented where relevant. This work has not been submitted for any other degree or professional qualification.

Eleni Zarogianni

Acknowledgements

I would first like to express my gratitude and offer many thanks to my first supervisor, Professor Stephen M. Lawrie, for his trust and continued guidance throughout the four years of this thesis and to my second supervisor Amos J. Storkey.

I would, also, like to offer my special thanks to Dr. Bill Moorhead for his invaluable help and support with the pre-processing of the brain scans. Special thanks go to Professor Eve Johnstone and Professor David Owens for their conceptualization of the Edinburgh High Risk Study. I am also very grateful to Professor Anita Riecher-Rossler, Professor Stefan J. Borgwardt and Dr. Renata Smieskova for trusting me with their data from the FePsy study and for their time and support.

Furthermore, I would like to thank my friends here in Edinburgh and my dearest friends back in Greece for their continued support, their understanding and patience when it was most needed. And finally, I am most grateful to my family, my parents Vassiliki and Christos and my sister Maria, for their ongoing support and belief in me; I could not have done any of these without them. I dedicate this thesis to them.

TABLE OF CONTENTS

Abstract

Over the past years early detection and intervention in schizophrenia have become a major objective in psychiatry. Early intervention strategies are intended to identify and treat psychosis prior to fulfilling diagnostic criteria for the disorder. To this aim, reliable early diagnostic biomarkers are needed in order to identify a high-risk state for psychosis and also predict transition to frank psychosis in those high-risk individuals destined to develop the disorder. Recently, machine learning methods have been successfully applied in the diagnostic classification of schizophrenia and in predicting transition to psychosis at an individual level based on magnetic resonance imaging (MRI) data and also neurocognitive variables.

This work investigates the application of machine learning methods for the early identification of schizophrenia in subjects at high risk for developing the disorder. The dataset used in this work involves data from the Edinburgh High Risk Study (EHRS), which examined individuals at a heightened risk for developing schizophrenia for familial reasons, and the FePsy (Fruherkennung von Psychosen) study that was conducted in Basel and involves subjects at a clinical high-risk state for psychosis.

The overriding aim of this thesis was to use machine learning, and specifically Support Vector Machine (SVM), in order to identify predictors of transition to psychosis in high-risk individuals, using baseline structural MRI data. There are three aims pertaining to this main one. (i) Firstly, our aim was to examine the feasibility of distinguishing at baseline those individuals who later developed schizophrenia from those who did not, yet had psychotic symptoms using SVM and baseline data from the EHRS study. (ii) Secondly, we intended to examine if our classification approach could generalize to clinical high-risk cohorts, using neuroanatomical data from the FePsy study. (iii) In a

more exploratory context, we have also examined the diagnostic performance of our classifier by pooling the two datasets together.

With regards to the first aim, our findings suggest that the early prediction of schizophrenia is feasible using a MRI-based linear SVM classifier operating at the single-subject level. Additionally, we have shown that the combination of baseline neuroanatomical data with measures of neurocognitive functioning and schizotypal cognition can improve predictive performance. The application of our pattern classification approach to baseline structural MRI data from the FePsy study highly replicated our previous findings. Our classification method identified spatially distributed networks that discriminate at baseline between subjects that later developed schizophrenia and other related psychoses and those that did not. Finally, a preliminary classification analysis using pooled datasets from the EHRS and the FePsy study supports the existence of a neuroanatomical pattern that differentiates between groups of high-risk subjects that develop psychosis against those who do not across research sites and despite any between-sites differences.

Taken together, our findings suggest that machine learning is capable of distinguishing between cohorts of high risk subjects that later convert to psychosis and those that do not based on patterns of structural abnormalities that are present before disease onset. Our findings have some clinical implications in that machine learning-based approaches could advise or complement clinical decision-making in early intervention strategies in schizophrenia and related psychoses. Future work will be, however, required to tackle issues of reproducibility of early diagnostic biomarkers across research sites, where different assessment criteria and imaging equipment and protocols are used. In addition, future projects may also examine the diagnostic and prognostic value of multimodal neuroimaging data, possibly combined with other clinical, neurocognitive, genetic information.

Lay Summary

One of the major goals in clinical psychiatry today is to diagnose individuals destined to develop a psychiatric disorder as early as possible and treat them accordingly, with the ultimate aim of either preventing disease onset or at least ameliorating clinical outcome and course. The work presented in this thesis focuses on examining the feasibility of achieving an early diagnosis for individuals considered at high risk for developing schizophrenia and other related psychoses.

To do so, a machine learning method called Support Vector Machine (SVM), was implemented and structural imaging data from individuals at high risk for schizophrenia and other related psychoses were used. Two cohorts of high risk patients were utilised; the Edinburgh High Risk Study (EHRS) cohort, which examined individuals at a heightened risk for developing schizophrenia for familial reasons, and a cohort of individuals from the FePsy (Fruherkennung von Psychosen) study that was conducted in Basel and involved subjects at a clinical high-risk state for psychosis.

Taken together, the findings produced suggest that machine learning is capable of distinguishing between cohorts of high risk subjects that later convert to psychosis and those that do not based on patterns of structural abnormalities that are present before disease onset.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ARMS	At-Risk Mental State
ARMS-T	At-Risk Mental State that transition to psychosis (FePsy study)
ARMS-NT	At-Risk Mental State that did not transition to psychosis (FePsy study)
CV	Cross-Validation
DSM	Diagnostic and Statistical Manual for Mental Disorders
EHRS	Edinburgh High Risk Study
GM	Grey Matter
HC	Healthy Control individuals
HR	High Risk
HR[ill]	High-risk subjects that developed schizophrenia (sample from the EHRS)
HR[symp]	High-risk subjects that did not develop schizophrenia, yet had psychotic symptoms (sample from the EHRS)
HR-T	High-risk subjects that developed schizophrenia
HR-NT	High-risk subjects that did not develop schizophrenia
ICD	International Classification of Diseases
LOO	Leave-One-Out
MRI	Magnetic Resonance Imaging
RFE	Recursive Feature Elimination
sMRI	structural Magnetic Resonance Imaging
SVM	Support Vector Machine
VBM	Voxel-based Morphometry

CHAPTER 1

Schizophrenia

1.1 Introduction

Schizophrenia is one of the most severe and debilitating mental disorders, posing a huge burden on the affected person and the society as well, as most people with schizophrenia are unable to continue in employment or education. It is estimated that approximately 1% of the population worldwide suffers from the disorder, which generally manifests itself in adolescence and early adulthood although its symptoms are chronic and devastating. Schizophrenia has a variable phenotypic expression although its most common symptoms primarily include a severe decline in cognitive function and abnormal social behaviour. Despite numerous years of scientific research, the aetiology and the underlying pathophysiological mechanisms of schizophrenia still remain elusive. The risk for developing schizophrenia, however, primarily involves a major genetic contribution and environmental factors, possibly interacting with genetic susceptibility.

1.2 General overview

Schizophrenia was first described by the German physician Emil Kraepelin. In 1887, Kraepelin introduced the term *Dementia Praecox* to describe a disease of early age of onset that was characterised by progressive decline of emotional and cognitive processes. Based on his clinical experience, Kraepelin observed various psychotic conditions that he thought were manifestations of a single disease entity (of a single aetiology) that ultimately resulted in cognitive deficits and executive dysfunctions (Kraepelin, 1919). Kraepelin distinguished schizophrenia from other types of psychoses, such as manic-depressive psychosis, and also developed a classification system of mental illnesses that influenced contemporary classification guidelines.

It was in the early twentieth century, however, that the term *Schizophrenia* was first coined by the Swiss psychiatrist Eugen Bleuler who not only dissociated the disorder from a deteriorating course, but also hypothesised that the disease primarily involved a loosening of thoughts and feelings, in the context of a ‘split mind’ (Bleuler, 1911).

The term schizophrenia (from the greek words: ‘σχίζειν’ =split and ‘φρην’=mind) has, however, led to much confusion about the nature of the illness, in that it is often mistakenly thought of as a split personality disorder; a condition totally unrelated to schizophrenia.

Bleuler believed that at the core of schizophrenia lies the loss of association between thought processes, emotion and behaviour. He introduced the concept of primary symptoms, which he thought were particularly characteristic of the disorder, and secondary symptoms of schizophrenia. Based on his clinical experience, he concluded that the four fundamental symptoms of schizophrenia included associational disturbances, ambivalence, affective disturbances and autistic behaviour (the so-called four A’s, now considered as negative symptoms) whereas hallucinations and delusions were deemed as secondary (or accessory) symptoms of the disorder.

1.3 Current diagnosis and clinical description

Our current conceptualization of schizophrenia derives fundamentally from the descriptions of Kraepelin, Bleuler and Schneider, who described a number of first-rank symptoms that are considered particularly characteristic of the disorder and now best describe the positive symptomatology (Schneider, 1959).

Diagnosis of schizophrenia is nowadays typically attained through structured interviews and meeting criteria in the Diagnostic and Statistical Manual for Mental Disorders, fifth edition (DSM-5) established by the American psychiatric Association (APA, 2013) or the International Classification of Diseases and Related Health Problems, 10th edition (ICD-10), established by the World Health Organization (WHO, 1994).

Both the ICD-10 and DSM-5 classifications are based on expert clinical observation and scientific research and provide operationalized criteria that are used to make a reliable diagnosis of schizophrenia. Tables 1.1 and 1.2 provide a summary of the

criteria needed to reach a diagnosis of schizophrenia. Differences between the DSM-5 and ICD-10 exist in terms of symptom duration with the former suggesting a more chronic course (symptoms have to be present for 6 months whereas in ICD-10 one month is required) and requiring an accompanying deterioration in functioning in order to reach a diagnosis. In addition, both systems require that other diagnoses such as mood disorders with psychotic symptoms, schizoaffective disorders and drug-induced psychoses should be first ruled out. Despite their discrepancies, however, definitions in both diagnostic systems are sufficiently similar to promote standardization across centers and ensure international agreement.

Although the use of standardized criteria is clinically relevant and enhances reliability in the diagnostic process, most of the attributes defining schizophrenia are primarily inferential and depend on self-reported subjective experience. The clinical presentation of the disease itself is complex, with patients manifesting heterogeneous symptomatology, comorbidity with other psychiatric disorders, variable illness course and adherence to treatment.

Two broad types of symptoms, including positive or psychotic symptoms and negative symptoms, characterize schizophrenia. Positive symptoms are associated with the presence of dysfunctions or abnormalities and include disturbances of perception and inferential thinking, such as hallucinations and delusions, disorganized thinking and speech. Patients with schizophrenia may hear voices that other people do not hear, may think that someone else is controlling their thoughts or reading their minds or even believe they are victims of a plot with the intention to harm them. They often have a hard time organizing their thoughts or speaking in a logical, coherent manner. These symptoms usually become apparent at the early stages of the disease, in adolescence or early adulthood and are usually mitigated with the use of anti-psychotic treatment, although recurrent relapses and remissions commonly go on throughout the patient’s life.

Negative symptoms generally succeed this florid psychotic phase and are mostly linked to a lack of motivation and emotional responsiveness, a loss of interest in oneself, social isolation and withdrawal. Current anti-psychotic medication does not seem to treat negative symptoms, which tend to persist for longer than positive

symptoms and are generally associated with a poorer outcome (Hunter & Barry 2012). Disturbances in cognitive functions form another cluster of signs and symptoms of schizophrenia, although they are not widely considered as distinct a dimension of the disorder as positive and negative symptoms. Impairments in executive functions, working memory and attention are generally signs that might precede the florid psychotic stage of the disorder and might be apparent in the prodrome to overt psychosis, and usually persist or even deteriorate during the course of the disease.

No single symptom can be considered pathognomonic for schizophrenia and patients usually present with different constellations of symptoms. Symptom heterogeneity across patients can be described in the context of distinct clinical subtypes that are captured in the previous DSM-IV and the ICD-10, and include among others: paranoid, disorganized, catatonic and undifferentiated schizophrenia. Clinical presentation of patients is not, however, limited to only one subtype, nor this subtype remains stable throughout the course of the illness, thus rendering the subtyping of schizophrenia diagnostically unstable and possibly unreliable (Tandon et al. 2009). As a result, descriptions of subtypes have been removed in the most recent DSM-5 (Tandon et al. 2013).

Comorbidity with other psychiatric disorders such as depression or comorbid substance-use disorders is common among schizophrenia patients. Depressive symptoms might be present at any point during the illness, with the majority of patients suffering from them in the pre-morbid (pre-psychotic) phase of the disorder or during acute psychotic exacerbations, and usually as an adverse side effect of anti-psychotic medication. Substance abuse, such as nicotine, alcohol or cannabis is highly prevalent among schizophrenia patients and is linked to poorer outcome, including increased psychotic symptoms and poorer adherence to treatment (Dixon 1999, Winklbauer et al. 2006).

Table 1.1 DSM-5 Diagnostic criteria for schizophrenia

Criterion A.
Two (or more of the following), each present for a significant period of time during a 1-month period (or less if successfully treated). At least one of these should include 1-3.

1. Delusions
2. Hallucinations
3. Disorganised speech
4. Grossly disorganized or catatonic behaviour
5. Negative symptoms, e.g. avolition, alogia, diminished emotional expression

Criterion B.
Social/ Occupational Dysfunction

Criterion C.
Duration of at least six months, with at least one month of active symptoms.

Table 1.2 ICD-10 Diagnostic criteria for schizophrenia

Either *at least one* of the syndromes, symptoms, and signs listed under (1) below, *or* at least two of the symptoms and signs listed under (2) should be present for most of the time during an episode of psychotic illness lasting for at least 1 month (or at some time during most of the days).

(1) At least one of the following must be present:

(a) thought echo, thought insertion or withdrawal, or thought broadcasting;
(b) delusions of control, influence, or passivity, clearly referred to body or limb movements or specific thoughts, actions, or sensations; delusional perception;
(c) hallucinatory voices giving a running commentary on the patient's behavior, or discussing the patient among themselves, or other types of hallucinatory voices coming from some part of the body;
(d) persistent delusions of other kinds that are culturally inappropriate and completely impossible (e.g., being able to control the weather, or being in communication with aliens from another world).

(2) *Or* at least two of the following:

(a) persistent hallucinations in any modality, when occurring every day for at least 1 month, when accompanied by delusions (which may be fleeting or half-formed) without clear
affective content, or when accompanied by persistent overvalued ideas;
(b) neologisms, breaks, or interpolations in the train of thought, resulting in incoherence or irrelevant speech;
(c) catatonic behavior, such as excitement, posturing or waxy flexibility, negativism, mutism, and stupor;
(d) negative symptoms, such as marked apathy, paucity of speech, and blunting or incongruity of emotional responses (it must be clear that these are not due to depression or to neuroleptic medication).

1.4 Epidemiology of schizophrenia

Schizophrenia is one of the most severe and devastating mental disorders, ranked by the World Health Organization as one of the top ten illnesses contributing to global burden of disease (Jablensky 1997). Epidemiological studies provide a useful tool in the understanding of the disorder by studying and describing the distribution pattern of schizophrenia in populations and across cultures and identifying factors that convey risk for the development of psychosis.

Based on epidemiological studies conducted over the last century, the prevalence of schizophrenia (i.e. the number of cases at any time point) in the general population is approximately 1% (Jablensky et al. 1992). Based on the only global study that directly generated incidence data (WHO-10 nation study, Jablensky et al. 1992), the annual incidence of schizophrenia (i.e. the number of new cases) was estimated to range from 16-40/100,000 per year using broad diagnostic criteria (ICD-9, WHO-1978) and 7-14/100,000 using narrow criteria (CATEGO class S+, Wing et al. 1974). Since incidence is a measure of the number of new cases with a disease over the number of cases who are at-risk for developing the illness over a specified period of time, this measure is highly dependent on the diagnostic criteria used, and thus incidence rates exhibit some variation across sites. A recent meta-analysis of all published studies from 1965 to 2001 reported a median incidence rate of 15.2 per 100,000 per year with 80% confidence interval rates (10^{th} to 90^{th} decile) ranging from 8-43 per 100,000 per year (McGrath et al. 2004). The notion that the incidence rates of schizophrenia vary across sites has also been supported by another meta-analysis of the same research group (McGrath et al.2008), which further reported that sex, urbanicity and migration (Cantor-Graae & Selten 2005) are associated with a higher risk for developing schizophrenia.

Gender differences in schizophrenia have been widely reported by a host of studies, showing differences in disease onset, course and outcome (Leung & Chue 2000). A recent meta-analysis reported higher incidence of schizophrenia in men versus women (Aleman et al. 2003). A recent review quantified a median male to female

rate ratio of 1.4: 1 (McGrath et al. 2004, McGrath et al. 2008). Differences in age of onset have been widely reported, with men having an earlier disease onset than women (Leung & Chue 2000, Castle et al. 1998, Goldstein et al. 1989, Gureje 1991), although this finding is not universally replicated (Folnegovic 1994, Addigton et al. 1996, Murthy et al. 1998). In addition, earlier studies reported a worse course of illness in male patients, who tend to exhibit more severe negative symptoms (Castle & Murray 1991, Gur et al. 1996, Riecher-Rossler & Hafner 2000) and have worse outcome resulting in higher rates of relapses and institutionalizations (Haro et al. 2006).

Women with schizophrenia usually have better social functioning, fewer admissions to hospitals (Morgan et al. 2008) and better response to treatment (Goldstein et al. 2002, Usall et al. 2007). A possible explanation for the better level of functioning might be the later age onset of schizophrenia in women that enables them to attain a certain level of social functioning by the time they develop full-blown symptoms and are first admitted to clinical services, thus contributing to a better clinical outcome. The effect of oestrogen in women has also been suggested as a protective factor in the development of schizophrenia (Hafner 2000).

1.4.1 Premorbid phase and onset

The typical onset is between the ages of 16 and 30, with rare cases developing schizophrenia by the age of 12 (childhood-onset schizophrenia) (Russel 1994) or after the age of 45 (late-onset schizophrenia) (Folsom et al. 2006). The onset of schizophrenia is often preceded by a premorbid phase where subtle and non-specific abnormalities in motor, cognitive and social functioning are observed (Figure 1.1; Schenkel & Silverstein 2004). The premorbid phase is usually followed by a prodromal phase that can last from several months or even years and it is characterised by non-specific changes in the person's behaviour and a noticeable decline in functioning. At this stage, academic, work and/or social functioning might show signs of impairment; the individual might report increased anxiety,

suspiciousness or a non-specific feeling that the world seems different. The individual is considered at-risk for developing psychosis and often presents with attenuated or transient psychotic symptoms; however these are not severe or frequent enough to warrant a diagnosis of schizophrenia or any other psychotic-related disorder. Onset of schizophrenia is usually accompanied by prolonged and/or acute episodes of psychosis that lead individuals to their first hospital admission and possibly the first administration of anti-psychotic treatment. There is often significant delay from the index individual and his family in seeking clinical management and treatment that can last up to 1-2 years (McGlashan 1999). Negative symptoms tend to occur later or become more severe over the course of illness (Fuller et al. 2003).

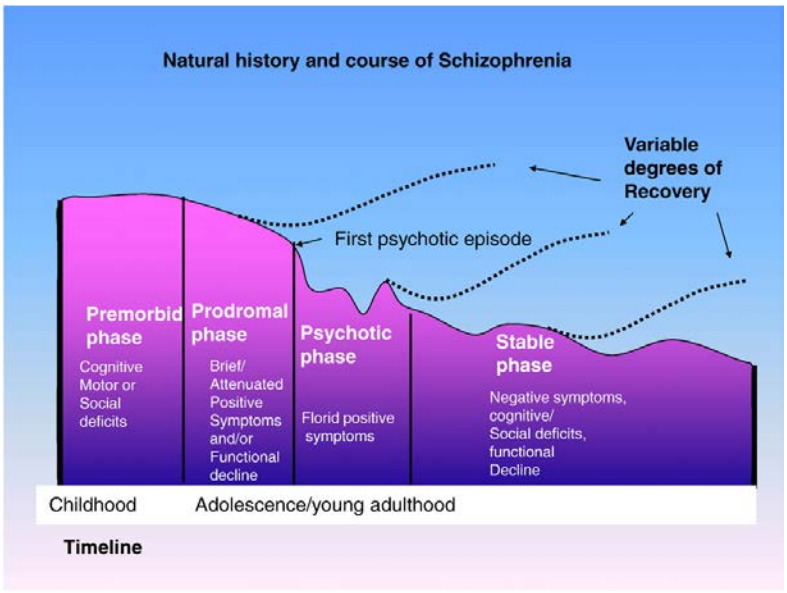


Figure 1.1 The stages of schizophrenia. Figure adjusted from Tandon et al. 2009.

The course of schizophrenia varies substantially across patients. After the first psychotic episode, the majority of individuals have one or more relapses in the course of illness that may lead to hospitalization. The first ten years of schizophrenia are usually characterized by repeated relapses of psychotic episodes of variable extent and severity, followed by varied periods of symptoms' remission (Ciompi 1980, Andreasen et al. 2005) and accompanied by a decline in cognitive functioning

over the course of illness. Over the long-term of the illness, positive symptoms seem to subside and become less devastating, partly on account of the anti-psychotic medication, and negative symptoms and cognitive dysfunction seem to persist.

1.4.2 Outcome

Despite advances in treatment, the outcomes for most schizophrenia patients remain poor. Only a small fraction of patients achieve full recovery. The majority of patients, after a long course of relapses and remissions, reach a stable phase in the illness trajectory where social isolation, impaired social and cognitive functioning remain. Most patients have trouble finding or keeping a job and being active in the community. Individuals with schizophrenia have higher risk of attempting and/or committing suicide (Hawton et al. 2005), increased incidence of comorbid medical or psychiatric illnesses (Leucht et al. 2007, Carney et al. 2006) and lower quality of life (Eack & Newhill 2007) than the general population.

The outcome of schizophrenia is influenced by several factors, including premorbid functioning, the characteristics of the illness symptomatology and course, and social variables. A favourable outcome in schizophrenia has been linked with an acute onset of the illness, a later age of onset, the female gender and absence of substance abuse. On the contrary, insidious disease onset, early age of onset, the presence of negative symptoms and poor compliance to treatment are factors linked to poor prognosis.

The outcome of schizophrenia is also influenced by socioeconomic variables. In the multinational World Health Organization study, patients in developing countries had a better outcome than patients in developed countries (Jablensky 1992).

1.5 Aetiology of schizophrenia

1.5.1 The neurodevelopmental model

The neurodevelopmental hypothesis of schizophrenia has been one of the predominant models in schizophrenia research over the past two decades. This model postulates that schizophrenia is the result of aberrative neurodevelopmental processes that commence long before the onset of clinical symptoms (Weinberger 1987, Murray & Lewis 1987) and are caused by genetic and environmental factors (presented in more detail in the following subsections) or a combination of both.

Converging evidence from a wide range of research studies supports the neurodevelopmental model. Patients affected with schizophrenia are more likely to have experienced pre- and/or perinatal adverse events (such as hypoxia during their birth or exposure to maternal virus when in utero) than healthy controls (Geddes & Lawrie 1995, Keshavan et al. 2006, Gilmore & Murray 2006, Byrne et al. 2007). Additionally, schizophrenia patients often exhibit increased rates of minor physical anomalies (Compton and Walker 2009, Ismail et al. 1998) and motor, cognitive and behavioural impairments (Weinberger 1996) long before illness onset (even present in childhood), thus further supporting the neurodevelopmental hypothesis. Subtle brain structural abnormalities have also been found in patients prior to illness onset (Lawrie et al. 2001b, Pantelis et al. 2003, Pantelis et al. 2005; to be discussed in more detail in Chapter 2).

To explain the gap between the early developmental insults and the onset of schizophrenia in adolescence or early adulthood, researchers have postulated that these early developmental abnormalities are likely linked to aberrant brain maturational processes (such as excessive synaptic pruning), possibly in synergy with some stressor, that ultimately lead to the development of psychotic symptoms and schizophrenia.

The controversy between a neurodevelopmental and neurodegenerative origin of

schizophrenia goes back to the years of Kraepelin and is still a matter of debate (Lieberman 1999, Woods 1998). However, the idea that these are not mutually exclusive and may both characterize the development and course of schizophrenia gains significant momentum nowadays.

1.5.2 Genetics

Genetic predisposition appears to be a major risk factor for the development of schizophrenia. A plethora of family, twin and adoption studies have examined the contribution of genetic factors to the development of schizophrenia. As already mentioned above, the lifetime risk for the development of schizophrenia is estimated around 1% in the general population, while having an affected family member substantially increases the risk for the disorder (McGuffin et al. 1995, Tandon et al. 2008). It is estimated that having a second or third degree relative increases the risk of developing schizophrenia to 2-6% (Gottesman et al. 1987), while a first-degree relative (e.g. sibling, parent) can increase the risk up to 15% (Tandon et al. 2008). A useful approach for elucidating the genetic underpinnings of vulnerability to schizophrenia is the study of monozygotic and dizygotic twin pairs who are discordant for schizophrenia (Sullivan et al. 2003). Dizygotic twins share on average 50% of their genes, so if one twin is affected, the other has 10-15% risk of developing schizophrenia while monozygotic twins discordant for schizophrenia, who share almost 100% of their genetic material have 40-50% risk of schizophrenia (Gottesman et al. 1987, Sullivan et al. 2003).

Despite the fact that the role of genetic risk is undisputed, it remains unclear which genes or genetic variations confer this risk for schizophrenia from one generation to the next. Converging evidence suggests that no single gene is sufficient or necessary for the development of schizophrenia but instead many, even thousands of genes might act in additive or interactive ways to mediate risk for the disorder (Walker et al. 2010). Findings from genome-wide association studies (GWAS) have implicated a large number of putative risk genes and susceptibility loci that may be linked to

abnormal neurodevelopmental processes associated with schizophrenia. Failure to replicate findings in genetic association studies has, however, seriously impeded the definite consideration of genes for their involvement in developing schizophrenia (Tandon et al. 2008). In one of the largest molecular genetic studies of schizophrenia that combined numerous GWAS genotypes into a single analysis, the authors found 108 independent genomic loci, some of which were not previously reported, that increase the risk for the development of schizophrenia (Ripke et al. 2014).

1.5.3 Environmental factors

A series of environmental factors have been postulated to convey an added risk for schizophrenia. Maternal virus infections, nutritional deficiencies during pregnancy and complications during birth delivery have been linked to an increased liability to schizophrenia (Maki et al. 2005, Tandon et al. 2008). Birth during winter or early-spring months (Torrey et al. 1997) and older paternal age have also been associated with a greater probability for developing schizophrenia (Malaspina et al. 2001). There is also evidence suggesting that urbanization and migration are significant risk factors for the development of schizophrenia (Boydell & Murray 2003, Hickling et al. 1999). Finally cannabis abuse during adolescence has been consistently linked to an increased risk for schizophrenia (Moore et al. 2007) that was estimated to be 2-times higher than the relative risk for schizophrenia on an individual level (Arseneault et al. 2004). Although a direct causational relationship between cannabis use and schizophrenia has not yet been found, it is postulated that cannabis can precipitate schizophrenia in vulnerable (or at-risk) individuals (Barnes et al. 2006). For a detailed review of the risk factors in schizophrenia, the interested reader is referred to Lawrie et al. 2011.

In general, it is highly likely that a combination of genetic and environmental factors interact to cause schizophrenia. It remains, however, elusive exactly which and how these factors are involved in the development of schizophrenia.

1.6 High-risk paradigms

As explained above, the onset of schizophrenia is usually preceded by a prodromal phase characterized by mild, sub-threshold positive symptoms and functional decline. Research into the early phases of psychosis may provide important clues to the pathophysiological mechanisms underlying transition to psychosis, without the confounding effects of anti-psychotic medication. Early intervention strategies during this prodromal stage are of great interest because they have been shown to provide better outcome and ameliorate the burden of psychosis (Marshall & Lockwood 2006).

Two research paradigms have been employed in order to explore vulnerability markers to psychosis. Utilising a clinical high-risk (HR) paradigm, individuals presenting with prodromal signs of psychosis are considered at an imminent, increased risk of making a transition to full psychosis (also often described as at an 'at-risk mental state' or ARMS) (Fusar-Poli et al. 2013a, Cornblatt et al. 2002, Phillips et al. 2000). In general individuals at clinical HR most often present with attenuated positive symptoms, brief limited intermittent psychotic symptoms (Riecher-Rossler et al. 2007, Yung et al. 2004), mild cognitive deficits (Brewer et al. 2006, Lencz et al. 2006, Riecher-Rossler et al. 2009) and/or might have a family history of psychosis with an accompanying reduction in functioning.

In familial HR paradigms, individuals are characterised on the basis of having at least one affected first or second-degree relative (Lawrie et al. 2001b, Cannon & Mednick 1993, Erlenmeyer-Kimling et al. 1997, van Haren et al. 2004). Schizophrenia is known to have a heritable component (Harrison & Weinberger 2005). Thus by examining individuals from multiply affected families diagnosed with schizophrenia, it is possible to evaluate whether and to what extent brain, clinical and cognitive abnormalities are mediated by a genetic risk to the disorder.

It should be noted, however, that this delineation of HR subjects is not straightforward as many individuals at familial HR often exhibit pre-psychotic symptoms and clinical HR subjects might also have a family history of the disorder.

A recent systematic review of clinical and familial HR cohorts reported overlapping brain changes in the frontal, temporal and cingulate regions (Smieskova et al. 2013). In addition, individuals in both HR paradigms also share a higher probability of developing schizophrenia than a healthy population, suggesting that several aspects of their risk might represent vulnerability markers for the development of the disorder.

1.7 Aims and outline of the present thesis

The overriding aim of the present thesis was to apply machine learning in order to identify predictors of transition to schizophrenia in subjects at high risk for developing the illness. Initially, data from the Edinburgh High Risk Study (EHRS), which were immediately available, were used with the intention of identifying neuroanatomical-based markers of transition and also replicating earlier findings in the EHRS literature (Johnstone et al. 2005). The second aim was to examine if our classification technique could generalize to other high risk cohorts and especially to high-risk individuals that were identified on the basis of having prodromal symptoms (clinical HR cohorts). To this aim, baseline neuroanatomical data from the FePsy (Fruherkennung von Psychosen) study (Riecher-Rossler et al. 2007) were obtained. Finally, the prognostic performance of our classifier was tested by pooling data from the EHRS and the FePsy datasets.

Overall the thesis spans two main areas of study; neuroimaging and machine learning. Therefore, relevant concepts from the fields of neuroimaging and machine learning are presented in Chapters 2 and 3, respectively. Chapter 2 gives an introduction to Magnetic Resonance Imaging (MRI), including an overview of the pre-processing steps and standard univariate methods of analysis. A review of MRI findings in schizophrenia is additionally presented. Chapter 3 provides an introduction to machine learning approaches and the steps of a machine learning-based classification pipeline. In addition, a literature review of machine learning studies in schizophrenia is presented. Chapter 4 presents the subject material in the

present thesis and describes the methodology used. Subsequent chapters constitute the main experimental sections of the present thesis. Chapter 5 addresses the first aim of the thesis and presents the results of applying our classification technique in distinguishing at baseline those high-risk subjects that later developed schizophrenia from those who did not, yet had psychotic symptoms. Chapter 6 addresses the second aim and presents results pertaining to applying the classifier to baseline data from the FePsy study. Chapter 7 addresses the final aim of the thesis and examines the performance of the classifier on the pooled dataset. Finally, Chapter 8 comprises an overall summary of the main findings presented in this thesis and a general discussion of limitations and considerations for future work.

CHAPTER 2

Background: Brain Imaging

2.1 Introduction

Recent advances in neuroimaging have enabled scientists to visualize and study the human brain *in vivo* and develop methods to understand its anatomy and function. Commonly used neuroimaging modalities include X-ray computed tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI). The work presented in this thesis will focus on MRI, which is described in the following subsections, after a brief introduction to human brain anatomy.

2.1.1 Neuroanatomy

The human brain is broadly divided into three parts: the cerebrum, the cerebellum and the brain stem. The cerebrum is the largest part of the brain and is composed of the cerebral cortex and several subcortical structures, such as the hippocampus and basal ganglia. The cerebrum is divided into two hemispheres, left and right, which are connected together by a flat bundle of neural fibers, the *corpus callosum*. Outermost to the cerebrum lies the cerebral cortex, which is divided into four lobes: frontal, parietal, occipital and temporal (Figure 2.1).

In general, each lobe in the cerebral cortex specializes in different functions. The frontal lobe, which lies at the front of the brain and behind the forehead, is responsible for reasoning and planning. It also plays an important role in long-term memory. Posterior or caudal (behind) to the frontal lobe, and dorsal (superior) to the occipital lobe is the parietal lobe, which is primarily responsible for visuo-spatial processing, recognition and navigation. Ventral (inferior) to the parietal lobe is the temporal lobe, which is mainly responsible for auditory processing and is associated with memory and speech. Finally, the occipital lobe, which lies at the back of the cortex, is the visual processing center of the brain.

At the back of the brain, underlying the occipital and temporal lobes is the cerebellum, also known as 'the little brain', which is responsible for movement coordination and balance.

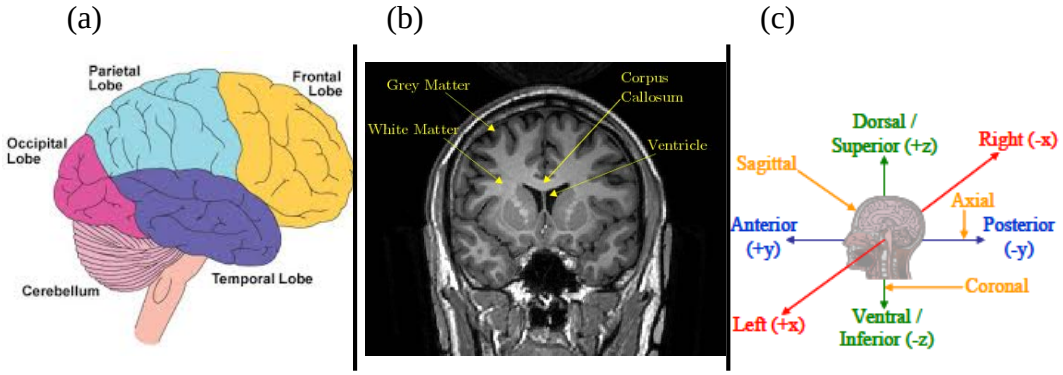


Figure 2.1 The human brain and anatomical planes and directions. (a) A sagittal view of the human showing the four lobes and cerebellum. (b) An axial slice of a structural MRI scan highlighting the grey matter, white matter and the ventricles (c) The anatomical directions and planes in current medical terminology.

The cerebral cortex has a typical grey color due to grey matter (GM), which is also distributed at the depths of the cerebrum (e.g. thalamus, hypothamalus etc.), the cerebellum and the spinal cord. The grey matter consists mainly of neuronal cell bodies that are responsible for neural processing and several functions. In contrast, white matter (WM) consists mostly of glial cells and myelinated axon tracts that transmit information (i.e. signals) from one part of the brain to the other and act supportively to the function of neurons (e.g. by providing nutrients to the neurons etc.). At the center of the brain are the ventricles, which are filled with cerebrospinal fluid (CSF) that facilitates the transmission of several substances across brain regions and acts like a 'cushion' to the cortex, by protecting it and absorbing pressure.

2.1.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) utilizes the phenomenon of nuclear magnetic resonance (NMR) of the hydrogen atom in order to produce high-quality, detailed images of internal body structures and any other tissue.

In order to obtain a structural MRI, the patient is placed inside a MRI scanner. A MRI scanner is a cylinder-shaped magnet that generates a powerful static magnetic field. The strength of the magnetic field determines to a great degree the quality of the produced images. Scanners currently used can generate magnetic fields that range from 1T to 4T.

MRI exploits the spin of the hydrogen nuclei. In the presence of a static magnetic field B_o, protons of the hydrogen nuclei tend to align parallel or anti-parallel to the direction of the magnetic field and rotate along the axis. A slightly larger number of protons spin (or *precess*) in the direction that is parallel to the field and have less energy than those that are anti-parallel, thus creating a net longitudinal magnetic field in the z-direction (Figure 2.2a).

When a radiofrequency (RF) pulse is introduced and emitted at the precession frequency of the hydrogen atom, lower-energy protons are excited and jump to a higher-energy state, where they align anti-parallel to the field and precess in phase. A new transversal magnetic field is thus created in the (x,y)-plane (Figure 2.2). When the RF pulse is then switched off, the excited protons gradually release their previously absorbed energy and return to the initial lower-energy state in their parallel orientation (*relaxation phase*). This results in a gradual increase in the longitudinal magnetic field, which is characterized by the longitudinal T1 relaxation time. In addition, in the absence of the RF pulse, protons begin to precess again in different phases. This causes a gradual decline in the net transversal magnetic field, which is characterized by the T2* relaxation time.

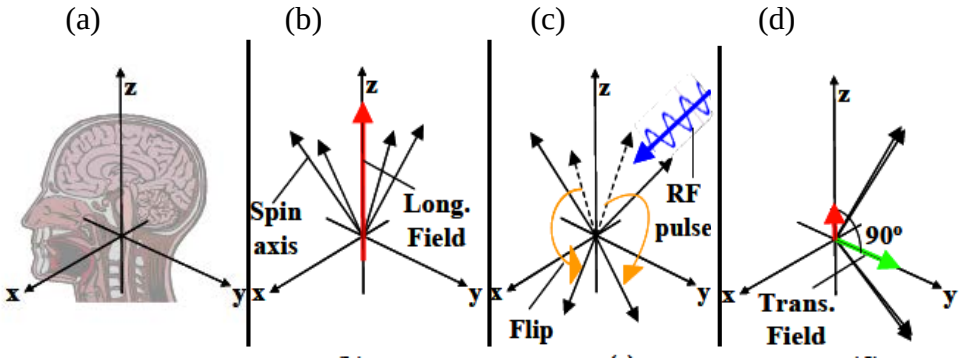


Figure 2.2 Behavior of hydrogen atom after an RF pulse. (a) The system of conventional axes: z–axis is defined by the line that connects toe to head of the subject. The xy–plane is parallel to the z–axis. (b) The net longitudinal magnetic field in the +z direction, before the introduction of the RF signal. (c) The application of an RF pulse flips some protons to the –z direction and this reduces the longitudinal magnetic field. (d) Application of RF pulse makes protons rotate in phase, creating a transversal (in xy–plane) magnetic field.

The construction of contrast images is based on the differing T1 and T2* relaxation properties of various brain tissues. In most structural MRI studies, a T1 contrast is used to obtain the images, while a T2* contrast is used for functional MRI.

Two imaging parameters are of great importance in MRI: the time to repeat (TR) and the time to echo (TE). TR represents the time interval between two consecutive applications of the RF pulse and TE the time lapse from the moment the RF pulse is applied and the moment when the signal is measured. For a more thorough description of the MRI one can refer to Hashemi et al., 2004.

2.1.3 Functional MRI

Functional MRI (fMRI) is used to visualize brain activation in response to an external stimulus or a certain task. Functional MRI builds on the same technology as MRI, with the difference that it exploits the ratio of oxygenated to deoxygenated blood instead of the hydrogen atom. It is considered an indirect measure of brain

activity, since fMRI does not capture neural activity per se but the hemodynamic response instead (Huettel et al. 2004, Lazar 2008).

When a brain region becomes active, oxygenated haemoglobin flows to that area in order to increase blood concentration. What is measured in an fMRI task is the difference of oxygenated to deoxygenated haemoglobin. Deoxygenated haemoglobin is more paramagnetic than oxygenated haemoglobin (the atoms of a paramagnetic material tend to align themselves in the presence of an external magnetic field, thus increasing the field strength). This difference induces change in the local magnetic field and thus affects the MR signal through the Blood Oxygenation Level Dependent (BOLD) contrast (Ogawa et al., 1990). The BOLD signal gives an indication of neuronal activity.

2.2 Image Processing

The present work is based on the analysis of brain imaging data that were acquired using MRI. Each MRI brain scan consists of volume units called *voxels*, in which an aspect of a specific function or structure is recorded.

In order to achieve spatial correspondence of voxels between subjects, normalization of brain scans to a common standard space is required. In addition, a certain number of pre-processing steps are usually performed before analysing the data and making inferences for specific effects. The scientific question to be investigated and the type of task at hand generally determine the processing procedures to be followed.

In this work, structural MRI data were pre-processed by employing voxel-based morphometry. A general overview to the voxel-based morphometry theory and methodology is given below. More details about specific steps (and parameters) followed in the pre-processing of the MRI scans in this thesis can be found in Chapter 4 - Materials & Methods.

2.2.1 Voxel-based Morphometry

The aim of voxel-based morphometry (VBM) is to identify differences in the local concentration of brain tissue (usually GM), after correcting for large-scale differences in gross anatomy and position (Ashburner & Friston, 2000, Mechelli et al. 2005). VBM involves spatially normalizing all the structural images to the same stereotactic space, extracting GM and WM from the normalized images, smoothing the normalized GM and WM segments and finally performing voxel-wise statistical comparisons to make inferences about group differences (Figure 2.3).

One important aspect of VBM is that it is not restricted to a specific brain region, like a region-of-interest (ROI) analysis that requires *a priori* assumptions, but examines the whole brain in an unbiased and objective manner.

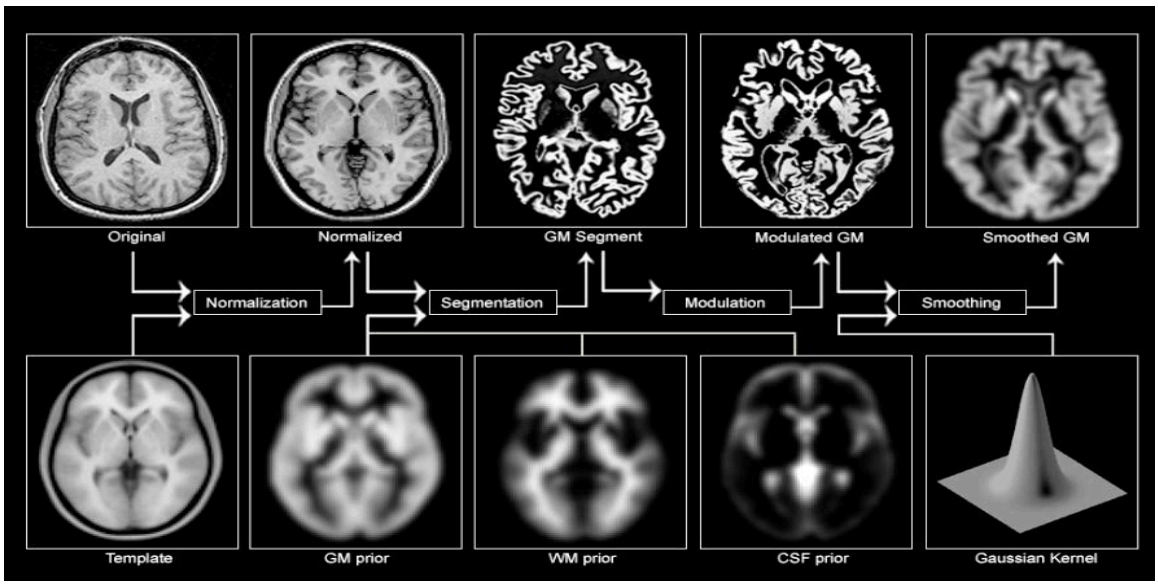


Figure 2.3 Overview of VBM pre-processing stream.

2.2.1.1 Spatial Normalization

Spatial normalization involves the registration of all images to the same template image, by minimizing the residual sum of squared differences. The first step in spatially aligning each brain image involves matching the image by estimating the optimum 12-parameter affine transformations (Ashburner & Friston 1997). The second step accounts for global non-linear shape differences, which are modeled by a linear combination of smooth spatial basis functions. This step involves estimating the coefficients of the basis functions that minimize the residual squared difference between the image and the template, while simultaneously maximizing the smoothness of the deformations. Spatial normalization does not aim to match every cortical feature exactly but merely correct for global brain differences.

2.2.1.2 Segmentation

The spatially normalized brain images are then segmented into grey matter, white matter and cerebrospinal fluid. Segmentation employs a mixture model cluster analysis to identify voxel intensities matching particular tissue types combined with an *a priori* knowledge of the spatial distribution of these tissues in normal subjects, derived from tissue probability maps. The segmentation step also incorporates an image intensity non-uniformity correction (Ashburner & Friston, 2000) to address image intensity variations caused by different positions of cranial structures within the MRI head coil.

A unified segmentation approach has also been adopted in currently used pre-processing streams (Ashburner & Friston 2005). Unified segmentation is a probabilistic framework in which image registration, tissue classification and intensity non-uniformity (bias) correction are all combined in the same generative model. In this way, the inherent circularity found in standard VBM approaches, where image registration requires an initial tissue classification and in turn tissue classification requires an initial registration, is here resolved.

2.2.1.3 Smoothing

The normalized, segmented images are then smoothed using an isotropic Gaussian kernel. The intensity in each voxel of the smoothed data is a locally weighted average of grey matter (or white matter) density from a region of surrounding voxels, the size of the region being defined by the size of the smoothing kernel (Ashburner & Friston, 2000). Smoothing also renders the data more normally distributed, thus increasing the validity of the following parametric statistical tests.

Additional pre-processing steps can be included in the standard VBM processing pipeline. Creation of GM and WM study specific templates and skull extraction can be performed prior to the initial normalization step so that errors pertaining to misclassification of tissue can be avoided.

In the optimized protocol suggested by Good et al. 2001, study-specific templates for GM and WM are created by averaging the smoothed normalized GM and WM images from all subjects in the study. Then an iterative procedure of normalization and segmentation steps is followed. First, the original structural MRI images in native space are segmented. The resulting GM and WM images are then spatially normalized using the GM/WM study-specific templates to derive the optimized normalization parameters. These parameters are then applied to the original, whole-brain structural images in native space before a second segmentation step takes place.

Details for the creation of study-specific priors and the exact steps of the pre-processing pipeline that were followed in this work can be found in Chapter 4 of Materials & Methods.

Following the pre-processing of structural MRI data, voxel-wise statistical analysis can then be performed.

2.2.2 Univariate methods of analysis

In univariate methods of analysis, each voxel is treated independently in order to examine an effect of interest. Parametric statistical models are assumed at each voxel, using the General Linear Model (GLM) in order to describe the data as a linear combination of experimental effects, potentially confounding variables and some error (Friston et al. 1995). Classical statistical inference is then used to test hypotheses that are expressed in terms of GLM parameters. Inferences in neuroimaging settings may be about structural differences when comparing one group of subjects to another (e.g. voxel-based morphometry- Ashburner & Friston 2000), or neurophysiological indices of brain function (fMRI- Friston et al. 1994).

2.2.2.1 Statistical Parametric Mapping

Statistical Parametric Mapping is the most prevalent approach to characterizing effects pertaining to structural and functional anatomy. Two steps are involved in the characterization of regionally specific effects: i) modelling the data, i.e. estimating parameters using a GLM and ii) making inferences via *t*- or *F*-statistics.

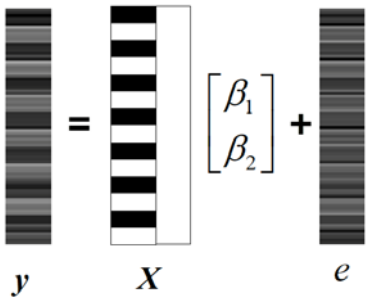


Figure 2.4 General Linear Model. The General Linear Model assumes that the acquired signal intensities in Y are the combination of three components: the design matrix **X** that contains the actual data for the two groups of interest along with confounds that model some source of noise. A weight assigned to each component of interest (β_1) or confound (β_2) and some error **e.** *Source: SPM-Zurich course notes.*

A GLM model that describes the signal intensities in voxel **Y** can be expressed in terms of some experimental parameters **B** and some residual variability (error) **E**. For characterising group differences in this voxel, this can be expressed as: $\mathbf{Y} = \mathbf{XB} + \mathbf{e}$, where **X** is the design matrix that contains the data for the two groups (Figure 2.4).

Parameters in **B** are determined for each voxel either using students *t*-Test or analysis of variance (ANOVA). Following the estimation of parameters, *t*- or *F*-scores are constructed from contrasts, which basically define the hypothesis (i.e. the scientific question) to be investigated. Statistical scores are then compared to the expected distribution under the null hypothesis, allowing the computation of a p-value for each voxel. P-values should be corrected for multiple comparisons to ensure that voxels did not reach statistical significance levels by chance on account of the large number of voxels that are treated independently. Standard Bonferroni correction could be used to control for multiple comparisons. However, given that it requires all tests to be independent, this adjustment is too harsh and possibly inappropriate in neuroimaging settings where voxels are highly correlated. In contrast, Random Field Theory (RFT) takes into account that contiguous voxels are not independent (Brett et al. 2003, Worsley et al. 1996, Worsley et al. 2004) and provides a more parisimonious approach to controlling for type I errors.

Type I errors occur when the null hypothesis is rejected when in fact it is true (also known as *false positive* error). For example, a brain region (or voxel) is shown to be activated during a particular task when it is actually not or a difference in experimental conditions is detected but such a difference does not truly exist. On the other hand, a type II error (or *false negative* error) occurs when the null hypothesis (e.g. no difference between two experimental groups or no voxel activation during a task) is accepted when it is actually false. While controlling for type I errors is a widely accepted practice in neuroimaging studies (through Bonferroni correction or RFT), type II errors are more rarely considered (Bennett et al. 2009). Although, the focus on controlling for type I errors is clearly necessary in that it ensures that most observed effects represent true differences, it could, also, result in a number of negative consequences, such as an increase in type II errors and deficient meta-analyses (Lieberman and Cunningham 2009). The key in neuroimaging studies is to

find a balance between the two types of error; a tricky endeavour in which one must consider the desired power and the purpose or nature of the study. This balance could be achieved by increasing the sample size in single neuroimaging studies, which is not always cost-effective and easy to perform, or by considering studies in the aggregate through systematic meta-analyses that can compensate for the low statistical power. Increasing sample size will result in less conservative thresholds being needed to control for false positives and provide more statistical power to detect the more subtle differences and as a result this will reduce false negative errors as well. Another possible solution would be to restrict the amount of multiple corrections to be performed by either focusing on specific regions of interest or by means of small volume corrections (SVC) that would allow for more liberal p-values.

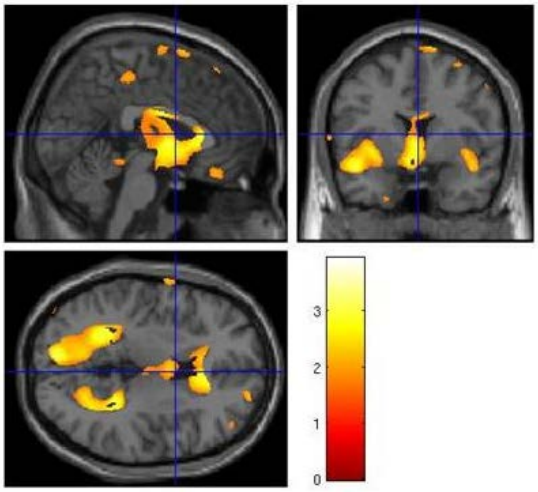


Figure 2.5 Example of an unthresholded T-map. Axial, coronal and sagittal views of T-map overlaid over a single subject anatomical image.

Finally, the results of statistical parametric analyses can be illustrated in 3D maps, often called SPM maps, where statistically significant differences in brain voxels can be highlighted (Figure 2.5).

In terms of neuroimaging packages, SPM software is currently the most prevalent one for the pre-processing and analysis of brain imaging data. It was first made available to the neuroimaging community in 1991 (Friston et al. 1991) and has been available ever since from the Wellcome Trust Centre for Neuroimaging at University College London (http://www.fil.ion.ucl.ac.uk/spm), along with detailed documentation and related publications. The latest version of SPM released is the SPM12. However, in this work I have mainly used SPM5 because I found this version to be relatively more stable to work with than the other SPM versions.

2.3 MRI findings in schizophrenia

Since the time of Kraepelin and Bleuler, it was strongly believed that schizophrenia was an organic brain disorder. Since then, brain imaging studies have attempted to unravel the pathophysiology of schizophrenia and reported a host of structural and functional brain abnormalities associated with the disorder.

2.3.1 Neuroanatomical abnormalities in established and first-episode schizophrenia

The first CT study of schizophrenia (Johnstone et al. 1976) reported abnormally enlarged lateral ventricles in patients with schizophrenia; a finding that has been highly replicated in most subsequent MRI studies (Shenton et al. 2001). Early MRI studies have also reported significant reductions in total brain volume in schizophrenia patients compared to healthy individuals (Shenton et al. 1997). In Lawrie & Abukmeil 1998, the authors reported about 3% global brain volume reduction in patients, mainly attributed to grey matter volume loss (of about 4%). In an extensive meta-analysis of regional brain volume studies in schizophrenia, Wright et al. 2000 reported that the mean cerebral volume of patients was 2% smaller than the mean volume for healthy controls in 58 studies that involved 1,588 schizophrenia patients.

Temporal and frontal lobe volume reductions have also been consistently reported in studies comparing schizophrenia patients and healthy controls that employed either ROI or VBM (Lawrie & Abukmeil 1998, Wright et al. 2000, Shenton et al. 2001, Zakzanis et al. 2000, Honea et al. 2005, Haijma et al. 2013). Medial temporal lobe structures, particularly the hippocampus (Velakoulis et al. 2006), amygdala (Lawrie & Abukmeil 1998, Nelson et al. 1998) and superior temporal gyrus (Honea et al. 2005, Sun et al. 2009b), were found to be greatly reduced among patients. In a recent meta-analysis conducted by Honea et al. 2005, 50% of the studies included reported grey matter deficits in schizophrenia in the left superior temporal, parahippocampal and inferior frontal gyrus. Findings in the parietal and occipital lobes have been less consistently replicated in schizophrenia research. Moderate evidence for abnormalities in the parietal lobe (in 60% of studies reviewed) has been reported in a review conducted by Shenton et al. 2001. A certain number of VBM studies and meta-analyses have reported volumetric reductions in parts of the parietal or occipital lobes (Honea et al. 2005, Ellison-Wright et al. 2008, Hulshoff Pol et. al 2001). Divergent evidence has been reported for findings concerning the cerebellum and the anterior cingulate. Two recent meta-analyses have reported decreased volume in the anterior cingulate in schizophrenia patients (Baiano et al. 2007, Fornito et al. 2009), whereas other studies reported increased volume in that region (Kopelman et al. 2005, McCormick et al. 2005). Similarly, some studies have reported volumetric deficits in the cerebellum in schizophrenia patients as compared to healthy individuals (Honea et al. 2005, Ellison-Wright et al. 2008), while others have reported grey matter increase in the area (Wilke et al. 2001, Suzuki et al. 2002).

Diversity in findings in schizophrenia studies might be partly attributed to methodological differences in the implementation of the pre-processing approaches selected and the choice of parameters in the VBM pipeline. Specifically it has been suggested that the smoothing kernel and the choice of statistical analysis (either voxel-level or cluster-level significance) can significantly impact results (Honea et al. 2005).

Differences in patient groups can be another possible explanation for the controversies in the structural abnormalities reported. Patient groups may vary with

respect to age, anti-psychotic treatment and/or treatment duration, symptom severity, presence of comorbidity and substance use.

Traditional meta-analyses, that pool together statistical findings from individual research studies are invaluable in summarizing the effect size of research findings and also identifying and explaining the heterogeneity in findings across different studies that may be due to the plethora and diversity of the participating cohorts. A recent meta-analysis conducted by Bora et al. 2011 reported that GM abnormalities in the bilateral insula, superior temporal and anterior cingulate gyrus and the thalamus were more widespread in studies consisting of more males than females or where more chronic cases of schizophrenia, with more severe negative symptoms were included.

Prospective meta-analysis efforts, such as those performed by the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium (Thompson et al. 2014) have the added benefit of standardizing the analyses across individual sites and thus promoting their consistency rather than the ad hoc aggregation of statistical results. The ENIGMA is one of the largest consortia to date that aims to integrate and meta-analyze data sets across the world in order to delineate the neurobiological, clinical and genetic underpinnings of various psychiatric disorders. In schizophrenia, the ENIGMA-Schizophrenia Working Group provides a framework that ranks brain measures based on their effect sizes for the comparisons between patients and healthy controls. In a recent meta-analysis of data from 2,028 patients and 2,540 controls, the ENIGMA-Schizophrenia Working Group found that deficits in hippocampal volume had the largest effect size in differentiating between patient and controls (van Erp et al. 2015), followed by deficits in the amygdala, thalamus and accumbens. Significant positive associations were also found between duration of illness and also age, with putamen and pallidum volume increases in schizophrenia patients.

Other studies tried to link the structural brain alterations to distinct clinical syndromes in schizophrenia, in an effort to delineate the neurobiological

underpinnings of clinical heterogeneity in schizophrenia. It was suggested that schizophrenia syndromes can be broadly subdivided into three different nosological profiles, each of which highlight the prevalence of either positive, negative or disorganized symptoms (Buchanan & Carpenter 1994, Andreasen et al. 1995, Grube et al. 1998). A significant number of studies have employed either ROI (Flaum et al. 1998, Gur et al. 2000, Crespo-Facorro et al. 2004) or VBM (Sigmundsson et al. 2001, Koutsouleris et al. 2008, Nenadic et al. 2010) approaches in order to evaluate whether these three symptom domains engage different patterns of structural abnormalities, thus linking symptoms to different neural surrogates. Specifically, Koutsouleris et al. 2008 reported that the three clinically defined schizophrenia subgroups possibly share common structural abnormalities compared to healthy controls in prefrontal and perisylvian structures but specific and extended GM density reductions affecting the medial prefrontal, limbic and temporal cortices bilaterally were observed in the disorganized symptom dimension. Positive symptoms were associated with pronounced alterations in perisylvian regions of the left hemisphere and extended GM density reductions in the thalamus while negative symptoms were mostly linked to extended alterations within orbitofrontal, medial prefrontal, lateral prefrontal and temporal cortices as well as limbic and subcortical structures (Koutsouleris et al. 2008). A more recent study employing an advanced morphometric analysis method that provides superior sensitivity and specificity compared to VBM approaches by accounting for the interrelatedness of spatial information, confirmed the prefronto-perisylvian grey matter reduction pattern found in the schizophrenia group as a whole, but also revealed marked brain volume changes within the three subgroups, including pronounced reductions in the cerebellum in the negative symptom dimension, marked GM volume losses in the ventro-medial prefrontal and occipitotemporal cortex associated with positive symptoms (Zhang et al. 2015). The disorganized symptom dimension presented a relative preservation of GM compared to the other symptom dimensions (Zhang et al. 2015).

Similar structural abnormalities to those described above in schizophrenia patients have also been identified in groups of patients in the first episode (FE) of the disorder (Kubicki et al. 2002, Steen et al. 2006). First-episode refers to individuals who have experienced their first psychotic episode, and are thus in the early phase of schizophrenia, and is operationally defined as the first psychiatric hospitalization (Lawrie et al. 2001b, DeLisi et al. 1991, Chua et al. 2003) or first contact with psychiatric services or even the first administration of anti-psychotic treatment (Cahn et al. 2002, Zipursky et al. 1998). In general most MRI studies corroborated a similar network of regions being affected in FE as in the chronic schizophrenia (Shenton et al. 2001). Specifically, MRI-based studies reported deficits in total brain volume (Vita et al. 2006, Steen et al. 2006), GM volume reductions in temporal and prefrontal areas including the anterior cingulate gyrus (Kubicki et al. 2002) and the thalamus (Meisenzahl et al. 2008a, Watson et al. 2012), volumetric deficits in the hippocampus (Velakoulis et al. 2006, Vita et al. 2006, Meisenzahl et al. 2008a, Watson et al. 2012, Steen et al. 2006) and an enlargement of the lateral ventricles in FE patients compared to healthy controls (Steen et al. 2006, Vita et al. 2006).

These structural abnormalities are however less pronounced in FE compared to the established state, possibly suggesting active disease processes around the time of onset, although genetic factors, substance misuse, antipsychotic drug treatment and other factors may be partly responsible (Meisenzahl et al. 2008a, Olabi et al. 2011). Additionally, the fact that more extended brain alterations are observed in chronic schizophrenia compared to FE suggests that additional cortical and sub-cortical brain regions become involved in the disease process in the advanced stages of schizophrenia (Meisenzahl et al. 2008a).

However, there is no consensus in findings in FE schizophrenia studies with regards to the brain regions involved. Kubicki et al. 2002 did not report any reductions in medial temporal lobe structures and interestingly enough, a meta-analysis by Vita et al. (2006) did not confirm volumetric reductions in the temporal lobe and amygdala in FE patients.

Despite any controversies in structural findings, the study of FE schizophrenia is useful because it allows the detection of regional effects of the disease at the time of

onset and represents a useful tool in assessing hypotheses about progressive brain changes in the longitudinal course of schizophrenia.

Converging evidence reveals that ongoing brain volume changes in schizophrenia might not be limited to the early, first-episode phase of the disorder but progress over the course of the illness. Longitudinal studies of schizophrenia have suggested continuous progressive lateral ventricle increases (Van Haren et al. 2008a, DeLisi et al. 2004, Hulshoff & Kahn 2008, Olabi et al. 2011), progressive whole-brain volume loss (Van Haren et al. 2008a) and brain tissue volume decreases, especially in frontal (DeLisi et al. 2004) and temporal GM volume (Hulshoff & Kahn 2008) in chronic patients with schizophrenia compared to healthy individuals. Ongoing volume reductions in the right caudate and thalamus have also been reported (Van Haren et al. 2007). No progressive volumetric changes in chronically ill schizophrenia patients in the hippocampus were found (Wood et al. 2001). More widespread and pronounced alterations in brain tissue were found to be associated with poor outcome, severe negative symptoms and more pronounced neuropsychological impairment (Hulshoff & Kahn 2008). Specifically, progressive ventricular enlargement was found to be associated with poorer outcome (Lieberman et al. 2001, Ho et al. 2003), greater frontal and temporal volume reductions were correlated with less improvement in negative symptoms (Gur et al. 1998, Mathalon et al 2001, Kasai et al. 2003) and poorer neurocognitive and executive functioning was associated with more pronounced frontal and parietal GM volume reductions (Lieberman et al. 2005, Hoet al. 2003) as patients progressed from their first episode to the chronic state.

The ascertainment of progressive brain changes is of fundamental importance as to whether schizophrenia is a neurodevelopmental or neurodegenerative disorder. The fact that ongoing brain alterations continue over the course of the illness in chronic patients suggests that one or more active pathophysiological processes take place during the illness. Identification of those pathophysiological processes is of high clinical relevance as it could advise strategies to stop or reverse the disease process and possibly provide better outcome. In this direction, longitudinal studies of healthy controls are highly relevant in order to distinguish normal brain changes, possibly associated to age, gender or other factor, from aberrant alterations associated with

schizophrenia.

A potential confounding factor in most schizophrenia studies is the effect of anti-psychotic medication since it is very difficult to establish whether (progressive) brain volume changes are a result of anti-psychotic treatment (or not) and to what extent (Van Haren et al. 2008b). Volume increases in the basal ganglia, and specifically the caudate nucleus, have been consistently linked to antipsychotic treatment (Chakos et al. 1994, DeLisi et al. 2006, Keshavan et al. 1994). Cumulative dose of anti-psychotic drug treatment was linked to progressive cerebral GM volume deficits, particularly in the superior medial frontal gyrus (Cahn et al. 2002) and also progressive decreases in frontal lobe volumes (Madsen et al. 1999). Another study compared the effect of typical (haloperidol) and atypical (olanzapine) anti-psychotic medication and reported progressive whole-brain GM volume loss in both first-episode and chronically ill patients receiving haloperidol compared to olanzapine, where no such effect was found (Lieberman et al. 2005). Administration of typical and atypical anti-psychotic medication seems to affect the basal ganglia volume differently as well, as volume decreases were reported when changing to atypical antipsychotic treatment (Chakos et al. 1995, Scheepers et al. 2001).

2.3.2 Neuroanatomical abnormalities in high-risk individuals

In order to minimize the confounding effect of anti-psychotic medication and shed light to the nature and extent of pathophysiological processes underlying schizophrenia researchers have begun to study anti-psychotic naive individuals at imminent risk of developing the disorder either due to sub-threshold clinical symptoms (clinical HR paradigms- Fusar-Poli et al. 2013) and/or increased genetic liability (genetic HR) (Cannon et al. 2005). A detailed presentation of the two HR paradigms can be found in Chapter 1.

Converging evidence suggests that baseline structural abnormalities, qualitatively similar to the established state, are already evident before the onset of schizophrenia

albeit to a lesser extent. The identification of neuroanatomical abnormalities already present in HR cohorts would allow the evaluation of correlates of vulnerability to psychosis, possibly reflective of its neurodevelopmental origin, and serve as a basis for the further distinction of markers of transition to psychosis (Bois et al. 2015, Borgwardt et al. 2011). Clinical high-risk studies employing VBM methodology have reported structural abnormalities in frontal, lateral temporal, medial temporal and limbic regions already present in HR subjects compared to healthy individuals (Borgwardt et al. 2007a, Borgwardt et al. 2007b, Borgwardt et al. 2008, Meisenzahl et al. 2008b, Mechelli et al. 2011, Dazzan et al. 2012). In Job et al. 2003, the authors found GM density reductions in the anterior cingulate (bilaterally) and left parahippocampal gyrus in individuals at HR for familial reasons against healthy controls.

A recent voxel-based meta-analysis concluded that individuals at HR, for both clinical and familial reasons, showed reduced GM volume in the right superior temporal gyrus, left precuneus, left medial frontal gyrus, right middle frontal gyrus, bilateral parahippocampal, hippocampal regions and bilateral anterior cingulate (Fusar-Poli et al. 2011) compared to healthy control subjects. Another meta-analysis of whole-brain VBM studies employing familial HR cohorts found GM reductions in the right superior frontal gyrus, left insula, thalamus and putamen but surprisingly increased GM volumes in the left medial frontal gyrus (Cooper et al. 2014).

A plethora of studies have also examined the high risk individuals, who go on to develop psychosis (HR-T) against individuals that do not make a transition (HR-NT) with the aim of identifying the neuroanatomical markers associated with transition to schizophrenia, and other related psychoses. Neuroanatomical deficits in HR subjects who later developed psychosis were found in clusters of brain regions spanning the frontal, orbito-frontal, temporal and medial temporal lobe structures, the cerebellum and the cingulate gyri (Borgwardt et al. 2007a, Borgwardt et al. 2007b, Borgwardt et al. 2008, Pantelis et al. 2003, Koutsouleris et al. 2009a, Mechelli et al. 2011, Dazzan et al. 2012). The meta-analysis conducted by Fusar-Poli and colleagues (2011) concluded that HR-T individuals showed GM volume reductions in the right inferior

frontal and the right superior temporal gyrus at baseline, compared to HR-NT subjects. Another meta-analysis reported subtle GM volume decreases in the cingulate, insular and prefrontal cortex and the cerebellum (Smieskova et al. 2010) occurring in HR-T subjects.

Longitudinal studies have also focused on the progressive changes underlying the onset of psychosis. In a study conducted by Pantelis et al. (2003), the authors reported continuous GM volume reductions in the left parahippocampal, fusiform, orbitofrontal and cerebellar cortices and the cingulate gyrus in HR-T individuals compared to HR-NT. Another study found longitudinal volume reductions in the superior frontal, orbitofrontal, inferior temporal medial and superior parietal cortex and in the cerebellum (Borgwardt et al. 2008). In the only genetic HR study, the authors reported progressive GM density reductions in the left inferior temporal gyrus, uncus and the right cerebellum (Job et al. 2005).

2.4 Functional MRI findings in schizophrenia

Functional MRI studies in schizophrenia have focused on identifying a network of brain regions that are responsible for abnormal functioning in the disorder. The goal behind these studies is to demonstrate how failure to activate a network of brain regions (and/or an over-activation of another network) leads to behavioural and cognitive deficits related to the illness. Functional MRI tasks in schizophrenia include motor, working memory, verbal learning and memory, emotion processing, and decision-making tasks. Several studies have examined the fronto-temporal connectivity, a circuitry that is considered to play an essential role in executive functioning, learning and memory and reported abnormal activation in a network of brain regions, particularly implicating the prefrontal cortex (Meyer-Lindenberg 2010, Fusar-Poli et al. 2007) and connectivity from it to the rest of the brain (Lawrie et al. 2002).

The majority of fMRI studies have demonstrated a reduced connectivity in patients in all subgroups of schizophrenia (chronic schizophrenia and first-episode patients, high-risk individuals) relative to healthy control subjects (Petterson et al. 2011) and an involvement of frontal brain regions with thalamic and cerebellar regions (Whalley et al. 2005) or temporal regions (Lawrie et al. 2002). A reduced activation in the prefrontal cortex and in the anterior cingulate was observed in schizophrenia patients compared to healthy controls in a meta-analysis of 41 executive fMRI tasks (Minzenberg et al. 2009) and in a working memory task (Glahn et al. 2005).

A recent meta-analysis by Fusar-Poli and colleagues (2007) found that neurophysiological abnormalities in the prefrontal cortex, particularly involving the dorsolateral, ventrolateral and anterior prefrontal cortex, are present in HR subjects. Similar functional dysfunctions are also found in first-episode patients to a greater extent (Fusar-Poli et al. 2007). Another meta-analysis focusing on genetic HR cohorts, reported increased activation in the right posterior superior temporal gyrus and hypo-activation of the left thalamus and cerebellum in HR subjects compared to controls (Cooper et al. 2014).

In a cross-sectional study of familial HR individuals, increased activation in the left parietal lobe and decreased activation in the anterior cingulate cortex was observed in HR subjects that subsequently developed schizophrenia relative to healthy controls (Whalley et al. 2006). Compared to HR-NT, HR-T individuals demonstrated smaller increases in activation with increasing task difficulty in the right lingual gyrus and bilateral temporal regions (Whalley et al. 2006).

2.5 Conclusions

The advent of MRI has provided promising leads towards the identification of the neurobiological processes underlying schizophrenia. In general, it appears that neuroanatomical and neurophysiological abnormalities occur at different stages of the disorder and progress as patients transit to a more chronic state of the illness. Active neurobiological alterations might also take place before the onset of schizophrenia; some of which may be specifically linked to a later onset of psychosis, as opposed to an increased vulnerability. The identification of markers of transition to psychosis is of high clinical relevance because early intervention strategies could provide a significant effect on clinical outcome.

However, the identification of neuroanatomical and neurophysiological markers in schizophrenia warrants some degree of concensus in imaging finding. As it has been presented and discussed in the previous sections of this chapter, this has not always been the case. Contrasting findings in the imaging literature on schizophrenia may reflect either differences in the clinical populations or methodological issues. Clinical populations often differ in the number of subjects included in the analysis which significantly influences results and statistical power of the analyses (Salmond et al. 2002). Additionally, clinical cohorts may differ with respect to age, duration of illness and symptom severity scales, the use and type of antipsychotic medication (typical or atypical), substance abuse and the existence of comorbid illnesses (such as depression). These factors have been shown to have modulatory effects on neuroimaging findings (Bora et al. 2011, Fusar-Poli et al. 2011, Haijma et al. 2013, Van Erp et al. 2015) and may partly account for the existing variability.

From a technical point of view, variation in the preprocessing and the image analysis methodology applied (Fusar-Poli et al. 2010), such as the size of the smoothing kernel, the method used for correcting for multiple comparisons or the significance and thresholding scheme, are likely to have influenced imaging findings (Fusar-Poli et al. 2011). Differences in image acquisition parameters and the quality of the acquired images (with respect to the signal to noise ratio) could significantly impact

on segmentation procedures, which is in turn a potential cause of varied brain measures and findings (Abdulkadir et al. 2011, Fusar-Poli et al. 2011). Such variation in image findings leads to an urgent need for replication studies and also meta-analyses that could estimate the effect sizes of those findings across individual studies and account for the confounding effect of clinical- and technical-related factors.

Another important anchor in the identification of biomarkers in schizophrenia is their potential for clinical translation which requires a shift away from considering differences at the group-level and towards making inferences at the individual level. Mass-univariate methods of analysing imaging data are, however, limited to making inferences at the group level. Moreover, region-of-interest (ROI) methods require *a priori* assumptions to be made about regionally specific effects and are thus confined to predefined brain regions. Voxel-based morphometry and other approaches to computational morphometry on the other hand, provide an unbiased, whole-brain approach to studying brain abnormalities but require brain averaging and can neither determine subtle and diffuse networks of neuroanatomical and neurophysiological abnormality across the brain nor capture individual deviations from the norm (Davatzikos 2004). To address these limitations, the neuroimaging community has turned to machine learning methods in an effort to detect the MRI correlates of clinical relevance and utility both because of their ability to examine voxels jointly and their potential for making inferences at a single-subject level.

CHAPTER 3

Background: Machine learning

Work in this chapter has been presented in:

Zarogianni E, Moorhead TWJ, Lawrie SM. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. Neuroimage Clin. 2013;3:279-289.

3.1 Introduction

In this chapter, I present the application of machine learning in the analysis of structural and functional MRI data in diagnosing schizophrenia, particularly for making an early prediction in people at high-risk of developing the disorder. Firstly, a brief overview of machine learning theory is given along with a description of the most common processing steps in the image analysis pipelines. Then, I present and discuss the studies that have employed machine learning in schizophrenia research and finally, I analyze the main practical challenges and limitations that machine learning methods suffer from, in the context of their potential integration into routine clinical practice, before concluding with future research directions.

3.2 Overview of machine Learning

Machine learning (ML) is a term used to describe a set of methods for detecting patterns in data that would enable reliable future predictions. There are two major methodological approaches: supervised and unsupervised machine learning techniques. In supervised learning, the goal is to find a mapping from the data instances x_i to a set of desired outputs y_i, given a set of labeled input-output pairs $D=\{x_i, y_i\}$, for i=1..N instances. Here, D is the training set, consisting of feature vectors x_i and their corresponding labels drawn from label set y_i and N is the number of the training instances. If y_i is a categorical or nominal variable drawn from a finite set, for instance $y_i = \{1, 2 ,....C\}$, then the problem is known as a classification

problem. In its simplest form where C = 2 (and thus y_i = {-1, 1}) this is a binary classification problem, whereas if C > 2, then there is a multi-class classification problem. On the other hand, if y_i is a real-valued (continuous) variable, the problem is known as regression. In unsupervised learning, on the other hand, the goal is to identify an inherent structure in the data in order to classify given data instances D= {x_i} into groups (clustering).

3.2.1 Classification pipeline

The following steps in the image analysis pipeline are common to most machine learning methods:

Preparation of the training set. The first step in an ML analysis is the creation of the training set. This procedure involves two main processes: i) feature extraction and ii) feature selection. Feature extraction involves the transformation of the original data set into a form that would be meaningful for the classifier to process. In the context of neuroimaging, this procedure entails the extraction of feature vectors corresponding to intensity values of voxels from each subject's scan. Feature selection involves a procedure for selecting those feature vectors that are better at discriminating between the classes and thus could facilitate and speed up the classification process. Feature selection can be performed either with a dimensionality reduction technique (such as Principal Component Analysis) or by constraining the research to specific brain areas for which the research team possesses prior knowledge about their likely involvement in the condition under investigation. Feature extraction is an obligatory step in the classification pipeline, but feature selection approaches are optional.

Model training and testing. In the model training step of the pipeline, the chosen algorithm has to learn the relationship between the training set and the labels

associated with it, while trying to optimize the algorithm's parameters in order to maximally discriminate between the groups. In the testing phase, the algorithm tries to predict the class label (in the case of classification) or the continuous variable (in the case of regression) of previously unseen data instances. It is very important that the algorithm generalizes well to new instances. That is, the testing set should not include instances of the training set to avoid circularity or data overfitting. Cross-validation techniques are a popular way to ensure this. In k-fold cross validation, the original data set is split into k non-overlapping sets and then the algorithm is trained using k-1 subsets and the left-out set is used in the testing phase. The procedure is repeated k times, so that every subgroup is used in the testing phase.

Performance evaluation. The final step is the evaluation of classification performance of the method. This usually includes measures such as sensitivity, specificity and accuracy. Sensitivity refers to the proportion of actual positive cases correctly identified (e.g the number of schizophrenia patients identified as in the ill group or class) while specificity refers to the proportion of the negatives cases correctly classified (e.g healthy controls correctly identified as being healthy). Accuracy refers to the overall amount of correct classifications across the groups.

3.3 Machine learning in psychiatry

Machine learning methods have already been applied in the analysis and interpretation of functional and structural MRI data (Pereira et al. 2009, LaConte et al. 2005, Lemm et al. 2011), in 'mind reading' paradigms (Cox & Savoy 2003, Haynes & Rees 2006), the classification of cognitive states (Mitchell et al. 2004, Mourao-Miranda et al. 2005), and in lie detection approaches (Davatzikos et al. 2005a). More recently, classification algorithms have been applied to diagnose neurological and psychiatric disorders (Kloppel et al. 2011, Orru et al. 2012, Bray et al. 2009), such as dementia (Kloppel et al. 2008a, Kloppel et al. 2008b, Davatzikos et

al. 2011), depression (Fu et al. 2008, Mourao-Miranda et al. 2011) and schizophrenia (Fan et al. 2008a, Davatzikos et al. 2005b, Koutsouleris et al. 2009b, Koutsouleris et al. 2012a). Multivariate pattern recognition techniques provide the possibility of making inferences about a subject's health status at an individual level and, thus, are well suited for clinical decision making purposes.

3.3.1 Machine learning in schizophrenia

In the past few years, an increasing number of studies have employed machine learning to investigate the neuroanatomical and neurophysiological correlates of schizophrenia. These studies can be divided into three main categories: (i) studies that examine the diagnostic power of machine learning in distinguishing between healthy controls (HC) and schizophrenia patients (SCHZ), (ii) studies which examine the potential of machine learning to make an early diagnosis of schizophrenia (prediction) by comparing scans at baseline of people at high risk (either for familial or clinical reasons) of making a transition to the disorder and (iii) studies which examine the performance of machine learning in predicting progression of the disease and response to treatment, usually by examining the baseline scans of first-episode (FE) patients with a later known clinical outcome or treatment response.

3.3.1.1 Diagnostic Studies of Schizophrenia

The first study to apply a sMRI-based classification method was conducted by Davatzikos et al. (2005b), who tested the performance of Support Vector Machine (SVM) in classifying 69 schizophrenia patients (46 men, 23 women) and 79 matched healthy controls (41 men, 38 women), reaching a 81% classification accuracy via leave-one-out cross-validation. The authors also tested individual men and women classifiers and observed similar classification results (85% accuracy for the male and 82% for the female classifier), possibly implying good generalisability of the MRI-

based diagnostic system. In another study by the same group, Fan et al. 2007 achieved an impressive 91.8% and 90.8% accuracy in distinguishing between the same 23 female SCHZ patients and 38 female HC and 46 male SCHZ patients and 41 male HC respectively. Here, the development of an adaptive regional feature extraction method, that automatically grouped morphological traits of similar classification power, along with a SVM-Recursive Feature Elimination method, that selected features with the highest discriminatory power, may possibly account for what still remains one of the best diagnostic performances observed in chronic schizophrenia diagnostic studies published to date. The researchers achieved this diagnostic performance by using just 39 features for the female and 44 features for the male individual classifiers. This diagnostic result was, however, obtained from a feature set that might be specific to this sample group and the result may well not generalize to other data samples. In the context of examining family members of schizophrenia, only one study has up to date investigated the role of genetic factors in the disorder, using MRI-based machine learning (Fan et al. 2008a). Fan et al. 2008a observed that unaffected family members share similar phenotypic patterns to their affected schizophrenia relatives. Although these initial results are encouraging, longitudinal studies are, however, essential in determining whether this endophenotypic pattern is present before disease onset and how it relates (if so) to transition to psychosis in unaffected relatives.

Evaluating a classifier on a totally independent cohort is of course the ideal way of examining the generalizability and robustness of the classifier (Nieuwenhuis et al. 2012, Schnack et al. 2014). Unfortunately, the consequent need for large data sets makes this endeavor very difficult. In an impressive two-stage study, Kawasaki et al. 2007 observed a 80% classification accuracy using a partial least squares model that was trained on 30 male HC and 30 male SCHZ patients and tested on a new, independent cohort of 16 male controls and 16 SCHZ patients. In a particularly large classification study employing an independent test set, diagnostic accuracy was however only about 70% (Nieuwenhuis et al. 2012), when testing a SVM classifier developed on 239 participants (128 SCHZ) on a completely independent sample of 277 subjects (155 SCHZ). The use of a larger validation set may partly account for the lower diagnostic accuracy, if one takes into account the possible inclusion of

more variable schizophrenia phenotypes in this larger group.

Table 3.1 Studies employing machine learning and structural MRI to distinguish patients with schizophrenia from healthy controls.

Author	**Sample (N, diagnostic classification)**	**ML methods and scanner field strength**	**Classifier's Performance (accuracy %)**
Davatzikos et al. (2005b)	HC=79, SCHZ=69 DSM-IV	SVM 1.5T	81.1
Fan et al. (2007)	HC_1=38 (females) SCH_1=23 (females) HC_2=41 (males) SCH_2=46 (males) DSM-IV	SVM-RFE 1.5T	HC_1 vs SCH_1= 91.8 HC_2 VS SCH_2=90.8
Kawasaki et al. (2007)	*Train set:* HC=30 SCHZ=30 (males) *Test set*: HC=16 SCHZ=16 (males) DSM-IV	DA & MLM 1.5T	80
Yoon et al. (2007)	HC=52, SCHZ=53 DSM-IV	SVM 1.5T	>90
Sun et al. (2009a)	HC=36, ROS=36 DSM-IV	SMLR 1.5T	86.1
Karageorgiou et al. (2011)	HC=47, ROS=28 SCID-I for DSM-IV	sMRI & Neuropsychological Data PCA-LDA 3T	92
Kasparek et al. (2011)	HC=39, FE=39 ICD-10	MLDA 1.5T	72
Greenstein et al. (2012)	HC=99, COS=98 DSM-IIIR/IV	RF 1.5T	73.7

Nieuwenhuis et al. (2012)	*Train set:* HC=111 SCHZ=128 *Test set*: HC=122 SCHZ=155 DSM-IV	SVM 1.5T	70.4
Zanetti et al. (2013)	HC=62, FE=62 DSM-IV	SVM 1.5T	HC vs FE =73.4
Borgwardt et al. (2012)	HC=22, FE=23 ARMS-T=16 DSM-IIIR	ensemble SVM 1.5T	HC vs FE=86.7 HC vs ARMS-T=80.7 FE vs ARMS-T= 80
Schnack et al. (2014)	*Train set:* HC=66, SCHZ=66 *Test set:* HC=43, SCHZ=46 DSM-IV	SVM 1.5 T: train set 3T: independent test set	HC vs SCHZ i) Cross validation: 90 ii) Independent test set:76

Abbreviations: ARMS-T, at-risk mental state with transition to sch izophrenia; COS, child-onset schizophrenia; DA, discriminant analysis; DSM-IV, Diagnostic and Statistical Manual of Mental Disorder *Fourth Edition*; DSM-IIIR, Diagnostic and Statistical Manual of Mental Disorder *Third Edition Revised ;* FE, first-episode schizophrenia patients; HC, healthy controls; ICD-10, the International Statistical Classification of Disease and Related Health Problems; LDA, linear discriminant analysis; MLDA, Maximum-uncertainty linear discrimination analysis; MLM, multivariate linear model; PCA, principal components analysis; RF, random forests; ROS, recent-onset schizophrenia; SCHZ, schizophrenia patients; SCID-I, Structural Clinical Interview; SMLR, sparse multinomial logistic regression; SVM, Support Vector Machine; SVR, Support Vector Regression; SVM-RFE, Support Vector Machine with Recursive Feature Elimination;

Several studies have, alternatively, employed fMRI in an attempt to establish the diagnosis in groups of people with schizophrenia and controls (Table 3.2). These studies have included various cognitive tasks (Costafreda et al. 2011, Yoon et al. 2012) or resting-state fMRI (Calhoun et al. 2006, Shen et al. 2010, Venkataraman et

al. 2012), in which the subject is simply instructed to remain still during scanning, not to think of anything in particular and not fall asleep. In recent fMRI studies, resting-state paradigms are often preferred to task-related approaches, as they are free from task-related confounds and easier for patient populations to perform, although they do have limitations (Morcom & Fletcher 2007). The diagnostic accuracy of resting-state fMRI-based classification methods ranged from 75% (Jafri & Calhoun 2006, Venkataraman et al. 2012) to 92% (Costafreda et al. 2011, Shen et al. 2010), suggesting that resting-state fMRI has the potential to be useful in clinical practice. Results should, however, be interpreted with caution since the sample sizes in most cases (Shen et al. 2010, Anderson et al. 2010) were very small, potentially introducing a bias to the classification (Demrici & Calhoun 2009).

A recent meta-analysis of 38 multivariate pattern recognition studies, including a total of 1602 schizophrenia patients and 1637 healthy controls and both structural and functional MRI studies, concluded that neuroimaging-based phenotypes can differentiate patients from controls with an overall sensitivity and specificity of 80%, with age and disease stage having significant effects on sensitivity and antipsychotic medication significantly impacting on specificity levels (Kambeitz et al. 2015).

Sample size is, also, an important consideration in neuroimaging-based studies and might also affect classification performance. Although counterintuitive, based on the studies presented here (see Tables 3.1, 3.2 and 3.3), classifiers using small sample sizes seem to have favored diagnostic performance (Fan et al. 2007, Kawasaki et al. 2007, Yoon et al. 2007, Sun et al. 2009b, Anderson et al. 2010, Yang et al. 2010) whereas in studies that emloyed larger participating cohorts, the classification accuracy was lower (Greenstein et al. 2012, Zanetti et al. 2013, Nieuwenhuis et al. 2012), possibly due to the fact that the latter studies have included more variable patient and/or control cohorts, exhibiting a wider range of phenotypic manifestations. However, despite this observation, it should be particularly noted that large sample sizes are needed for building (and testing) reliable and robust models as well as encompassing the range of clinical profiles of schizophrenia patients presented in routine clinical practice.

Differences in the image analysis and classification pipelines might, also, partly

explain such variation in findings. The introduction of refined feature selection methods can boost classifiers' performance, as it was observed in Fan et al. 2007, compared to a previous study of the same group (Davatzikos et al. 2005b). The choice of the machine learning method is another crucial factor in the performance of the diagnostic model as well. Notably, SVM's tend to provide better classification results (Pereira et al. 2009) (Table 3.1) than other pattern recognition methods, although, a direct comparison between the machine learning methods used in the presented studies and classification performance cannot be performed due to other differences in the imaging and clinical characteristics of the samples used.

The clinical characteristics of patients may play a significant role in the observed fluctuations in accuracy across diagnostic studies (Greenstein et al. 2012, Zanetti et al. 2013). Machine learning in FE schizophrenia studies seem to deliver worse diagnostic performance (Kasparek et al. 2011, Zanetti et al. 2013, Yoon et al. 2012) than studies of established schizophrenia (see Tables 3.1 and 3.2), possibly due to the less pronounced brain alterations in the former group (Kambeitz et al. 2015), although diagnostic accuracies can be as high as 92% (see Table 3.1). As was previously mentioned, the first-episode stage of schizophrenia is characterized by less marked brain changes than in chronic schizophrenia, and this could partly account for the accuracy fluctuations observed. In addition, comorbid disorders and patient recruitment procedures may, also, have an effect on the sensitivity of the classifier in detecting disease-specific patterns. For instance, Zanetti et al. (2013) recruited a population-based sample of FE patients with comorbid substance use disorders, using epidemiological methods in order to ensure representativeness of 'real-world' individual cases, and observed just 73.4% accuracy in classifying them against HCs. In childhood-onset schizophrenia (COS), only one study examined the neuroanatomical correlates in 98 COS subjects (all below the age of 13) versus 99 HCs (Greenstein et al. 2012) and observed moderate diagnostic accuracy (73.7%), possibly due to the young age of their patients and the fact that their unconsolidated brain structure may hinder the detection of clear, concrete brain patterns that would facilitate classification. Factors associated with the use of anti-psychotic drug treatment are, also, a serious consideration because medication may have an effect on brain structure (Pantelis et al. 2003, Navari & Dazzan 2009) possibly even up to a

point that the sensitivity of the classifier to detect morphological abnormalities specifically associated with schizophrenia diagnosis is compromised.

A possible way to control for the confounding effect of anti-psychotic medication would be to remove from any analyses those brain regions that are known to be affected by anti-psychotic medication, as seen in the study conducted by Nieuwenhuis et al. 2012 where the authors masked out the stratium and tested the diagnostic accuracy of a SVM-based classifier by excluding (and including) this brain region. However, given the contrasting findings on the effects of anti-psychotic medication on brain structure (Smieskova et al. 2009), and especially gray matter (Shepherd et al. 2012), it would be challenging to decide which brain regions should be left out.

Another analysis-based solution to controlling for any type of counfounding variable would be to stratify the participating cohorts into corresponding sub-groups (for instances users and non-users of anti-psychotic medication, or typical and atypical anti-psychotics users), estimate the effect on the aggregated data and on each stratum and then calculate the pooled estimate across strata (Tripepi et al. 2010). This stratification technique is a very informative strategy as it describes how the effect of the explanatory variable on the outcome of interest varies across subgroups of subjects with different characteristics.

3.3.1.2 Early Diagnostic Studies of Schizophrenia

Several recent neuroimaging studies have shown structural and functional abnormalities in subjects at high-risk of developing schizophrenia compared to healthy controls as well as compared to established patients (Lawrie et al. 2008, Smieskova et al. 2010, Mechelli et al. 2011). To date, there are no biological markers for the identification of emerging psychosis, which is currently identified by clinical symptomatology. The early identification of those high-risk individuals who are most likely to develop psychosis is of high potential clinical value, as early intervention and treatment planning could alleviate symptoms burden or even prevent disease

onset (Marshall & Lockwood 2006, Riecher-Rossler et al. 2006). Job et al. 2005 were the first to assess the predictive value of grey matter reductions in genetic high-risk subjects regarding the possible transition to schizophrenia but as previously discussed, they used univariate analysis methods, with their known limitations. More recently, machine learning has been applied in the context of making an early diagnosis of schizophrenia and even to predict disease transition at individual level (Table 3.3), by identifying the neuroanatomical correlates of vulnerability to psychosis in individuals at high-risk of developing the disorder mainly due to clinical reasons.

Table 3.2 Studies employing machine learning methods and functional MRI in diagnosing schizophrenia.

Author	**Sample (N, diagnostic classification, fMRI paradigm)**	**ML methods and scanner field strength**	**Classifier's Performance (accuracy %)**
Jafri et al. (2006)	HC=31, SCHZ=38 DSM-IV Resting-state paradigm	ICA & NN 3T	76
Calhoun et al. (2008)	HC=26, SCHZ=21 DSM-IV AOD task	ICA 1.5T	*SCHZ vs N-SCHZ*: sensitivity=92 specificity=98 *HC vs N-HC*: sensitivity=95 specificity=88
Shen et al. (2010)	HC=20, SCHZ=32 DSM-IV Resting-state paradigm	Unsupervised classifier based on C-Means 1.5T	92.3
Yang et al. (2010)	HC=20 , SCHZ=20 DSM-IV AOD task	FMRI & genetic data SVM 3T	87
Anderson et al. (2010)	HC=6, SCHZ=14 DSM-IV	ICA & RF 3T	85

	Resting-state paradigm		
Castro et al. (2011)	HC=54 , SCHZ=52 DSM-IV AOD task	ICA & Composite kernels with RFE 3T	95
Costafreda et al. (2011)	HC=40, SCHZ=32 DSM-IV Verbal fluency task	SVM 1.5T	SCHZ vs HC: 92
Fan et al. (2011)	HC=31, SCHZ=31 DSM-IV Resting-state paradigm	ICA & SVM 1.5T	85.5
Venkataraman et al. (2012)	HC=18, SCHZ=18 DSM-IV Resting-state paradigm	RF 3T	75
Yoon et al. (2012)	HC=51, FE=51 DSM-IV Cognitive control task	LDA 1.5 T	61.8

Abbreviations: AOD, auditory oddball discrimination; BD, bipolar disorder; DSM-IV, Diagnostic and Statistical Manual of Mental Disorder *Fourth Edition*; FE, first-episode schizophrenia patients; HC, healthy controls; ICA, independent component analysis; LDA, linear discriminant analysis; NN, neural networks; N-BD, non-bipolar subjects; N-HC, non-healthy controls; N-SCHZ, non- schizophrenia subjects; RF, random forests; SCHZ, schizophrenia patients; SVM, Support Vector Machine;

Koutsouleris et al. (2009b) were the first to apply multivariate pattern recognition to evaluate individual vulnerability to psychosis and predict disease onset. In their work, a SVM classifier was built upon structural MRI data of individuals in early (ARMS-E, n=20) and late at-risk mental state of psychosis (ARMS-L, n=25) and a group of matched healthy controls (HC_1, n=25). The performance of the classifier was validated by distinguishing sMRI data derived from baseline scans of individuals with subsequent transition to schizophrenia (ARMS-T, n=15), those who did not make the transition (ARMS-NT, n=18) and matched healthy controls (HC_2, n=17). Three group and pairwise classifiers were constructed, all achieving classification

performance above 80% (with the exception for the binary classifier HC_1 vs ARMS-L =78%). In the most critical in terms of clinical utility, the ARMS-T vs ARMS-NT pairwise classifier achieved an accuracy of 82%, suggesting the potential of a MRI-based system in predicting transition to schizophrenia. In a follow-up study, Koutsouleris and colleageus (2012a) emphasized the predictive potential of SVMs in classifying an independent cohort of 22 HC, 16 ARMS-T and 21 ARMS-NT subjects. The authors, here, constructed a robust classification method, based on SVM ensemble classifiers that performed feature selection, model learning and predictive ensemble learning wrapped in a nested cross-validation framework. The critical ARMS-T vs ARMS-NT pairwise classifier showed slightly improved classification results compared to their previous work (Koutsouleris et al. 2009b), whereas diagnostic performance was lower in the pairwise HC vs ARMS-NT classifier (66.9% accuracy as opposed to 86% in Koutsouleris et al. 2009b), possibly due to greater heterogeneity in the control sample.

In an effort to identify neuroanatomical markers of transition to psychosis across clinically defined high-risk populations, Koutsouleris et al. (2015b) extended their previous single-site investigations (Koutsouleris et al. 2009b, Koutsouleris et al. 2012a) by pooling two independent cohorts of subjects with ARMS recruited at two different early recognition centres. In this study, the authors constructed an ensemble SVM classifier by using baseline structural MRI data from a pooled data set of 33 ARMS-T and 33 ARMS-NT subjects while an independent group of 7 ARMS-NT subjects was used to further validate the classification. The classifier's performance was evaluated by cross-validation and classification of the independent test set (see Table 3.3) and achieved a balanced accuracy of 80.3% in the pooled data set (sensitivity=75.8%, specificity=84.8%) and 80.4% (sensitivity=75.8%, specificity=85%) in the entire dataset (N=73), suggesting the existence of a neuroanatomical signature across research centres, irrespective of between-site differences. Additional Kaplan-Meier survival analyses, also, enabled the prediction of time to transition, thus facilitating an individualized risk staging that has added benefits for clinical management.

Despite the fact that these studies have delivered very encouraging results in the

context of prediction of disease transition, it should be borne in mind that the at-risk mental state sample in those studies involved symptomatic, help-seeking individuals (Koutsouleris et al. 2012a) and it is therefore unclear if these classification results could generalize to asymptomatic high-risk groups as well.

3.3.1.3 Predicting disease progression and treatment response

Prediction of disease progression is also of interest and potential clinical utility in established cases of schizophrenia, with a view to establishing the prognostic context and/or therapeutic responsivity of psychosis. Based on neuroanatomical pattern classification methods, studies reported poor to modest diagnostic performance (Table 3.3) in predicting the outcome of psychosis in FE schizophrenia patients at baseline. In this context, Mourao-Miranda et al. (2012a) used a linear SVM to predict clinical outcome from baseline sMRI scans of 100 FE psychosis individuals, who at 6-year follow-up were classified as having a continuous, episodic or intermediate course and a group of 91 matched HCs. Although classification accuracy was less than 75% in all contrasts, (see Table 3.3), this result serves as a promising starting point in predicting subsequent course type at the individual level. In another study, Zanetti et al. 2013 failed to predict 1-year outcome of FE schizophrenia patients. Despite the fact that the authors presented a robust method for feature generation and feature selection, their SVM classifier (based on the method proposed in Fan et al. 2007), achieved 58.3% accuracy in predicting clinical outcome of 15 FE patients with a subsequent remitting course versus 21 first-episodes with a subsequent non-remitting course. Differences in data samples (and/or data sample selection procedures) and in the duration of follow-up might partly explain the accuracy discrepancies observed between the two studies.

A key determinant of prognosis in psychosis is diagnosis, both because schizophrenia tends to have a worse outcome than bipolar disorder, and because these conditions tend to respond differently to treatments. Early studies have shown the possibility of distinguishing group activation patterns on fMRI in schizophrenia and bipolar disorder (McIntosh et al. 2008), but little SVM work has thus far been

done in this vein, especially at first presentation when it might be most valuable.

In the context of predicting response to treatment in schizophrenia, only one study that I am aware of has thus far employed machine learning to do so. Khodayari-Rostamabad et al. (2010) used kernel partial least squares regression in order to predict response to clozapine in chronic schizophrenia subjects, based on pre-treatment electroencephalography (EEG) data, providing 85% accuracy in identifying responders and non-responders to the medicine.

Table 3.3 Studies using machine learning to predict transition, progression and treatment response in schizophrenia.

Author	**Sample(N, diagnostic classification)**	**ML methods and scanner field strength**	**Classifier's Performance (accuracy %)**
Koutsouleris et al. (2009)	HC_1=25, HC_2=17 ARMS-E=20, ARMS-L=25, ARMS-T=15, ARMS-NT=18 at inclusion:DSM-IV at follow-up: ICD-10	Structural MRI SVM 1.5T	HC_1 vs ARMS-E vs ARMS-L = 81 HC_2 vs ARMS-T vs ARMS-NT = 82
Khodayari-Romastabad et al. (2010)	*Train set:* SCHZ=23 R=12, NR=11 *Test set:* SCHZ=14 at inclusion: DSM-IV post-treatment evaluation: PANSS	EEG kernel PLSR	R vs NR= 85
Koutsouleris et al. (2010)	HC=28,ARMS=25 ARMS-T=12, ARMS-NT=13 at inclusion: DSM-IV at follow-up: ICD	Structural MRI SVR 1.5T	HC vs ARMS: r = 0.83 HC vs ARMS-T vs ARMS-NT: r= 0.83
Koutsouleris et	HC=22, ARMS-T=16,	Structural MRI	HC vs ARMS-T= 92.3

al. (2012a)	ARMS-NT=21 at inclusion: APS, BLIPS at follow-up: classification criteria by Yung et al. 1998	ensemble SVM 1.5T	HC vs ARMS-NT= 66.9 ARMS-T vs ARMS-NT = 84.2
Mourao-Miranda et al. (2012a)	HC=28, EP-PS=28 CON-PS=28, INT-PS=32 at inclusion: ICD-10 at follow-up: WHO Life Chart	Structural MRI SVM 1.5T	EP-PS vs CON-PS= 70 CON-PS vs HC=67 EP-PS vs HC= 54
Zanetti et al. (2013)	R-FE=15, NRsub-FE=21 at inclusion: DSM-IV (SCID) at follow-up: DSM-IV	Structural MRI SVM 1.5T	R-FE vs NRsub-FE=58.3 HCsub vs NR-FE=64.3
Koutsouleris et al. (2015b)	*Train set:* ARMS-T=33, ARMS-NT=33 *Test set:* ARMS-NT=7	Structural MRI ensemble SVM 1.5T	ARMS-T vs ARMS=NT i) Cross-validation: 80 ii) Independent test set: Specificity: 85 iii) Overall BAC: 80

Abbreviations: ARMS, at-risk mental state; ARMS-E, at-risk mental state early; ARMS-L, at-risk mental state late; ARMS-T, at-risk mental state with Transition to schizophrenia; ARMS-NT, at-risk mental state without transition to schizophrenia; APS, Attenuated Psychotic Symptoms; BAC, Balanced accuracy; BLIPS, brief limited intermittent psychotic symptoms; CON-PS, continuous psychotic; DSM-IV, Diagnostic and Statistical Manual of Mental Disorder *Fourth Edition*; EP-PS, episodic psychotic; HC, healthy controls; ICD-10, the International Statistical Classification of Disease and Related Health Problems; INT-PS, intermediate psychotic; NR, non-responders; NRsub-FE, subgroup of non-remittent first-episodes; partial least squares regression; PANSS, positive and negative syndrome scale; PSLR, partial least squares regression; R, responders; R-FE, remittent fist-episodes; SCHZ, schizophrenia patients; SCID, Structured Clinical Interview; SVM, Support Vector Machine; SVR, Support Vector Regression; WHO, world health organization.

3.3.2 Discussion

Studies published so far demonstrate promising leads for the development of neuroimaging machine learning-based tools that could assist in establishing the diagnosis and prognosis of schizophrenia and therefore be useful in clinical practice. Machine learning methods are advantageous compared to standard univariate statistical methods, in that they have the potential to make inferences about effects of interest at a single-subject level and can detect subtle and widespread neuroanatomical and functional differences that span over large networks of brain regions, by virtue of their multivariate nature.

The development of a MRI-based machine learning system could well aid in the identification of objective biological markers for schizophrenia, and could thus help overcome the subjectivity in traditional clinical assessments. There are, however, significant hurdles to be overcome before their integration of machine learning into clinical practice is possible. The classifiers' performance is a key element for the potential integration of machine learning into clinical decision-making. As a general observation, diagnostic classification performance in psychiatry may not supersede clinical expertise in the foreseeable future, no matter the techniques employed, since training a classifier requires prior knowledge of a subject's clinical status (Orru et al. 2012). Where imaging and machine learning could still provide added clinical value is where early diagnosis, prognosis of long-term outcome and treatment response are difficult to predict. For example, the identification of high-risk individuals, likely to convert to schizophrenia is of high clinical value as a means to inform early treatment strategies that could result in better outcomes for the patients. It is, however, evident from the early diagnosis studies thus far (see Tables 3.1-3.3) that classification accuracy in the early detection of schizophrenia and predicting clinical course is not as high as in diagnostic schemes. This is probably explained by the fact that in the diagnosis of established groups of patients from controls, neuroanatomical and functional patterns of differentiation are more clearly and strongly established than in same group subjects who do or do not go on to show an outcome of interest and therefore present a more difficult classification problem.

It should be borne in mind that a classifier with high sensitivity and high specificity is desirable, and that overall accuracy is important, but the relative value of high and low sensitivity and specificity could have different implications in patients' clinical management, in different clinical scenarios, depending on the availability of treatment and the seriousness and frequency of adverse effects. Moreover, for an individualized patient high positive/negative predictive power is the most critical consideration (Lawrie et al. 2011). Furthermore, classification performance is primarily affected by the sample size. The limited number and nature of patient populations in SVM neuroimaging-based studies means that these encouraging early results may not generalize well to other patient groups. Recruiting patients for research studies can be difficult and patients with co-morbid conditions are often excluded, resulting in a limited representation of the various phenotypes across the spectrum of schizophrenia. Despite the fact that several machine learning methods can deal effectively with small sample size (Pereira et al. 2009), a limited number of data samples can cause model overfitting, resulting in poor generalization of the method to independent data sets. In such cases, cross-validation frameworks are often employed, to partition the original data set. However, cross-validation schemes should be performed with caution, as in many cases data samples in the validation (or training) set are also present in the testing set, seriously biasing the classifier's performance as a result. As a general rule, the greater the complexity of a method, the higher is the risk for overfitting the data (Mourao-Miranda et al. 2012a). Ideally, data for validation and training should be derived from completely independent cohorts, thus eliminating the need for performing cross-validation and the ensuing danger of poorly conducting one, and ensuring the robustness and reliability of the classification methodology (Kawasaki et al. 2007, van Haren et al. 2012).

The need for large data sets could be addressed with pooling data from multiple research centres (Mechelli et al. 2011). The existence of a well-validated training dataset to be shared between neuroimaging centres is likely to be of importance for standardizing classification accuracy across laboratories. In addition, future multi-site studies could provide the possibility for encompassing more heterogeneous clinical populations, demonstrating a range of clinical manifestations of a disorder (Borgwardt et al. 2012), for example subjects with various transition rates to

psychosis or subjects of lower diagnostic certainty, which could thus provide a more realistic mirroring of everyday psychiatric practice. Multi-site projects for the development of biomarkers for the Alzheimer's disease already exist (Alzheimer's Disease Neuroimage Initiative - Mueller et al. 2005) while only recently, multi-centre projects have launched for the early diagnosis and management of early psychosis; the PSYSCAN (http://psyscan.eu) and PRONIA projects (http://www.pronia.eu). Data sharing among research centres faces, however, its own difficulties. Different scanners, imaging parameters and protocols result in varying image intensity and susceptibility profiles that will require careful consideration and compatibility solutions. One promising approach is however to generate metrics from individual scans that can then be compared to reference data sets (Tijms et al. 2012).

Equally important, future studies should test the efficacy of machine learning in making a diagnosis of psychiatric disorders apart from schizophrenia, such as bipolar disorder, borderline personality disorder, depression, autism etc. Initial studies have already used machine learning to differentiate schizophrenia from bipolar disorder (and HC subjects) both by employing structural (Schnack et al. 2014) or functional MRI data (Costafreda et al. 2011, Cahloun et al. 2008) and delivering very encouraging diagnostic results. In the same context, another recent study has employed a neuroanatomical-based pattern classification technique to make a differential diagnosis of schizophrenia and mood disorders, namely major depression and bipolar disorder (Koutsouleris et al. 2015a). In order to explore the hypothesis that major depression, bipolar disorder and schizophrenia might represent different stages along the same neurobiological continuum (Lin et al. 2013), the authors first examined whether schizophrenia patients (n=158) could be distinguished from major depression patients (n=104) at the single-subject level and then quantified differential diagnostic decision values in order to characterize independent cohorts of clinically-defined HR and FE individuals (n=112 in total) and patients with bipolar disorder (n=35). A cross-validation scheme delivered an 76% balanced accuracy in the schizophrenia versus major depression classification task, whereas the trained model assigned 74% of bipolar patients to the major depression group, possibly suggesting that major depression and bipolar disorder share similar neuroanatomical signatures that are differentiated from schizophrenia (Koutsouleris et al. 2015a). However,

replication of these early findings in studies that include larger samples and more cases across a putative psychosis spectrum is necessary in order to identify patterns that differentiate between these psychiatric disorders.

From a methodological point of view, novel methods for feature selection and decision making of the classifiers could be introduced in order to improve diagnostic power in schizophrenia studies. For example, ensemble-learning methods could be introduced in order to improve the generalization ability of a classifier. Ensemble classifiers can achieve better predictive performance than single classifiers, by combining multiple weak learning models that decide upon the classification of a new instance through majority voting (Polikar 2006). Some well-known ensemble learning methods, such as bagging and random subspace methods have already been used in neuroimaging settings to identify biological markers for prodromal Alzheimer's disease (Fan et al. 2008b, Liu et al. 2012), reporting excellent diagnostic results. Ensemble learning could be a useful approach in data fusion studies as well, where a single classifier could be built and trained for each imaging modality and/or clinical measures (such as neurocognitive measures) separately and outputs from each classifier could be combined to classify new instances. An example of this approach is the study of Yang et al. (2010), who developed SVM-based ensemble classifiers of genetic and fMRI data and combined them to a single module that decided upon classification of testing samples via majority voting, achieving better diagnostic accuracy than either SVM ensembles alone (87% for the combined module, 74% for the genetic data classifier and 83% for the fMRI classifier). Future studies could, also, possibly address the problem of 'tuning' a machine learning method to fit into neuroimaging settings. Refinements in the SVM method, for example, already exist. The SVM-Recursive Feature Elimination (SVM-RFE), a very popular method that performs feature selection during training and recursively removes data instances, and has already been successfully employed in cancer classification (Guyon et al. 2002), and SVM-Sequential Minimal Optimization (SVM-SMO) which facilitates and speeds up the classifier's training, are methods yet to be validated for their efficacy in neuroimaging settings.

3.4 Conclusions

The application of machine learning methods for the purposes of diagnosing or making a prediction of psychosis onset has already demonstrated very encouraging results. The main advantage of machine learning methods, over standard univariate ways of analysing and interpreting neuroimaging data, is that they allow inferences to be made at subject-level; a key-feature in clinical practice. There are however, important difficulties yet to be fully considered and overcome, before their translation into routine clinical practice. The optimal means of multi-centre analyses, fusing imaging modalities and integrating various sources of information are critical considerations. Finally, once suitable techniques have been developed, they will ideally need to be tested, preferably in randomized control trials to ensure that they are acceptable and useful to clinicians and patients.

CHAPTER 4

Materials and Methods

4.1 Introduction

In the previous chapters, the concepts of neuroimaging and machine learning were presented along with a critical review of the application of machine learning in schizophrenia. As has been described in Chapter 1, the overriding aim of the present thesis is to apply machine learning in order to identify predictors of transition to schizophrenia in subjects at high risk for developing the illness. Initially, data from the Edinburgh High Risk Study (EHRS), which were immediately available, were used in order to examine the capability of SVM in identifying neuroanatomical markers that predict schizophrenia onset, and then data from the FePsy (Fruherkennung von Psychosen) study, which comprised clinical high-risk subjets, were put to the test with the intention of replicating earlier findings and examining the generalizability of the method in distinguishing clinical high-risk cohorts.

In this chapter, the subject material of the EHRS and the FePsy studies is presented and the pre-processing and classification methodology is extensively described. To help the reader in the understanding and comparing the EHRS and FePsy study, Tables 4.1 and 4.2 provide a summary of the main cohort information and the main imaging findings in the two studies.

4.2 Subjects

4.2.1 Edinburgh High Risk Study

The Edinburgh High Risk Study (EHRS) was a prospective, longitudinal study in which young people from multiply affected families with schizophrenia and matched groups of controls without family history were recruited. Comprehensive details of the recruitment process can be found in previous papers (Hodges et al. 1999, Johnstone et al. 2000). The original idea for this study was conceived by Professor Johnstone. The purpose of the EHRS was to determine the features that distinguish

high risk individuals who go on to develop schizophrenia from those who do not, and to compare relevant variables in affected and unaffected members of the high-risk sample with matched controls.

The study took place between 1994 and 2004. High-risk subjects were selected on the basis of having two or more first or second-degree relatives with a confirmed diagnosis of schizophrenia using the OPCRIT (Operational Criteria Checklist) computer program (McGuffin et al. 1991). Potential high-risk participants were identified by examining case-notes of patients with schizophrenia. Upon completion of recruitment, a total of 229 high-risk individuals were identified throughout Scotland, from which 156 participants provided complete data at baseline. Ten to fifteen percent of the high-risk group was predicted to develop schizophrenia by the age of 30 on the basis of the known frequency of the disease in individuals with this degree of heredity, and the actual occurrence of schizophrenia by this age (Johnstone et al. 2002a). Based on this, comparable groups of control subjects were determined and recruited: 36 healthy control subjects and 37 individuals in their first episode of the illness, both with no known family history of schizophrenia. All study participants were given written information about the study and time to consider before formally consenting taking part in the study. They were, also, aware that they could withdraw at any time. The study was approved by the Lothian Research Ethics Committee for Psychiatry and Pathology.

The high risk participants recruited had an age range between 16-24 years at baseline in order to ensure that the period of maximum risk of onset of schizophrenia would be covered. High-risk participants were followed up for up to 10 years during which they underwent a series of clinical, neuropsychological and neuroimaging assessments every 18-24 months.

Table 4.1 Outline of the EHRS and the FePsy study.

	EHRS	**FePsy**
Cohort	Familial High Risk	Clinical High Risk
Inclusion criteria	2 or more 1rst or 2nd degree relatives with schizophrenia	1. Basel Screening Instrument for psychosis (BSIP) 2. PACE criteria: - attenuated psychotic symptoms (APS) - brief limited intermittent psychotic symptoms (BLIPS) - genetic risk plus 2 risk factors
Age at study inclusion	16-25 years	18-45 years
Follow-up	10 year; assessments every 18-24 months	Up to 4-years; Monthly during 1rst year, months interval (2nd and 3rd year), annually thereafter
Psychopathology	Present State Examination (PSE); five-score psychopathological scale	Brief Psychiatric rating Scale (BPRS); Scale for the Assessment of Negative Symptoms (SANS)
Transition Criteria	PSE, score 4. Further validated using the ICD-10.	As defined in PACE criteria; further validated by ICD-10.
Converters and non-converters numbers	- 17 converters (with complete clinical assessments and at least 1 MRI scan) *Non-converters:* -57 HR[symp] -57 HR[well]	- 16 converters - 19 non-converters - 2 HR subjects with no follow-up information
Anti-psychotics (baseline)	All anti-psychotic naïve.	30 subjects, some time prior to study inclusion.

HR[symp]: High risk individuals that did not develop schizophrenia but yet exhibit psychotic or partially-held psychotic symptoms; HR[well]: High risk subjects that did not develop schizophrenia and did not have any psychotic symptoms.

4.2.1.1 Psychopathology

Psychopathology was assessed at entry and follow-up by the Present State Examination (PSE, Wing et al. 1974), which involved a structured interview with a psychiatrist in order to identify any psychotic symptoms. The PSE is a very detailed instrument giving a standardised assessment of a wide range of symptomatology and therefore would be helpful in evaluating the extent of any psychopathology shown by the high-risk participants and controls. Based on the PSE profiles, a five-score psychopathological scale system (Johnstone et al. 2000, Johnstone et al. 2002a) was administered by three experienced clinicians (Professor Johnstone, Professor Owens and Professor Lawrie). A score of 4 was assigned for definite schizophrenia based on the PSE, and was further validated by the ICD-10 (World Health Organization, 1993). A score of 3 was assigned for any fully-rated psychotic symptom; PSE items 55-92 (including thought reading, echo broadcast auditory, visual or other hallucinations, delusions of control, misinterpretation, reference, persecution, grandiosity influence or other) or PSE behavioural items 128, 129 and 135-137 (including blunted affect, incongruous affect, neologisms or idiosyncratic use of words, incoherence of speech, flight of ideas). A score of 2 was assigned if any of the features in 3 were partially held or present to a mild degree plus symptoms 49-54 (perceptual disorders other than hallucinations), and behavioural items 108, 109, 118, 125, 126 (self-neglect, bizarre appearance, behaves as if hallucinated, suspicious, perplexed) fully rated and items 133 (muteness) partially or fully rated. A score of 1 was assigned in cases where none of the above features existed but any other fully rated PSE item (such as tension, depression, anxiety, instability, obsessions etc.).

Finally a score of 0 was assigned if none of the above symptoms were observed. In essence:

4= schizophrenia

3= fully rated psychotic symptoms

2=partially rated psychotic symptoms

1= fully or partially rated non-psychotic symptoms

0= no symptoms

Subjects were further categorized according to the presence or absence of psychotic symptoms as individuals with fully or partially held psychotic symptoms (HR[symp], scores 2 and 3 combined), individuals without psychotic symptoms (HR[well], scores 0 and 1 combined) and individuals with schizophrenia (HR[ill], score 4).

All subjects were antipsychotic-naive at study-entry and at follow-up or until they were clinically diagnosed with schizophrenia. From those HR subjects who provided complete clinical assessments and had a MRI scan, 17 were diagnosed at follow-up with schizophrenia (after an average of 929 days, SD=138) based on the ICD-10. Once diagnosis of schizophrenia has been established, these subjects did not undergo any further assessments. However, clinical management of those patients ensured that their diagnoses remained unchanged until the end of the study and no other psychotic diagnoses were recorded.

Among the rest, 57 subjects experienced psychotic or possibly psychotic symptoms but were never ill enough to be given formal diagnosis of schizophrenia, up until the end of the 10-year follow-up period. However, these symptoms were too transient or mild to satisfy operational definition for schizophrenia or any related psychotic illness. The rest of the HR subjects remained well, with no symptoms. A summary of the main cohort characteristics and study criteria can be found in Table 4.1.

4.2.1.2 Main findings in the EHRS

In terms of demographic details, there were no significant differences between high-risk individuals and healthy controls with respect to age, sex, parental social class, alcohol and cannabis use at baseline (Johnstone et al. 2000). However, the high-risk group showed poorer educational and employment attainment than the control group (Johnstone et al. 2000).

Early studies in the ERHS have reported differences in clinical, psychopathological, neuropsychological and neurological indices between the high-risk group and the control groups (Johnstone et al. 2000, Cosway et al. 2000, Johnstone et al. 2002a, Owens et al. 2006 Lawrie et al. 2001a, Miller et al. 2002a, Miller et al. 2002b, Byrne et al. 2003, Byrne et al. 1999). Additionally the performance of high-risk participants on some neuropsychological tests remained poor over time but did not deteriorate relative to controls, which may provide evidence of a stable trait deficit and a possible cognitive marker for schizophrenia. In terms of psychopathology, high risk individuals showed more symptomatology than controls, mostly including partial and definite psychotic and non-psychotic symptoms (Johnstone et al. 2002a, Johnstone et al. 2002b). These findings might indicate that what is inherited by high risk individuals due to genetic liability is not the disorder itself, but a state of vulnerability manifested by fairly widespread neuropsychological impairments.

Significant predictors of schizophrenia have, also, been identified from a battery of behavioural and neuropsychological tests administered to high-risk individuals at baseline (Johnstone et al. 2005). Baseline memory and learning, as tested using the Rey Auditory Verbal Learning Test (RAVLT, Rey 1964) was significantly worse in high risk than control subjects (Byrne et al. 1999, Byrne et al. 2003) and particularly poor in those who developed the disorder (Whyte et al. 2006). Notably, over-time performance in the RAVLT was significantly improved across all HR and HC individuals and those HR that developed schizophrenia but also remained significantly different between the groups, possibly reflecting a familiarity with the test procedures at repeated assessments (Whyte et al. 2006). Previous studies have shown that the Rust Inventory of Schizotypal Cognitions (RISC, Rust 1998) and the

Structured Interview for Schizotypy (SIS, Kendler et al. 1989) were able to discriminate between schizophrenia patients and healthy controls. In the EHRS, baseline assessments of the SIS (total score, Miller et al. 2002a), RISC (Miller et al. 2002b) and the RAVLT (Byrne et al. 2003) were able to distinguish at baseline those high-risk subjects that will go on to develop schizophrenia from the high-risk subjects who do and do not develop psychotic symptoms (Johnstone et al. 2005). Specifically, those individuals that developed schizophrenia differed those who did not on social anxiety and withdrawal, verbal learning and memory and other schizotypal features (Johnstone et al. 2005). In relation to studies that included high-risk populations on the basis of symptomatic criteria or a combination of familial risk and symptomatic criteria, the EHRS cohort comprised non-help seeking individuals that presented much lower transition rates to schizophrenia. Most of these high-risk subjects were asymptomatic and functioned at a similar level to controls in terms of employment or further education. It can, also, be concluded that transient or partially-held psychotic symptoms in this familial cohort occurred in many more individuals that might be anticipated to develop schizophrenia. In contrast, in clinical high-risk studies the presence of sub-threshold psychotic symptoms was associated with later development of psychosis.

A significant number of studies have also reported a plethora of neuroimaging abnormalities within the groups of the EHRS (Lawrie et al. 2001b, Lawrie et al. 2008, Owens et al. 2006). Significant reductions in the amygdalo-hippocampal complex and the volume of the thalamus were reported in high-risk subjects against controls (Lawrie et al. 1999). A previous VBM study showed significant grey matter reductions in the anterior cingulate and medial prefrontal lobes in high-risk versus healthy controls with more marked changes shown in first episode patients (Job et al. 2002, Job et al. 2003). Job et al. 2005 found reductions in the concentration of grey matter in the temporal and prefrontal lobes in the high-risk group compared to the healthy controls but these were more pronounced in those subjects with a liability to develop psychotic symptoms. However these findings do not establish whether brain structure changes as a result of psychotic symptoms, or whether symptoms and changes in brain structure occur in unison. Additional grey matter density reductions

were presented in the left para-hippocampal uncus, fusiform gyrus and right cerebellar cortex in those that made a transition to schizophrenia against those who did not but had psychotic symptoms (Job et al. 2005). Progressive reductions over two successive MRI time-points (18-months apart) in the three previously specified regions showed significant diagnostic properties in predicting transition to schizophrenia, reporting positive predictive values as high as 60% and thus possibly suggesting their use as part of a positive predictive test for schizophrenia (Job et al. 2006). These findings could be clinically relevant as part of a predictive test for schizophrenia in people at enhanced risk for familial reasons, for positive predictive power and in combination with other neurocognitive and behavioural predictive measures that were shown to have strong negative predictive power (Johnstone et al. 2005). It should be noted, however, that while this might be the case, the values of sensitivity and specificity are not high enough for diagnosis of schizophrenia.

Finally, in the context of associating brain structure with schizotypal measures, as assessed by the RISC and SIS, and verbal and learning memory RAVLT assessment, significant associations between grey matter density in the left superior temporal gyrus and the RISC measure was found for those high-risk subjects that made a transition to schizophrenia while the RAVLT was significantly correlated with grey matter density reductions in the right parahippocampal gyrus in high risk subjects that exhibited transient, isolated or partial symptoms (Lymer et al. 2006).

Overall, the findings from the EHRS literature presented above have many strengths including the unprecendently large number of familial high risk individuals, all of whom were unmedicated at study-entry and at follow-up or until they met operational criteria for schizophrenia. However, as seen in all studies that involved high-risk subjects that later developed schizophrenia, the sample size was small and therefore findings require replication. Moreover, studies presented in this section only include findings from VBM analysis and are thus restricted to limitations inherent to this technique. It is widely recognised that VBM is sensitive to systematic shape differences, so the choice of the parameters in spatial normalization and the template used for that purpose can have an impact on the resulting brain measures (Job et al. 2003). The choice of kernel in the smoothing step can also obscure differences of lesser spatial extent and affect results. Finally, the use of small volume

corrections (SVC) and the size of resels along with the inclusion and exclusion of specific confounding variables such as age, sex and total intracranial brain volume in the modelling stage of VBM could partly explain the heterogeneity with previous EHRS studies and other neuroimaging studies in the field.

4.2.2 The FePsy study

The FePsy (Fruherkennung von Psychosen) study is a prospective, longitudinal study that aims to identify and investigate individuals considered to be at high risk of psychosis and matched groups of individuals at their first psychotic episode. Subjects were recruited through a specialized clinic for the early detection of psychosis at the Psychiatric Outpatient Department, University Hospital in Basel, Switzerland. The FePsy study started in March 2000 and inclusion of high risk participants was completed by February 2004.

The study was approved by the local ethics committee of the University of Basel and written informed consent was obtained for each participant.

4.2.2.1 Screening and psychopathology

For screening purposes, subjects were assessed using the Basel Screening Instrument for Psychosis (BSIP, Riecher-Rossler et al. 2008). The BSIP is a 46-item checklist based on variables which have been reported as risk factors or predictors of psychosis (Riecher et al. 1990, Riecher-Rossler et al. 2006, Riecher-Rossler et al. 2007), such as DSM-III-R-'prodromal' symptoms, social decline, drug abuse, previous psychiatric disorders or genetic liability for psychosis. The BSIP checklist was used in combination with the Brief Psychiatric Rating Scale (BPRS expanded version; Lykoff et al. 1986, Ventura et al. 1993) in order to evaluate the severity of pre-psychotic signs. All assessments were conducted by experienced psychiatrists

who underwent regular training.

Subjects were identified as high-risk or an at-risk mental state (ARMS) if they met (one or more of) the following inclusion criteria that corresponded to the Personal Assessment and Crisis Evaluation (PACE) criteria (Yung et al. 1998), which have been already employed in other studies using individuals with an at-risk mental state (Phillips et al. 2002, Pantelis et al. 2003):

i) Attenuated psychotic-like symptoms

ii) Brief limited intermittent psychotic symptoms (BLIPS)

iii) genetic risk: a first or second-degree relative with a psychotic disorder plus at least 2 further risk factors for or indicators of beginning psychosis according to the BSIP screening instrument.

Inclusion because of attenuated psychotic symptoms required that change in mental state had to be present at least several times a week and for more than 1 week's duration (a score of 2 or 3 on the BPRS hallucination item or 3 or 4 on BPRS items for unusual thought content or suspiciousness).

Inclusion because of BLIPS required scores of 4 or above on the hallucination item or 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items of the BPRS, with each symptom lasting less than 1 week before resolving spontaneously. A more detailed description of these ARMS criteria can be found in Riecher-Rossler et al. 2007.

Additionally, negative symptoms were assessed using the Scale for the Assessment of Negative Symptoms (SANS; Andreasen 1989), which was used in combination with the BSIP. The SANS assessment is a well-recognised rating scale for the assessment of negative symptoms in schizophrenia and consists of 19 items, which are grouped into five domains or factors (affective flattening, alogia, avolition-apathy, anhedonia-asociality, and inattention). Further assessments to elicit neuropsychological and psychopathological indices were performed. However, their description is outside the scope of the present thesis, and additional information can

be sought in previous papers of the FePsy study (Gschwandtner et al. 2003, Riecher-Rossler et al. 2009).

Exclusion criteria were age below 18 years, insufficient knowledge of German, IQ <70, previous episodes of schizophrenic psychosis (treated with major tranquillizers for more than 3 weeks), a clearly diagnosed brain disease or substance dependency (except for cannabis dependency), or psychotic symptoms within a clearly diagnosed depression or borderline personality disorder.

In total, 37 ARMS individuals were recruited. Thirty of the 37 ARMS individuals never received antipsychotic medication. Seven participants had been administered low doses of antipsychotic medication for behavioural control by the referring psychiatrist or general practitioner (2 participants on olanzapine, 2 Chlorprothixene and 3 risperidone) at some time prior to study inclusion.

Matched groups of healthy controls and first-episode patients were recruited as well. In short, 22 healthy controls (HC) with no personal or family history of any psychiatric disorder were recruited from the same geographical area as the ARMS group through local advertisements and were matched to the ARMS sample group-wise for age, gender, handedness, and education level. The first-episode group (FE) consisted of 25 individuals who met operational criteria for first episode psychosis as described in Yung et al. 1998. More details regarding inclusion and operational criteria for the HC and FE group can be found in Riecher-Rossler et al. 2007.

Transition to psychosis was monitored by means of the transition criteria defined in Yung et al. (1998) and based on the BPRS scale; that is BPRS scores of 4 or above on the hallucination item or scores of 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items; symptoms had to occur daily and persist for more than 1 week to be deemed a conversion to frank psychosis. In subjects who met these criteria the diagnosis was determined by an interview using ICD-10 research criteria at the time of transition, corroborated by a subsequent assessment at least one year post transition using the Operational Criteria for schizophrenia/schizoaffective disorder (OPCRIT) checklist for psychotic and affective illness (McGuffin et al. 1991). Based on these, the ARMS group was

subsequently divided into 16 ARMS subjects that made a transition to psychosis (ARMS-T) and 19 ARMS individuals that did not convert (ARMS-NT) (no follow-up information was available for two ARMS subjects). See Table 4.1 for an outline of the FePSY study compared to the EHRS.

Table 4.2 The main structural imaging findings in the EHRS and FePsy literature.

	EHRS	**FePsy**
HR vs. HC	↓ anterior cingulate (R+L) medial frontal lobe (R) middle temporal gyrus (L) postcentral gyrus (L) limbic lobe (L) parahippocampal gyrus(R+L) thalamus (R+L)	↓ insula (L) superior temporal gyrus (R+L) hippocampus (R) amygdala (R) posterior cingulate gyrus precuneus
HR-T vs. HR-NT	↓ cerebellum (R) inferior temporal gyrus (L) (para)hippocampal uncus (L)	↓ insula (R) inferior frontal gyrus superior temporal gyrus (R) anterior cingulate (R+L) ↑ parahippocampal gyrus (R+L) fusiform gyrus (R+L) medial occipital gyrus (R+L) thalamus(R+L) supramarginal gyrus (R)

R: right; L: left.

4.2.2.2 Main Findings in the FePsy study

In the FePsy study, individuals with an ARMS did not differ significantly from healthy controls with respect to age, sex, ethnicity, educational level and total brain volume. The ARMS subjects that later developed psychosis showed at baseline more pre-psychotic symptoms than ARMS subjects who did not, but less severe symptoms than the FE group, namely in all four BPRS subgroups (see Appendix IV, Figure 1)

but also for the BPRS global score measuring general psychopathology (Riecher-Rossler et al. 2007). In the negative symptomatology scale as measured by the SANS, the anhedonia/ asociality scale was significantly different in those ARMS that developed psychosis versus those who did not (Riecher-Rossler et al. 2007). It was, also, shown that a stronger weighting of clinical and cognitive variables in an intergrated model could predict transition to psychosis in the ARMS population (Riecher-Rossler et al. 2009). Specifically, suspiciousness (attenutated psychotic symptom/ BPRS subscale; see Appendix IV, Table 1), anhedonia/asociality (negative symptoms, SANS subscale) and reduced speed in information processing task (Test for Attentional Performance; TAP Go/NoGo false alarms; Zimmermann and Fimm 1993) could achieve an overall predictive accuracy of 80.9%, with a sensitivity of 83.3% and a specificity of 79.3% (Riecher-Rossler et al. 2009).

At as structural level, ARMS individuals showed significant volumetric deficits in a cluster of regions that included the left insula, superior temporal gyrus, cingulate gyrus and precuneus against control groups of healthy individuals and first-episode patients (Borgwardt et al. 2007b). Within the ARMS group, those individuals that later made transition to psychosis demonstrated grey matter volume deficits in the right insula, inferior frontal and superior temporal gyrus compared to the AMRS individuals that did not develop psychosis (Borgwardt et al. 2007b), suggesting that transition to psychosis is associated with alterations in regional grey matter volume, particularly in the inferior frontal and medial temporal cortex as seen in the EHRS (Job et al. 2005). In addition, ARMS-T individuals showed relatively greater grey matter volume in the parahippocampal gyrus, the parietal and posterior temporal cortex, and the thalamus contrary to their initial hypotheses. The finding of larger parahippocampal gyri is of particular interest as in the EHRS literature transition to psychosis was associated with a progressive reduction in medial temporal volume in high risk individuals due to familial reasons (Job et al. 2005).

Another study compared only the ARMS-T group against healthy controls and found grey matter volume reductions in the posterior cingulate gyrus, the precuneus, paracentral lobule bilaterally and in the left superior parietal lobule while compared to first-episodes, ARMS-T individuals showed greater volume in the temporal gyrus bilaterally, possibly suggesting that temporal lobe abnormalities seem to occur later

in the course of illness (Borgwardt et al. 2007a).

Table 4.2 provides a summary of the main MRI findings in the EHRS and FePsy study.

4.3 Preprocessing of the MRI scans

Preprocessing of the MRI scans was based on the VBM methodology, as described in Ashburner & Friston (2000) and Good et al. 2001, and was performed using the SPM5 toolbox (*http://www.fil.ion.ucl.ac.uk/spm/software/spm5*), running in Matlab version 7 *(The MathWorks, Natick, MA)* on a linux machine operating GNOME version 2.2.8.2. The same preprocessing pipeline was followed for both the EHRS and the FePsy dataset. Study-specific templates and study-specific brain tissue a-priori maps were constructed for each dataset separately, as described below.

4.3.1 EHRS study-specific templates and priors

A study-specific template was constructed from all 146 HR and 36 control structurally normal MRI scans. Since this group contained scans from subjects of mixed outcome (i.e., those without psychotic symptoms (HR[well]), those with isolated or partial psychotic symptoms (HR[symp]) and those who were later diagnosed with schizophrenia (HR[ill]), it was believed to represent the entire study population and therefore minimised bias for spatial normalisation. The scans were normalised to the generic SPM T1 template using 12-point linear affine transformation to minimise the residual sum of squares differences between the images and the template. A study-specific T1 template was created from the mean image calculated from all the normalised T1 images and smoothed at 8-mm full-width at half maximum (FWHM).

To generate study-specific brain tissue a priori maps, the normalised images were segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF)

using an SPM cluster analysis with a modified mixture model and the SPM-derived GM, WM and CSF a-priori probability maps. A brain tissue mask was produced using the "Xtract brain" function. This removes the extra-cerebral voxels from both the GM and WM segmented images using a series of dilation functions and adds together the segments forming a binary image of extracted brain tissue. Multiplication of this image with the original segmented images removes the extra-cerebral voxels. Mean images were calculated and extracted GM and WM segments were smoothed at 8-mm FWHM to produce study-specific a priori maps.

4.3.2 FePsy study-specific templates and priors

Study-specific templates and customized prior probability maps were constructed based on all 82 subjects (i.e. 35 ARMS, 22 HC and 25 FE). Again, scans were normalised to the generic SPM T1 template using 12-point linear affine transformation to minimise the residual sum of squares differences between the images and the template. A study-specific T1 template was created from the mean image calculated from all the normalised T1 images and smoothed at 8-mm full-width at half maximum (FWHM).

The normalised images were, then, segmented into GM, WM and CSF using the generative segmentation approach in SPM5 that involves image registration, bias correction and tissue classification all combined in the same probabilistic model (Ashburner & Friston, 2005; for more details see chapter 2). To generate study-specific brain tissue a priori maps, mean images for the normalized GM, WM and CSF segments were produced and then smoothed at 8-mm full-width at half maximum (FWHM).

4.3.3 Image acquisition

Details of the scanning protocols are given in relevant subsections of Chapters 5 and 6.

4.3.4 Image processing pipeline

Before any pre-processing, I manually set the origin of the images to the anterior commissure using the 'Display' function in SPM. The pre-processing stages for both datasets were identical and were as follows. Firstly, T1 brain scans were segmented (using the unified segmentation approach in SPM5) in native space into GM, WM and CSF using study-specific images for GM, WM and CSF, after which the SPM brain extraction function returned a tissue mask for each scan. These masks were, then, applied to the original T1 images in order to remove non-brain tissue. T1 brain images were, then, spatially normalized to the study-specific T1 template using a 12-parameter linear affine transformation. Bilinear interpolation was used to resample the normalized images and write MNI-normalized images into the stereotactic space at a 1x1x1mm voxel resolution. These normalized images were again segmented using the study-specific a priori templates and spatially normalized segments for GM, WM and CFS were returned. Finally, the spatially normalized, segmented images were smoothed with an 8mm full-width at maximum (FWHM) isotropic Gaussian kernel to further 'remove' differentiation patterns between subjects.

4.4 Pattern Classification Analysis

4.4.1 Support Vector Machine

This section gives a brief introduction to Support Vector Machine (SVM). Despite the fact that one cannot claim that SVM is universally the best classification method

to date, it has, however, been widely used in neuroimaging-based studies of various psychiatric disorders (Kloppel et al. 2008a, Kloppel et al. 2008b, Mourao-Miranda et al. 2011, Fu et al. 2008, Davatzikos et al. 2011) including schizophrenia (Fan et al. 2007, Fan et al. 2008a, Davatzikos et al. 2005b, Koutsouleris et al. 2009b, Koutsouleris et al. 2012a, Nieuwenhuis et al. 2012, Mourao-Miranda et al. 2012a, Zanetti et al. 2013, Schnack et al. 2014, Zhang et al. 2015, Koutsouleris et al. 2015a) and has thus far delivered rather satisfactory classification results. For a critical review on the application of SVM in psychiatric disorders, see Orru et al. 2012. Therefore, the reason for selecting this classification method lies, on the one hand, on its wide use by the neuroimaging community and, on the other hand, on the fact that it can deal effectively with high dimensional data and deliver good classification performance. The diagnostic and prognostic accuracy of the SVM models implemented in schizophrenia studies has ranged from 70% (Mourao-Miranda et al. 2012a, Nieuwenhuis et al. 2012) to over 90% (Fan et al. 2007, Yoon et al. 2007, Costafreda et al. 2011, Koutsouleris et al. 2012b), which is very impressive considering the high-dimensionality of the brain imaging data, with thousands of voxels being used as input in the classifiers. For a detailed description and discussion of the variability in the performance of the SVM models in schizophrenia research, one can go back to Chapter 3, section 3.3.1. The rest of this section comprises of a description of the rationale and the mathematical background of SVM models.

The SVM methodology is based on statistical learning theory as originally proposed by Vapnik and Chervonenkis (1971). In the original implementation described in Vapnik (1995) SVM was a linear classifier but other modifications exist that describe the classification for non-linear cases (Boser et al. 1992).

The main goal of SVM is to construct a decision boundary that would allow the classification of individual observations into two (or more) distinct classes, while simultaneously trying to maximise the margin between those classes (training stage). A hypothetical example of a binary SVM classifier is illustrated in Figure 4.1. The optimal separating hyperplane is determined by maximising the distance between the nearest data instances of opposite classes (i.e. the support vectors; Figure4.1b). Once

determined, the optimal hyperplane can be used to classify new, previously unseen data instances (testing phase).

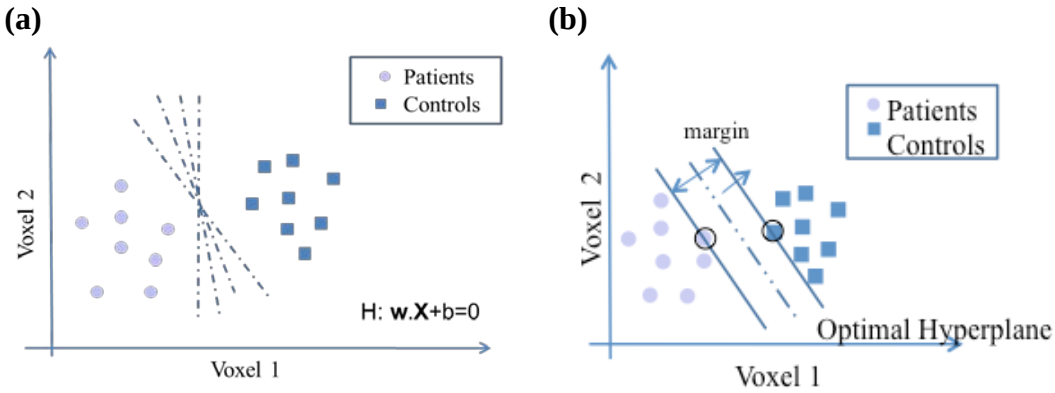


Figure 4.1 Representation of a linear, binary SVM classifier. (a) Illustration of the classification problem between two groups (i.e. circles represent patients, squares represent healthy controls) for the simplified case of two voxels. The dashed lines represent a subset of possible separating hyperplanes, described by a weight vector **w** and an offset b. (b) The optimal separating hyperplane is the one with the largest margin of separation between the two groups and is described as a function of f(x)= w*x+b, where w is a weight vector that is normal to the hyperplane, b is an offset and b/ ||w|| is the distance from the hyperplane to the origin. Points in the dashed lines represent the support vectors. During the training phase, the SVM classifier computes the optimal decision function f(x) and in the testing phase, this decision boundary is applied to new data instances.

Here, a linear support vector machine (SVM) classifier was used for the classification task.

4.4.1.1 Linear Support Vector Machines

A linear binary SVM implies that data instances of two classes can be separated by either a line (in case of two dimensions) or a hyperplane (in case of more than two dimensions). Each data instance is represented by an n-dimensional vector of

features and each instance belongs to only one of the two classes (i.e. +1 for the positive class; patients and -1 for the negative class, i.e. control subjects).

Now, let's assume that we have l training instances, where each data instance x_i has n attributes, and belongs to one of the two possible classes y_i = {-1, +1}. This set of instance-label pairs is called *training set* and is of the form:

$$(x_i, y_i) \text{ where } i = 1....l,\ y_i \in (-1,+1)^l \text{ and } x \in R^n \quad (1)$$

In the training phase of the SVM, the classifier computes a decision function (hyperplane) that maximises the distance between the opposite classes. This linear decision function is based on the *linear discriminant* function of the form: $f(x_i) = \vec{w} \cdot \vec{x}_i + b$, where w is a weight vector, normal to the hyperplane, b is a bias term and $\frac{b}{\|\vec{w}\|}$ is the distance from the hyperplane to the origin (Figure 4.1). As it is observed in Figure 4.1, the margin between the classes is equal to $\frac{1}{\|\vec{w}\|}$, however maximising this quantity is equivalent to minimising $\frac{1}{2}\|\vec{w}\|^2$.

For each class (y=+1, y=-1), one needs to solve:

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \text{ when } y_i = +1 \quad (2)$$

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \text{ when } y_i = -1 \quad (3)$$

Equations (2) and (3) can be combined to: $y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1$

This problem now becomes a primal optimisation one and we therefore need to:

$$min \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \quad \geq 0 \quad \forall_i \quad (4)$$

The above is an optimization problem with a convex quadratic objective and only

linear contraints and its solution gives the optimal margin classifier. The constraint in this formulation ensures that this maximum-margin SVM classifies data instances correctly and no data points of the one class lie on the opposite class. In practice, however, data are not often 'exactly' linearly separable. A greater margin can, thus, be achieved by allowing some misclassification. To allow some errors in the data classification the optimisation problem, in Equation (4), now becomes:

$$min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \quad \text{s.t.} \quad y_i\left(\vec{x}_i\cdot\vec{w}+b\right)\geq 1-\xi_i \quad \xi_i \geq 0, \forall_i \tag{5}$$

where $C > 0$ is a regularisation parameter of the error term (or else a penalty parameter that controls the trade-off between having no training errors and allowing some misclassification) and ξ_i is a positive slack variable that allows misclassification of points.

If $0 \leq \xi_i \leq 1$, then the corresponding data point lies in the margin, while if $\xi_i \geq 1$ the data point is misclassified. This formulation is called soft-margin SVM.

In order to solve ths constrained optimization problem one can, also, use the Lagrangian multipliers *a* and thus obtain the dual formulation of the optimization problem expressed as:

$$max \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j a_i a_j \vec{x}_i^T \cdot \vec{x}_j \quad \text{subject to} \quad \forall a_i \geq 0\,,\ \sum_{i=1}^{l} y_i a_i = 0 \tag{6}$$

This is also a convex quadratic optimization problem but with N variables (a_i, i=1,.., *L*) where *L* is the number of samples. The dual formulation leads to the expression of w as a linear combination of the training vectors in the form: $w = \sum_{i=1}^{l} a_i y_i \vec{x}_i$. The dual formulation is computationally more efficient as the decision function is only

computed by a subset of the training data for which: $0 \leq a_i \leq C$, and additionally there is no need to access the original data but only dot products. Data instances from this subset are called *support vectors* and are data points that lie upon the separating hyperplane (see Figure 4.1). For a complete description of the derivation of the dual formulation please see Appendix II.

Additionally, the dual formulation problem lends itself to the kernel trick (Aizerman et al. 1964), where instead of computing the ordinary dot product between data instances; one can only compute the kernel function:

$$max \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j a_i a_j \mathrm{K}(\vec{x}_i^T, \vec{x}_j) \quad \text{subject to} \quad \forall a_i \geq 0 \ , \ \sum_{i=1}^{l} y_i a_i = 0 \qquad (7)$$

The use of kernels is highly efficient in order to get optimal margin classifiers in high-dimensional space and especially in cases where the data are not linearly seperable. However, the mathematical description of non-lineal SVM models is outside the scope of this chapter.

Having computed the decision function, each new previously unseen data instance x' can be classified to one of the two classes by computing the decision function $sgn\left(\vec{w} \cdot \vec{x}' + b\right)$.

4.4.1.2 LIBSVM

There are a wide range of SVM toolboxes available. We used LIBSVM library for the implementation of the linear SVM classifier (Chang & Lin 2011), running in Matlab version 7. This toolbox requires that each data instance is represented as a vector of real numbers, so transformation of the original data set into LIBSVM

format was necessary before running the method (Hsu et al. 2010). In addition, scaling to [0, 1] range was performed in order to ensure that attributes in greater numeric ranges would not dominate those in smaller ranges.

A 'grid-search' on the *C* parameter using cross-validation was also performed. Various *C* values were tried and the one that gave the best cross-validation accuracy was picked. More details on the C parameter can be found in the following results chapters of this thesis.

4.4.2 Feature Extraction

As described in Chapter 3, feature extraction is a fundamental and obligatory step in every pattern recognition image analysis pipeline.

Prior to using the SVM, all smoothed and normalized GM maps were mapped to the Automated Anatomical Labeling (AAL) brain atlas (Tzourio-Mazoyer et al. 2002). The AAL template is an atlas that provides a parcellation of the human brain in 116 brain regions and serves as a basis for the correspondence between sets of co-ordinates and their anatomical labels (Figure 4.2). For its construction, the AAL atlas was based on the spatially normalized MNI single-subject MRI brain template, which was obtained by repeatedly scanning a young man (27 times) and averaging each MRI acquisition to produce the final MRI brain template. For a detailed description of the delineation and parcellation process, one can see the Tzourio-Mazoyer et al. (2002) paper.

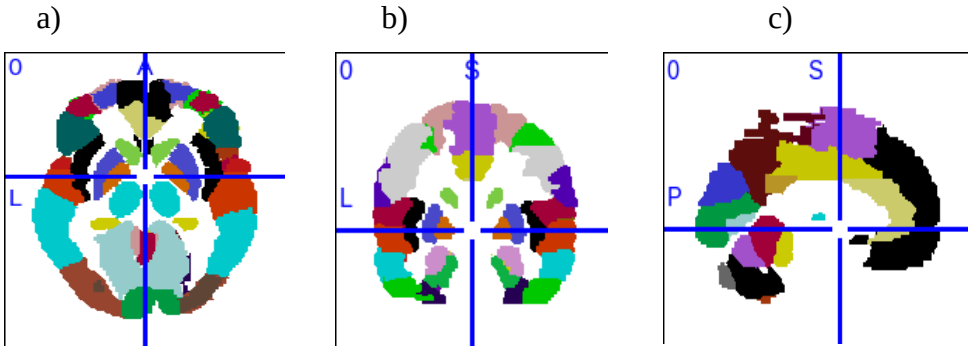


Figure 4.2 Illustration of the AAL brain atlas. a) axial view, b) coronal view and c) sagittal view.

A script in C programming language (Kernighan & Ritchie 1978) was implemented in order to extract the GM density volumes corresponding to the 116 pre-defined brain regions of the AAL template from each smoothed and normalized GM map in the EHRS and the FePsy datasets. Specifically, each subject's smoothed and normalized GM image was mapped to the AAL-atlas template, where in each AAL-specified region the intensity values of the voxels that comprised the brain region were summed up and their mean intensity was derived. These (mean) GM density values were, then, used as input features into the SVM classifier.

For reasons of disc memory and speed efficiency, it was decided to use a ROI-based approach instead of a voxel-based one. The C code for mapping and extracting mean intensity values was already available in the lab. There are a number of advantages to choosing such an approach, which include a faster calculation of the separating SVM hyperplane, since the number of features in the ROI-based classification is significantly lower than in the voxel-based case, a faster and more efficient feature selection and interpretation of the involved brain regions. On the other hand, because the signal intensity is averaged in the AAL-defined ROI, this could possibly lead to loss of information since there is no way one could detect differences or significant information in voxels or voxel clusters (and subregions) within the ROI.

4.4.3 Feature selection

In machine learning, feature selection is an important step in the analysis pipeline for a number of reasons: meeting computational and time constraints, facilitating the interpretation and visualization of results and enhancing the generalization performance (Guyon & Elisseeff 2003). A variety of feature selection methods exist. Filters are the simplest and most computationally efficient feature selection methods, in which features are ranked and eliminated based on a selection criterion during the pre-processing stage and just before the data enter the classifier. Filter methods are, therefore, independent of the classifier. The main disadvantage of filters is that these methods do not examine features jointly with respect to how significant they are in the classification task (as a group) but rather rank them independently of one another. Thus, filter methods are not well suited for situations where apart from removing redundancy, one is also interested in increasing classification performance because a group of features might be able to provide better accuracy than another and filter methods cannot examine this in any way. To this end, embedded methods are better suited because they aim to find a combination or a subset of features that would provide maximum classification accuracy. To do that, embedded methods utilize the classifier to exmamine the quality of the features by incorporating variable selection as part of the training process of the classifier and by implementing a ranking criterion that is bound to the cross-validation performance.

In the present work, a series of feature selection methods was tried out. Initially filter methods, such as the F-score (Chen and Lin, 2006) and the Pearson Correlation coefficient, which is closely related to the t-Test criterion (Guyon & Elisseeff 2003), were examined. A detailed descrition of these filter methods can be found in Appendix III. However, both of these methods failed to increase the classification accuracy (both when tried with linear and nonlinear SVM classifiers), which was one of the reasons for applying feature selection in the first place, apart form removing redundancy.

An embedded feature selection method, called Recursive Feature Elimination (RFE),

was then put to the test. It was observed that the classification performance was significantly improved after implementing the Recursive Feature Elimination, as opposed to not incorporating it in to the SVM classifier. However, for the sake of consistency, the performance of the SVM classifier with and without RFE will be given in the following results chapters. A detailed descrition of the RFE methodology is given below.

4.4.3.1 Recursive Feature Elimination

Recursive Feature Elimination (RFE) is one of the most effective feature selection methods, proposed and originally implemented by Guyon et al. (2002) for gene selection in cancer classification. It is a backward sequential feature elimination approach, embedded in an SVM classifier that selects features based on their influence on the determination of the maximum margin hyperplane.

The algorithm starts with all feature variables in the training set and gradually removes one feature at a time (or for speed reasons, more than one feature could be eliminated). At each step, a linear SVM classifier is trained using a subset of the original feature set and the coefficients of the **w** vector are used to compute a feature ranking score. For the linear case, the feature ranking criteria is the: $c_i = (w_i)^2$, for all features i. Based on this ranking criteria, the feature with the smallest ranking score is eliminated from the feature set and the method iterates in the same way until a specified number of features have been eliminated or until the classifier reaches its best classification performance (see Panel 1). However, in this implementation, the RFE method continued until all features were eliminated from the feature set while at each elimination step the performance on the validation step was recorded so that the (smaller) subset of features that achieved the highest performance on the validation set was determined and retrieved. Feature elimination based on this ranking criterion corresponds to eliminating that feature whose removal changes the objective function the least, thus, having the least contribution on the margin definition. Recall from the previous section, the objective function is $J = \frac{1}{2}\|w\|^2$, which justifies the use of (w_i^2)

as the ranking criterion. This approach is based on linear SVM, although other implementations of the methodology exist for non-linear classification problems (Rakotomamonjy 2003).

Panel 1. Algorithm for the Recursive Feature Elimination.

1. Start: training examples $X_0 = [x_1, x_2, ...x_k, ..x_l]^T$
   class labels $y = [y_1, y_2, ..y_k, ..y_l]^T$
   ranked feature subset R=[],
   surviving feature subset S=[1,2,.., n]
2. Repeat until S=[]
   a) train a linear SVM with features in set S
   b) compute the weight coefficients
   c) compute the ranking score $c_i = (w_i)^2$ for each feature in S
   d) find the feature with the smallest ranking score $f = argmin(c_i)$
   e) Update the R=[f, R] and S=S-[f].
3. Output: Ranked feature list R

4.4.4 Classification performance

Measures of classification performance are needed in order to assess the generalizability of the classifier to independent testing cases. The most widely used metric is the accuracy, which measures the proportion of correctly predicted data to the total number of the test data.

In the following subsections I present the performance metrics employed in this study and also describe the technique of cross-validation, which was used in order to assess the classifier's performance.

4.4.4.1 Performance metrics

The performance of a classification method can be visualised using a confusion matrix (Table 4.1), which contains information about the actual and predicted instances classified by the chosen classifier (Kohavi & Provost, 1998). The correctly labelled instances are located on the diagonal and consist of the *true positives (TP)* that refer to correctly classified patients and the *true negatives (TN)* that represent the correctly labelled controls. Instances that are incorrectly labelled are divided in *false positives (FP)* and *false negatives (FN)* that describe the proportion of control subjects incorrectly labelled as patients and the number of patients incorrectly identified as controls respectively.

		Predicted	
		Positive	Negative
True	Positive	True Positives (TP)	False Negatives (FN)
	Negative	False Positives (FP)	True Negatives (TN)

Table 4.1 Confusion matrix of a binary classifier. The confusion matrix aims to classify positive cases (i.e. patients) and negative examples (i.e. control subjects).

The accuracy is the proportion of the total number of correct classifications and is given by the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\,\%$$

In cases where the two classes are unbalanced, the accuracy is not a very reliable measure of the classification performance. It is, thus, important to report other performance metrics.

Sensitivity refers to the proportion of the actual cases in the positive class correctly classified and is given by:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\ \%$$

Specificity refers to the proportion of actual cases in the negative class correctly classified and is computed by:

$$Specificity = \frac{TN}{TN + FP} \times 100\ \%$$

Positive predictive value (PPV) gives the probability that a person identified by the classifier as having a disease actually has the disease:

$$PPV = \frac{TP}{TP + FP} \times 100\ \%$$

Negative predictive value (NPV) refers to the probability that a person identified as free of a disease truly does not have it and is given by the:

$$NPV = \frac{TN}{TN + FN} \times 100\ \%$$

4.4.4.2 Cross-validation

In cases where the data set is large enough, one would be advised to split the set into a training set and a holdout set that would serve for evaluating the classifier's performance (test set). However, it is not always 'affordable' to set aside a portion of the original data set for testing purposes, especially when this data set consists of limited cases; as is often the case in neuroimaging studies. To this end, cross-validation is usually performed (Kohavi 1995).

Here, a leave-one-out (LOO) cross-validation (CV) approach was employed. Leave-one-out CV can effectively deal with over-fitting and provide an almost unbiased estimate of the generalisation error. It involves a repeating procedure, where in each trial data from all but one subject (k-1 of k subjects) are used for training and the left-

out subject is used for testing the classifier's performance. This procedure is repeated k times, until all data instances have been used in the test set and then classification accuracy is averaged across all runs.

4.4.5 Permutation Testing

Permutation testing was used to evaluate the probability of obtaining accuracy levels higher than the ones obtained during the cross-validation procedure. An illustration of how permutation tests work is given in Figure 4.3. This test is used to derive a p-value that gives an estimate of the statistical significance of the classifier and determines whether classification accuracy exceeded chance levels (50%). After having completed the RFE procedure, and using the final set of selected features, I permuted the class labels 1000 times, by randomly assigning patient and control labels to the training subjects. I then counted the number of times the permuted accuracy exceeded the one obtained for the true labels and by dividing this number by 1000, the p-value for the classification was derived.

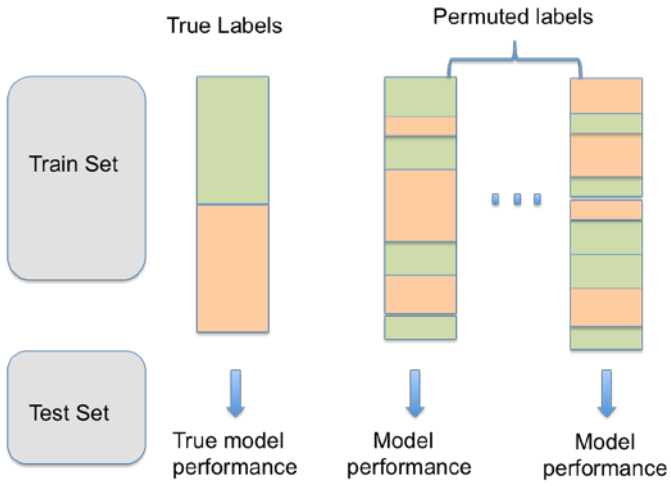


Figure 4.3 Permutation testing. The significance of a classifier's performance is assessed using permutations tests. The 'true' labels are randomly permuted to obtain "baseline" model performances. The "true" model performance is then compared to the baseline level obtained from the permutations. Typically, a model performance is significant if the performance obtained by chance does not exceed or is equal to the true model performance more than 5% of the time ($p < 0.05$).

CHAPTER 5

Individualized prediction of schizophrenia in subjects at high familial risk

5.1 Introduction

As seen in previous chapters, schizophrenia is associated with a number of brain abnormalities. However, it remains unclear whether these brain abnormalities occur prior to disease onset, or in high-risk individuals coming from multiply affected families or only those high-risk subjects that go on to develop schizophrenia. The EHRS was conceptualized with the aim of shedding light to those questions and identify predictors of vulnerability and transition to schizophrenia. Findings from previous EHRS studies using univariate methods of data analysis have been discussed in previous chapters.

The scope of this chapter is to investigate the utility of baseline structural MRI data in predicting transition to schizophrenia in high-risk subjects, who took part in the EHRS and were identified at a heightened risk for developing psychosis for familial reasons, using a Support Vector Machine (SVM) approach. Additionally, the diagnostic performance of the classifier was examined by combining baseline neuroanatomical information with baseline schizotypal and neurocognitive variables that were previously reported as good predictors for the development of schizophrenia. It was hypothesized high risk subjects who later go on to develop schizophrenia could be distinguished at baseline from HR subjects who do not and that a combination of structural brain imaging, schizotypal and neurocognitive data would enhance predictive performance compared to either individual measures alone.

5.2 Background

Psychiatric research interest has recently shifted from studying schizophrenia to attempting an early detection of the disorder before diagnosis of full-blown psychosis can be established. Making an early diagnosis of schizophrenia and other psychoses is, however, quite challenging mainly because of the underlying difficulty of identifying an 'at-risk' state for the disorder (Riecher-Rossler et al. 2006).

As seen in previous chapters, a plethora of risk factors for psychosis have been identified; the most predominant of which include the existence of subclinical psychotic symptoms and genetic risk. Individuals conceptualised as at clinical HR (or as having an 'at-risk mental state'- ARMS) present with early signs of psychosis such as attenuated positive symptoms or brief limited intermittent psychotic symptoms (Fusar-Poli et al. 2013a, Phillips et al. 2000). In familial HR paradigms, individuals are characterised on the basis of having at least one affected first or second-degree relative (Cannon & Mednick 1993, Erlenmeyer-Kimling et al. 1997, Mirsky et al. 1995). Several aspects of the familial and clinical HR paradigms, however, overlap as many individuals at familial HR often exhibit pre-psychotic symptoms and clinical HR subjects might also have a family history of the disorder.

Nevertheless, both HR paradigms are associated with an increased risk for development of psychosis and share various neurocognitive and neuroanatomical alterations (Smieskova et al. 2013), some of which might even be predictive of future disease transition (Lawrie et al. 2001b, Brewer et al. 2006, Smieskova et al. 2013).

Neurocognitive dysfunctions are cardinal to schizophrenia (Heinrichs & Zakzanis 1998) and some of them can even be detected early in the course of the disease (Fusar-Poli et al. 2012a, Seidman et al. 2006). Previous studies reported significant neurocognitive impairments in the domains of attention, working memory, executive function, verbal learning/fluency and memory in HR compared to healthy control (HC**)** subjects (Seidman et al. 2006, O'Connor et al. 2009, Brewer et al. 2005, Pukrop et al. 2006, Byrne et al. 2003, Gschwandtner et al. 2003, Wood et al. 2003). Reduced baseline performance in verbal fluency and memory was also found to distinguish HR subjects who later converted to psychosis from those who did not, both in clinical (Fusar-Poli et al. 2012a, Pukrop et al. 2007, Lencz et al. 2006) and familial HR paradigms (Johnstone et al. 2005, Erlenmeyer-Kimling et al. 2000, Whyte et al. 2006). Furthermore, prediction performance of multivariate models was shown to improve, up to 80% accuracy (Lencz et al. 2006), when neurocognitive functioning and symptom severity scores were integrated in a stepwise risk assessment (Riecher-Rossler et al. 2009, Lencz et al. 2006). Studies employing prognostic index, in a multilevel context that combines individualized risk estimation and stratification,

have also reported remarkable results in predicting transition to psychosis. The combination of data from different domains such as neuropsychological data and clinical HR criteria (Ruhrmann et al. 2010) or neurophysiology measures with premorbid adjustment (Nieman et al. 2014), has resulted in positive predictive value of up to 83% (Ruhrmann et al. 2010) and as a second step enabled an individualized risk classification in terms of magnitude and time to transition, that has an added benefit for informing targeted prevention strategies.

As seen in previous chapters, VBM studies have shown that schizophrenia is associated with grey matter (GM) volume reductions in frontal, superior temporal and anterior cingulate gyri and medial temporal lobe structures (Bora et al. 2011). Subtle GM reductions in prefrontal, medial temporal, limbic and temporo-parietal regions were also found across the whole HR population (Smieskova et al. 2010, Borgwardt et al. 2011, Meisenzahl et al. 2008b, Job et al. 2003, Pantelis et al. 2003) suggesting that they might be correlates of vulnerability to psychosis (Fusar-Poli et al. 2011). Transition to schizophrenia was found to be associated with further GM volume reductions in temporal, cingulate and cerebellar regions in clinical HR samples (Borgwardt et al. 2006, Koutsouleris et al. 2009a, Pantelis et al. 2003). In our own EHRS, progressive GM reductions in the inferior temporal gyrus, uncus and right cerebellum were reported in those HR subjects that later transited to schizophrenia versus those who did not (Job et al. 2006).

What is more needed in clinical practice, however, is moving from a characterisation of group differences to models that could simultaneously capture individual variations from the norm while making inferences at a single-subject level. To this end, pattern classification techniques, including the Support Vector Machine (SVM), have emerged as powerful brain image analysis tools and have already been applied in diagnosing various neurological and psychiatric disorders (Kloppel et al. 2012, Orru et al. 2012, Davatzikos et al. 2005b, Kloppel et al. 2008a, Mourão-Miranda et al. 2011). As seen in the literature review in Chapter 3, SVM has achieved high diagnostic accuracy, ranging from 77% (Koutsouleris et al. 2012b) to 84% (Koutsouleris et al. 2012a) in predicting future transition to psychosis in ARMS subjects, either using neurocognitive (Koutsouleris et al. 2012b) or neuroanatomical

data (Koutsouleris et al. 2012a, Koutsouleris et al. 2009b).

However, these studies have only focussed on using a single type of data at a time. A limited number of studies have examined the diagnostic performance of combined data in distinguishing schizophrenia or first episode patients (Pettersson-Yeo et al. 2013) against HC, mostly by combining genetic data with either functional or structural MRI (Yang et al. 2010) or clinical information (Struyf et al. 2008), and have produced quite encouraging results (Sui et al. 2012). To our knowledge, only one study has combined structural MRI with neurocognitive data (Karageorgiou et al. 2011), reporting good classification performance (89% sensitivity and 93% specificity) in the context of distinguishing recent-onset schizophrenia patients from HC subjects.

5.3 Materials and Methods

5.3.1 Subjects

The subject material was gathered as part of the Edinburgh High Risk Study (EHRS). A detailed description of the EHRS recruitment and assessment procedures was given in Chapter 4, section 4.2.1, and can also be found in previous papers (Hodges et al. 1999, Johnstone et al. 2000).

Briefly, 160 individuals aged 16-25 years, with no previous history of psychiatric problems, were drawn throughout Scotland and were identified as HR on the basis of having two or more relatives affected with schizophrenia. Subjects were followed up for up to 10 years during which they underwent a series of clinical, behavioural and neuroimaging assessments every 18 months.

Psychopathology was assessed at entry and follow-up by the Present State Examination (PSE; Wing et al. 1974) and allowed the classification of subjects into five categories: 0, no symptoms; 1, non-psychotic symptoms; 2, partially held psychotic symptoms; 3, definite but isolated and/or transient psychotic symptoms

and 4, diagnosis of schizophrenia, which was further validated by the ICD-10.

All subjects were antipsychotic-naive at study-entry and at follow-up or until they were clinically diagnosed with schizophrenia. From those HR subjects who provided complete data and had a MRI scan, 17 were diagnosed at follow-up with schizophrenia (after an average of 929 days, SD=138). Among the rest, 57 subjects experienced psychotic or possibly psychotic symptoms (points 2 and 3 combined) but these were too transient or mild to satisfy operational definition for schizophrenia or any related psychotic illness (Johnstone et al. 2000). The rest of the HR subjects remained well, with no symptoms.

In this study, the aim was to contrast the two groups who exhibited psychotic symptoms and were the most difficult to discriminate against on any basis: the one group of subjects that surpassed the clinical threshold for schizophrenia (HR[ill]) and the other remaining below the threshold for clinical diagnosis (HR[symp]) but yet exhibited psychotic symptoms. In that way, our study groups could indirectly relate to other studies in this area, which study ARMS individuals who typically present psychotic symptoms, although these were identified based on different assessment criteria than the present study.

To build the classifier, 17 of the 57 HR[symp] subjects were randomly selected to contrast with the 17 HR[ill] subjects (Table 5.1), a practice seen in other studies in this area (Koutsouleris et al. 2015b, Mourao-Miranda et al. 2012a, Koutsouleris et al. 2010, Koutsouleris et al. 2009b), primarily in order to alleviate class imbalance issues that can result in a classification bias towards the majority class (Japkowicz 2000, Akbani et al. 2004, Weiss 2004). The data of the remaining 40 subjects were used to further validate the classification (Table 5.2- $HR[symp]_{test}$), after training and cross-validation. A more detailed discussion on the reasons for choosing this technique to manage the class imbalance can be found in section 5.4.4 of this chapter.

5.3.2 Schizotypal and Neurocognitive Measures

We aimed to include in our analysis those variables that constituted the best predictors of schizophrenia in the EHRS literature (Johnstone et al. 2005), namely the Rust Inventory of Schizotypal Cognitions (RISC) questionnaire (Miller et al. 2002b) and the Rey Auditory Verbal Learning Test (RAVLT).

Briefly, the RISC is a self-completed, 26-item questionnaire that evaluates schizotypy and schizotypal cognitions, rather than overt psychotic symptoms, primarily associated with the positive symptoms of schizophrenia (Rust 1988). The areas tapped include ritualistic thinking, psychotic symptoms such as delusions of grandeur, schizophrenic symptoms such as auditory hallucinations and defense mechanisms such as reaction to disturbing ideas. A few indicative examples of the cognitions measured in the RISC are: 'I never use a lucky charm' and 'sometimes I get a weird feeling that I am not really here'. Here, a uni-dimensional test score was calculated for each participant at study-entry through the cumulative summation of the item scales.

The RAVLT instrument evaluates verbal memory and learning, by repeatedly (5 times) presenting the individual with a list of words and asking to recall as many as he/she can (Byrne et al. 2003). The final test score produced is equal to the total number of words recalled across all five trials. Previously, an equal mis-classification costs cut-off on the RISC score showed 94% negative and 50% positive predictive power while the same approach to the RAVLT trials 1-5 total score showed figures of 85% and 11% respectively (optimal cut-off points: 39.5 and 48.5 respectively; Johnstone et al. 2005).

5.3.3 Image Acquisition and Preprocessing

In the present thesis, only baseline neuroimaging scans were considered. In the EHRS study, all baseline scanning was performed on a 1.0 T Siemens Magnetom scanner (Erlangen, Germany). After localisation and a double spin-echo sequence to identify any gross brain abnormalities, T1-weighted MRI scans were acquired using a three-dimensional Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence with the following parameters: TR=10 ms, TE=4 ms, TI=200 ms, relaxation delay time=500 ms, flip angle= 12^{o}, FOV= 250mm x 250mm, resulting in 128 contiguous, 1.88-mm thick "slices". To correct for inhomogeneity of the RF coil and any scanner changes over time, an oil phantom was scanned immediately after each subject, using the same coil and in the same orientation as the subject's head (Job et al. 2002). An example of a typical structural MRI scan of a high-risk individual in the EHRS is given in Figure 5.1.

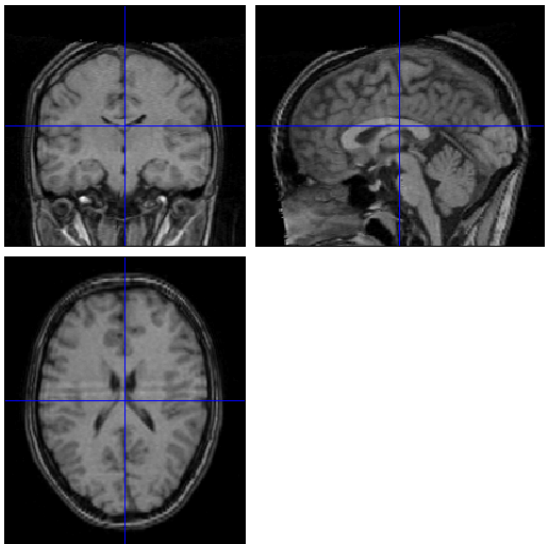


Figure 5.1 A typical MRI scan in the EHRS dataset.

A detailed description of the steps in the preprocessing pipeline can be found in Chapter 4, section 4.2.

Table 5.1 Baseline sociodemograpchic and behavioural assessment variables of the study groups

	HR[ill]	**HR[symp]**	***P***
Number of participants	17	17	
Mean age at baseline (SD)	20.07(2.37)	20.03(2.6)	ns[a]
Male (%)	11(64.7)	11(64.7)	ns[b]
Mean RISC score (SD)	39.88(10.6)	25.23(11.75)	<0.01[a]
Mean RAVLT, trials 1-5 (SD)	47.64(7.49)	53.41(7.45)	ns[a] **
Mean WAIS-IQ (SD)	98.64(12.93)	98.98(14.7)	ns[a]
Handedness			ns[b]
Right	16	15	
Left	0	1	
Mixed	1	1	
Social Class of origin			ns[b]
I and II	2	3	
III and IV	13	10	
V and VI	2	3	
unclassifiable	0	1	
Cannabis use at baseline			ns[b]
None	8	12	
Occasional	5	4	
Frequent	4	1	
Smoking cigarettes			ns[b]
None	7	9	
<10	6	4	
10-20	3	2	
>20	1	2	
Symptoms severity (PSE rating)			ns[b]
No psychotic symptoms	2	7	
Neurotic symptoms only	4	3	
Partially held psychotic symptoms	9	6	
Isolated and/or transient psychotic symptoms	2	1	

HR[ill]: individuals at high familial risk who developed schizophrenia during follow-up period; HR[symp]: individuals at high familial risk who remained well but developed psychotic symptoms during follow-up period; IQ, Intelligence Quotient; RISC, Rust Inventory of Schizotypal Cognitions; RAVLT, Rey Auditory Verbal Learning Test; WAIS-R, Wechsler Adult Intelligence Scale- Revised. Social class of origin was based on the

father's occupation at the time of the subject's birth using the Occupational Classification of the Registrar General (HMSO 19991). [a] Student's *T*-test. [b] Fisher's exact test. [**]effect size was r=0.54

Table 5.2 Socio-demographic and behavioural variables for the two HR[symp] groups

	HR[symp]	**HR[symp]$_{test}$**	***P***
Number of participants	17	40	
Mean age at baseline (SD)	20.03(2.6)	21.07(4.59)	ns[a]
Male (%)	11(64.7)	14(35)	ns[b]
Mean RISC score (SD)	25.23(11.75)	33.45(13.39)	ns[a]
Mean RAVLT, trials 1-5 (SD)	53.41(7.45)	51.15(11.59)	ns[a]
Mean WAIS-IQ (SD)	98.98(14.7)	96.42(12.64)	ns[a]
Handedness			ns[b]
Right	15	34	
Left	1	4	
Mixed	1	2	
Social Class of origin			ns[b]
I and II	3	7	
III and IV	10	22	
V and VI	3	10	
unclassifiable	1	1	
Cannabis use at baseline			ns[b]
None	12	26	
Occasional	4	10	
Frequent	1	4	
Smoking cigarettes			ns[b]
None	9	21	
<10	4	12	
10-20	2	4	
>20	2	3	
Symptoms severity (PSE rating)			ns[b]
No psychotic symptoms	7	13	
Neurotic symptoms only	3	13	
Partially held psychotic symptoms	6	13	
Isolated and/or transient psychotic symptoms	1	1	

Note: please see legend of Table 5.1

5.3.4 Multivariate Pattern Classification Analysis

A detailed description of the pattern classification analysis pipeline can be found in Chapter 4, section 4.3.

5.3.4.1 Support Vector Machine

A linear Support Vector Machine (SVM) was used for the classification task. A detailed description of the linear SVM classifier can be found in Chapter 4, section 4.3.1.1. Briefly, a binary SVM classifier works by finding an optimal decision function that separates the two classes. The classification procedure consists of two phases: training and testing. During the training phase, the classifier is trained by providing examples of the form (x_i, y_i), where x_i represent a spatial patter (e.g. grey matter map) and y_i represent class labels (here, HR[ill] and HR[symp]) in order to determine a hyperplane that optimally separates the two groups. Once the decision function is learned, it can be used to predict the class of a new test example.

As described before in chapter 4, the linear SVM has a slack variable that controls for the desired amount of misclassifications. Here, a 'grid-search' on the *C* parameter was performed, as recommended by Hsu et al. 2010, in order to identify the optimal value for the parameter (that is the value that produced the highest accuracy in the cross-validation). However, the grid search was only performed in the inner cross-validation loop of the nested cross-validation scheme (described in more detail below), whereas in the outer cross-validation loop the default *C* value was used (i.e. *C*=1).

Apart from the linear SVM, non-linear SVM classifiers were also tried out. Briefly, when the relationship between the class labels and the features is nonlinear, one can choose to map the original data into a higher dimensional space where the data samples are linearly separable. To do that, the kernel trick is employed (recall from section 4.4.1.1); the most widely used nonlinear kernel is the Radial Basis Function (RBF). In the following Results' section, the results pertaining to the use of the

nonlinear RBF SVM will be briefly presented.

5.3.4.2 Feature Extraction

As previously described, prior to using the SVM, all smoothed and normalized GM maps were mapped to the Automated Anatomical Labeling (AAL) brain atlas (Tzourio-Mazoyer et al. 2002) and GM density volumes corresponding to the 116 brain regions of the template were returned. The 116-length vector of GM density values was then concatenated with the baseline RISC and RAVLT scores and was used as input to the classifier. To ensure commensurability of the different data types, the features were scaled to [0 1] in the training set and then the same scaling was applied to the testing set.

5.3.4.3 Feature Selection

To identify the most significant features in the classification task, recursive feature elimination (RFE; Guyon et al. 2002) was applied. As previously described, RFE is a backward feature selection technique, that recursively removes features from the original dataset with the aim of removing the most redundant features, while keeping the most significant. During RFE, a linear SVM classifier is trained and the method removes one (or more) features based on a feature-ranking criterion that is computed as the square of each weight vector coefficient (wi^2).

Here, apart from removing the least significant features, I was also interested in improving diagnostic performance. A nested leave-one-out cross-validation (LOO-CV) framework was, thus, employed (Mourão-Miranda et al. 2012, DeMartino et al. 2008) in which one subject was removed from the original data set to comprise the test set and then a second split where the remaining subjects were again repeatedly repartitioned to form a validation and a training set was performed (Figure 5.2). The nested LOO-CV provided an unbiased estimate of the expected diagnostic accuracy on new cases.

The testing points had no involvement whatsoever in the RFE and the training process. The feature set that produced the maximal accuracy on the validation set was selected and applied to the testing set. The final accuracy was calculated as the mean accuracy over all test subjects.

The feature set that contained the most significant features in the classification process and produced maximal accuracy was returned. Each feature is associated with a weight vector w_i that gives an indication of the relative importance of each feature in predicting one class over the other. Given the labels +1 and -1 for the HR[ill] and HR[symp] groups respectively, a positive weight vector implies a higher importance of this feature in classifying the HR[ill] group whereas a negative value means this feature is more important in classifying HR[symp] individuals over HR[ill] ones. Figure 5.3 presents an illustration of the most discriminating regions based on their weight vectors (Figure 5.3).

Finally, it should be noted that the implementation of the RFE methodology for nonlinear SVM classifiers was fairly complicated and was not attempted; therefore results pertaining to the nonlinear SVM classifier (and specifically RBF SVM) include a single leave-one-out cross-validation loop. The lower predictive accuracy of the RBF-SVM classifiers and the entailed difficulty in implementing the RFE method were the main reasons for opting out of the nonlinear SVM and continuing with the linear classifiers.

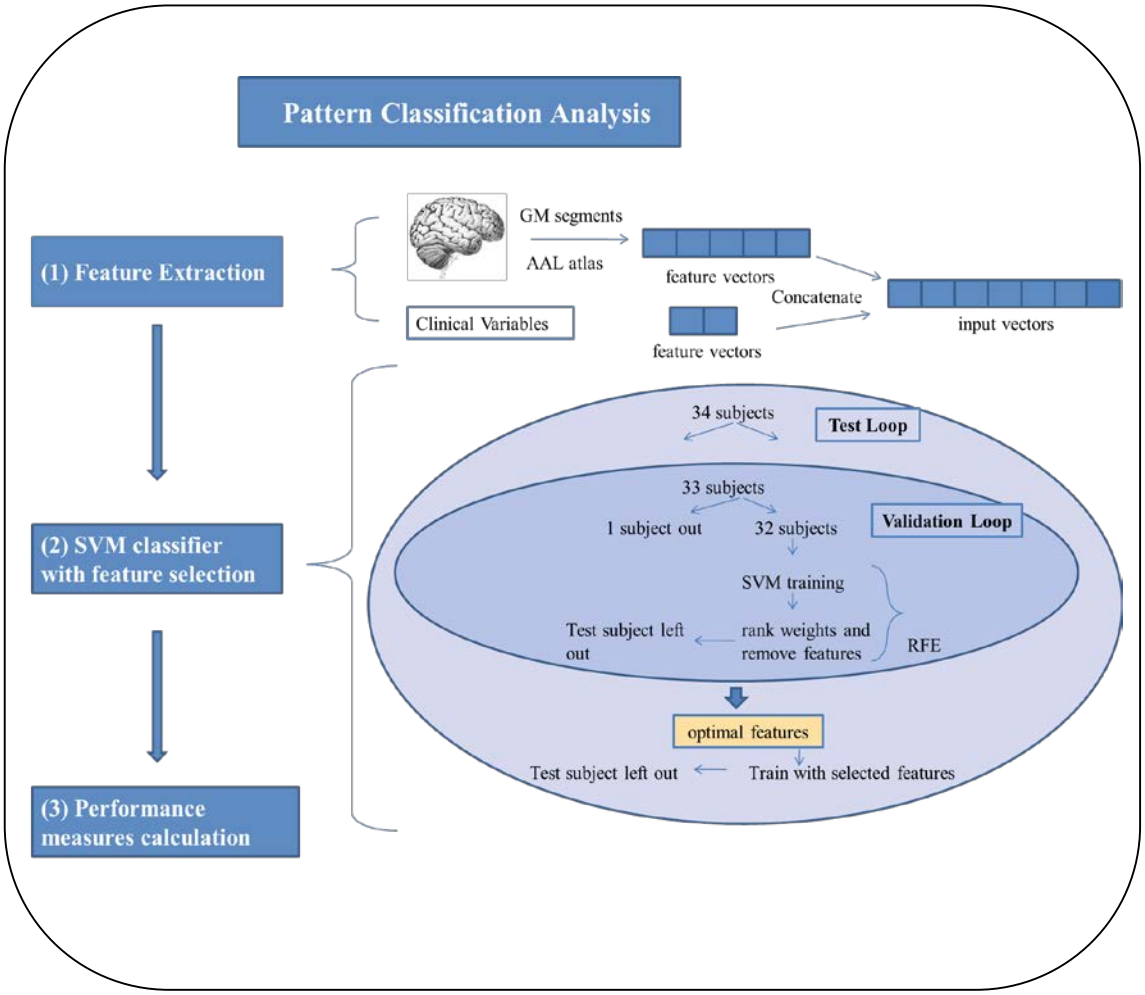


Figure 5.2 Pattern classification analysis pipeline. (1) Feature extraction: after standard scan preprocessing, GM segments were mapped to the AAL atlas and transformed into feature vectors containing GM volume information. In the second analysis, baseline clinical variables were concatenated to form the SVM input vectors. (2) Nested LOO-CV SMV-RFE. A nested LOO-CV was employed where I repeatedly excluded one subject from our data set to comprise the test set and the remaining subjects (33) were again repeatedly repartitioned in an internal validation loop where one subject was left out for validation and the rest (32) formed the training group. In this loop, RFE was repeatedly performed and the mean accuracy on the validation group at each elimination level was recorded until all features were removed. The feature set that produced the maximum accuracy on the validation set was selected and applied to the testing set of the outer testing loop. (3) Mean accuracy, sensitivity and specificity measures were returned.

5.3.4.4 Permutation testing

Permutation testing was performed in order to determine whether our classifier has produced these accuracy levels simply by chance. The rationale and implementation of the permutation testing was described in Chapter 4, section 4.4.5

5.4 Results

Demographic details for the two study groups are presented in Table 5.1. Student's T-tests were performed on the continuous variables (i.e. age, RISC, RAVLT and IQ measures) and Fisher's exact tests were performed on the rest, categorical variables. There were no significant differences in age, handedness, socio-economic background of origin, IQ, smoking, cannabis use and symptom severity at study-entry between the study groups (Table 5.1), except for the baseline RISC score which differed between the HR[ill] and HR[symp] group ($p<0.01$).

Additionally, I compared the 17 HR[symp] subjects against the 40 $HR[symp]_{test}$ subjects that were used to further validate the classification and found no significant differences in any of the reported variables (Table 5.2).

5.4.1 SVM classification based on structural MRI data

The nested LOO-CV linear SVM-RFE framework was applied to neuroanatomical data alone and a combination of neuroanatomical, schizotypy and neurocognitive data.

The SVM approach achieved 88% accuracy in predicting disease conversion based on structural MRI data alone (Table 5.3). Thirteen out of 17 subjects in the HR[ill] and all subjects in the HR[symp] group were correctly assigned to their group, (sensitivity/specificity: 76%/100%; PPV/NPV: 100%/81%; permutation test $p=0.001$).

Without the RFE method, the simple LOO CV linear classifier gave a classification accuracy of 63.5% (sensitivity/specificity: 57%/70%). When applied to the same 17 HR[ill]-17 HR[symp] dataset, the RBF-SVM classifier, using the entire feature set, gave a mediocre performance in predicting transition to schizophrenia, with an accuracy of 55.8%, a sensitivity of 64.7% and a specificity of 47.05%.

The spatially distributed network that discriminated between the two groups was quite extensive and consisted of GM abnormalities in a spatially distributed network covering all four lobes and the cerebellum. Although, with these methods, it is not possible to make local inferences on the discriminating regions, it is noteworthy that the anatomical regions with the highest contribution to the discrimination between groups include the cerebellum, the lateral and medial temporal lobe, the amygdala, the medial frontal lobe, the putamen and the superior parietal lobe covering the supramarginal gyrus bilaterally and extending to the right fusiform gyrus (Table 5.4). Table 5.4 presents a list of the most discriminating regions in the classification task, namely the brain regions with the highest (absolute) weight value that contributed relatively higher to the decision function.

Discrimination maps showing the spatial pattern by which the groups differ are illustrated in Figure 5.3A. It must be emphasized that the discrimination map should not be interpreted as a standard statistical parametric map resulting from a mass-univariate statistical test but rather as a spatial representation of the decision boundary, where no local inferences should be made based on the SVM weights.

.

Table 5.3 Diagnostic performance of the classifier, using sMRI data only.

sMRI analysis	TP	TN	FP	FN	Sens (%)	Spec (%)	BAC (%)	FPR (%)	PPV (%)	NPV (%)
HR[ill] vs HR[symp]	13	17	0	4	76.4	100	88.2	0	100	80.9
HR[symp]test	-	27	13	-	-	67.5	-	32.5	-	100
Overall	13	44	13	4	76.4	77.2	76.8	22.8	50	91.6

Note: sMRI-analysis refers to the classification analysis when only baseline GM volume data were considered. The diagnostic performance was evaluated by means of sensitivity

(Sens), specificity (Spec), balanced accuracy (BAC), false positive rate (FPR) and positive/negative predictive value (PPV/NPV).

Only 13 out of 40 independent HR[symp]test subjects were wrongly assigned to the HR[ill] group (specificity 67.5%, FPR 32.5%). The diagnostic performance in the entire dataset (the 17 HR[ill] and 57 HR[symp]) attained a balanced accuracy of 77% (Table 5.3).

Table 5.4 List of the most discriminative regions for the HR[ill] vs HR[symp] contrast using sMRI data alone.

Lobe	Region/Hemisphere	**w**
Negative weights		
Cerebellum	Cerebellum_4_5_R	-0.8256
	Cerebellum_3_R	-0.4076
	Cerebellum_6_L	-0.299
	Vermis_10	-0.1135
Temporal	Temporal_Pole_Sup_R	-0.0298
	Temporal_Pole_Mid_L	-0.1987
	Temporal_Inf_R	-0.1093
	Fusiform_L	-0.2405
Frontal	Frontal_Inf_Oper_R	-0.4115
	Frontal_Mid_L	-0.3032
	Frontal_Sup_Medial_R	-0.1166
Parietal	Paracentral_Lobule_R	-0.6526
	Paracentral_Lobule_L	-0.5691
Occipital	Occipital_Sup_R	-0.4378
Limbic	Amygdala_R	-0.2295
	Cingulum_Mid_L	-0.5633
Basal ganglia	Putamen_L	-0.4387
	Pallidum_R	-0.2149
Positive weights		
Cerebellum	Cerebellum_4_5_L	0.2036
	Cerebellum_Crus1_R	0.1993
Temporal	Fusiform_R	0.6462
	Temporal_Inf_L	0.5056
Frontal	Frontal_Med_Orb_L	0.1159

	Precentral_R	0.2116
Parietal	Precuneus_R	0.3113
	Precuneus_L	0.1127
	Parietal_Inf_R	0.0191
Limbic	Hippocampus_R	0.6802
	Cingulum_Post_L	0.6145
Basal Ganglia	Putamen_R	0.3042
	Pallidum_L	0.2891

Inf, inferior; L, left hemisphere; Mid, middle; Med, medial; Orb, orbital; Oper, opercularis; Post, posterior; R, right hemisphere; Sup, superior; w, weight vector of corresponding features in the classification process.

Note: The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=HR[ill]; -1=HR[symp] group), a positive weight means the weighted average in that voxel was higher for the HR[ill] group, and a negative weight means the weighted average was higher for HR[symp] group.

5.4.2 SVM classification based on combination of sMRI and behavioural data

Subsequent inclusion of baseline RISC and RAVLT scores delivered higher diagnostic accuracy than before. The proposed method achieved 94% accuracy in predicting at baseline subsequent transition to schizophrenia in HR for familial reasons subjects (Table 5.5). All subjects of the HR[ill] group were correctly assigned to their group while 2 subjects of the HR[symp] group were wrongly classified as HR[ill] (sensitivity/specificity: 100%/88%; PPV/NPV: 89%/100%; permutation test p<=0.001). Only 8 out of 40 independent HR[symp]test subjects were wrongly classified as HR[ill] (specificity 80%, FPR 20%). The diagnostic performance in the entire dataset attained 91% balanced accuracy (Table 5.5).

Again without the RFE, the linear classifier gave a mediocre performance in distinguishing between HR[ill] and HR[symp] with a 56.5% accuracy (sensitivity/specificity: 65%/48%). Conversely, the RBF-SVM classifier applied to

the 17HR[ill] and 17HR[symp] individuals gave a higher predictive performance with the introduction of the RISC and RAVLT variables, with an accuracy of 73.5%, a sensitivity of 70.5% and a specificity of 76.5%.

Baseline RISC and RAVLT scores were selected by the SVM-RFE method, implying their contribution in the discrimination process. The neuroanatomical decision function consisted of a less extensive spatial pattern than before that involved, however several cortical and subcortical brain structures and the cerebellum (Figure 5.3B – Table 5.6). Specifically, the regions contributing more to the classification of HR[ill] subjects included the left superior temporal lobe extending to the left fusiform gyrus, the left thalamus, right superior and inferior frontal lobe regions, the insula bilaterally and parts of the cerebellum bilaterally, whereas regions with a higher weighted average for the HR[symp] group were the right putamen, right hippocampus and fusiform gyrus and parts of the cerebellum. Please see Table 5.6 for a list of the most discriminating brain regions that jointly formed the decision boundary of the classifier.

Table 5.5 Diagnostic performance of the classifier, using sMRI data combined with baseline bevavioral variables

sMRI-behavioural analysis	TP	TN	FP	FN	Sens (%)	Spec (%)	BAC (%)	FPR (%)	PPV (%)	NPV (%)
HR[ill] vs HR[symp]	17	15	2	0	100	88.2	94.1	11.7	89.4	100
HR[symp]test	0	32	8	0		80		20		100
Overall	17	47	10	0	100	82.5	91.2	17.5	62.9	100

Combined sMRI-behavioural refers to the results when a combination of baseline GM, RISC and RAVLT data was employed in the analysis. Again, diagnostic performance was evaluated by means of sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), false positive rate (FPR) and positive/negative predictive value (PPV/NPV).

Table 5.6 List of the most discriminative regions for the classification of HR[ill] vs HR[symp] in the combined analysis of baseline sMRI, RISC and RAVLT data

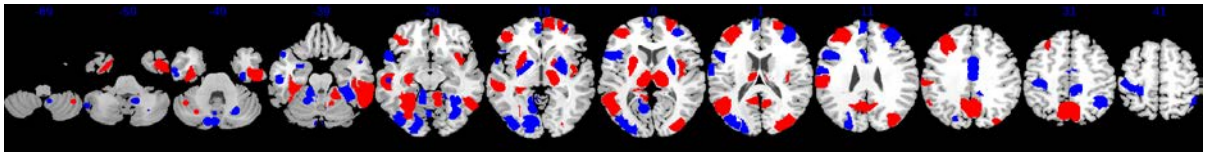
Lobe	Region/Hemisphere	**w**
Negative weights		
Cerebellum	Vermis_8	-0.5159
	Cerebellum_6_L	-0.3129
	Cerebellum_7b_R	-0.1477
Temporal	Temporal_Sup_R	-0.5579
	Temporal_Pole_Sup_L	-0.2867
	Fusiform_L	-0.4874
Frontal	Frontal_Inf_Tri_L	-0.5111
	Frontal_Inf_Orb_R	-0.3109
	Frontal_Sup_Medial_R	-0.2992
	Frontal_Sup_L	-0.2873
	Frontal_Mid_Orb_R	-0.1477
	Frontal_Inf_Oper_R	-0.0377
Parietal	Paracentral_Lobule_L	-0.3965
Occipital	Occipital_Mid_L	-0.262
	Cuneus_L	-0.2818
Limbic	Thalamus_L	-0.2449
perisylvian	Insula_R	-0.3669
	Insula_L	-0.2763
Positive weights		
Cerebellum	Cerebellum_Crus2_L	0.3704
	Cerebellum_4_5_L	0.2634
	Cerebellum_Crus1_R	0.263
Temporal	Fusiform_R	0.3337
Frontal	Frontal_Sup_Orb_L	0.3542
	Precentral_R	0.3201
Parietal	Parietal_Inf_R	0.0393
	Parietal_Inf_L	0.3
Occipital	Lingual_L	0.373
	Occipital_Sup_L	0.357
Limbic	Hippocampus_R	0.169
	Cingulum_Post_L	0.333
Basal ganglia	Putamen_R	0.356
	Pallidum_L	0.6429

Inf, inferior; L, left hemisphere; Mid, middle; Med, medial; Orb, orbital; Oper, opercularis; Post, posterior; R, right hemisphere; Sup, superior; w, weight vector of corresponding features in the classification process.

Note: The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial

representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=HR[ill]; -1=HR[symp] group), a positive weight means the weighted average in that voxel was higher for the HR[ill] group, and a negative weight means the weighted average was higher for HR[symp] group.

A



B

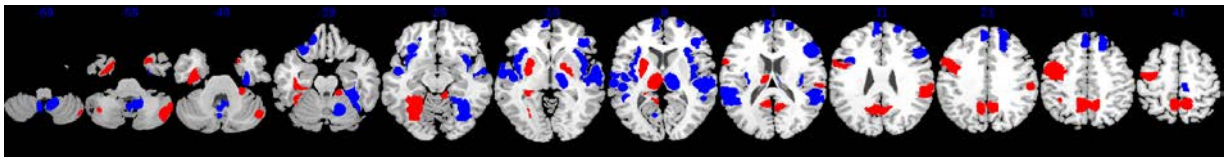


Figure 5.3 Discrimination maps for the classification of HR[ill] vs HR[symp]: a) just baseline MRI data were considered and b) baseline MRI were combined with RISC and RAVLT variables. The colours represent the weight of each feature in the classification function (the red scale represents positive weights and the blue scale represents negative weights). The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=HR[ill]; -1=HR[symp] group), a positive weight means the weighted average in that region was higher for the HR[ill] group, and a negative weight means the weighted average was higher for HR[symp] group. Therefore, the discrimination map should not be interpreted as a standard statistical parametric map resulting from a mass-univariate statistical test to find group differences, and no local inferences should be made based on the SVM weights.**Note**: features correspond to GM volume measures in the AAL-defined brain regions, and not voxels.

5.4.3. Voxel-based morphometry

For comparative purposes, I conducted a two-sample T-test in SPM5, contrasting the 17 subjects of the HR[ill] and the 17 subjects in the HR[symp]. Group differences were assessed using the General Linear Model (GLM) on the smoothed normalized GM segments, examining results at different statistical thresholds. No significant volumetric between-group differences in GM were observed when conventional VBM analysis was employed using a family error rate (FWE) of $p<0.05$. Even at a *p*-value uncorrected for multiple comparisons of $p<0.001$, very few differences in very small clusters could be detected using VBM. Lowering the VBM threshold to $p<0.05$ (uncorrected) resulted in significant differences in voxels covering the right fusiform, left cerebellum, right inferior and left superior parietal lobe, right supramarginal gyrus, left superior temporal and frontal lobe, the left anterior cingulate and the right precentral gyrus. The output maps for the VBM analysis are presented in Figure 5.4. Overall, the regions identified in the VBM analysis were similar to regions identified by our classification method.

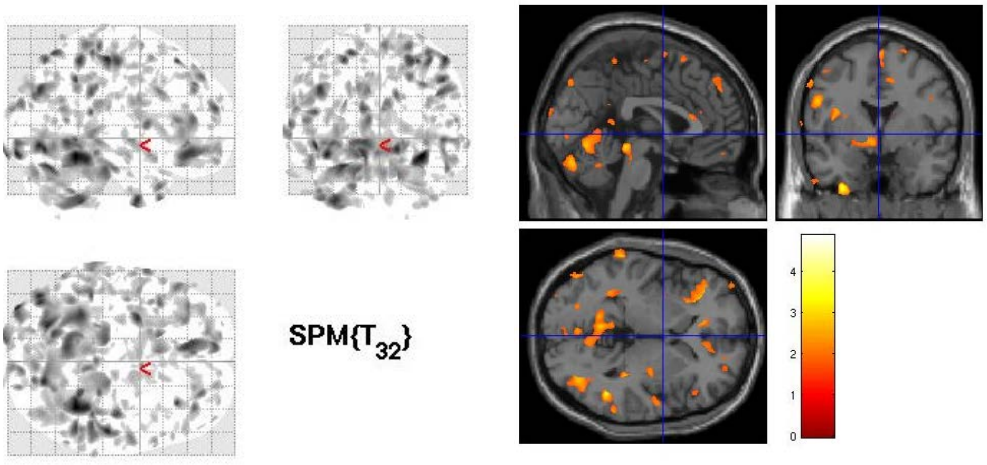


Figure 5.4 Results of the conventional VBM analysis for the smoothed and normalized GM segments (*p*-value<0.05, uncorrected).

5.4.4 Additional analyses

For purposes of clarity and thoroughness, classification results for other analyses that were conducted are reported in this section. Initially, I have ran the nested LOO-CV SVM-RFE method without performing any sampling and by using the total group of 57 HR[symp] subjects against the 17 HR[ill] subjects. A classification bias towards the majority class (i.e. HR[symp] group) was observed, in that the classifier assigned every subject as HR[symp], resulting in very good specificity of 92% but zero sensitivity (accuracy: 70.3%), both when MRI data were used alone and when they were combined with baseline RISC and RAVLT variables.

The class imbalance issue arises when there are more data examples in one class and fewer occurrences in the other class and it is a problem widely known in the machine learning community (Japkowicz 2000, Japkowicz & Stephen 2002, He & Garcia 2009) and more recently acknowledged in imaging-based studies that employ machine learning methods (Dubey et al. 2014, Cuingnet et al. 2011, Yuan et al., 2012). Class imbalance problems can be addressed in two ways (see Japkowicz & Stephen, 2002 for a review on this matter).

One way is to adopt a model-based approach where the classifier is assigned different misclassification costs for each class. In soft-margin SVM, there is a C penalty that controls the trade-off between maximizing the margin and reducing misclassification error. By assigning a higher misclassification cost for the minority class instances than the majority class instances, the effect of class imbalance could be reduced. However, in the EHRS dataset, I did not observe any change in the behaviour of the classifier and again our model ended up with very good specificity (90-95%) and zero sensitivity. Other studies have also reported that this so-called reweighting (or cost sensitive learning) approach did not produce satisfactory results (Duchesnay et al. 2011,[1]).

1 http://stats.stackexchange.com/questions/94295/svm-for-unblanced-data

Using a data-driven approach is another way to deal with class imbalance issues. That is, one can either choose to oversample the minority class or undersample the majority class (He & Garcia 2009, van Hulse et al. 2007) or do both (Akbani et al. 2004). Oversampling techniques randomly replicate (with or without replacement) existing (training) data from the minority class until a class balance is reached (Japkowicz 2000, Japkowicz and Stephen 2002). There are also oversampling techniques that generate new, synthetic data by randomly interpolating pairs of nearest-neighbour data instances (Chawla et al. 2002).

The alternative to oversampling is undersampling. The simplest version is the random undersampling technique that removes data samples from the majority class at random until the classes are balanced. Cluster-based sampling techniques aim to group together data instances in the majority class into a number of clusters that is equal to the size of the minority class (Yen and Lee 2006). Here, I chose to perform random undersampling because it was the most straightforward technique to alleviate the class imbalance and also because other authors and imaging labs working with data in schizophrenia have done this before (Koutsouleris et al. 2009b, Koutsouleris et al. 2010, Koutsouleris et al. 2015b, Mourao-Miranda et al. 2012a).

Therefore, an equal number of HR[symp] subjects was selected to match the 17 individuals in the HR[ill] group and the rest of the HR[symp] subjects were used to further validate the classifier.

Additionally, in order to alleviate the difference in RISC scores across the equally-sized groups, I have repeatedly and randomly downsampled the original HR[symp] group that consisted of 57 subjects to groups of 17 subjects in order to match the HR[ill] group and repeated the entire SVM-RFE procedure 100 times in order to create a distribution of accuracies. In the case where only the neuroanatomical imformation in the AAL-defined brain regions were considered, the method produced an average accuracy of 98.1% (sensitivity/specificity: 96.9%/99.4%, p-value= 0.09) whereas when baseline neuroanatomical information was combined with baseline RISC and RAVLT variables the mean accuracy of the method across all 100 runs was 98.7% (sensitivity/specificity: 98.4%/99.1%, p=value=0.1).

5.5 Discussion

To the best of the author's knowledge, this is the first study to evaluate the feasibility of an individualised prediction of psychosis in a cohort of subjects at HR for familial reasons, by combining baseline neuroanatomical data with baseline schizotypal and neurocognitive features using a multivariate pattern recognition technique. The high accuracy of our classification method was obtained by leave-one-out cross-validation which provides an almost unbiased estimate of the generalizability. The diagnostic performance of our classifier was further validated by reliably classifying the $HR[symp]_{test}$ subjects.

The neuroanatomical pattern associated with transition to schizophrenia was quite extensive and was primarily associated with grey matter abnormalities covering the frontal, orbito-frontal and occipital lobe regions bilaterally as well as parts of the superior and medial temporal lobe regions, the left inferior parietal lobe and parts of the cerebellum, in keeping with previous VBM studies that reported similar baseline neuroanatomical reductions in converters versus non-converters located in inferior frontal and superior temporal brain regions (Fusar-Poli et al. 2011), the hippocampus, cingulate cortex and the cerebellum (Smieskova et al. 2010, Pantelis et al. 2003, Job et al. 2006).

A conventional VBM analysis failed to detect any significant between-group differences after performing corrections for multiple comparisons. At a lower VBM threshold of $p<0.05$, uncorrected for multiple comparisons, however, differences in brain regions similar to those detected by the SVM classifier were found.

However, as stated before, VBM analyses consider each voxel as a spatially independent unit and cannot provide predictive value at a single-subject level. On the contrary, SVM is a multivariate technique that examines voxels jointly and considers inter-regional correlations. Therefore, individual brain regions may display high discriminatory power either because there is a large difference in volume between the study groups in that region or this region is highly inter-correlated with regions in a spatially distributed network of brain regions. For this reason, the neuroanatomical

maps derived by SVM should be interpreted as a spatially distributed pattern where all its constituent parts (i.e. brain regions) contribute to the classification rather than making assumptions on effects locally.

The neuroanatomical pattern that distinguished the HR[ill] from the HR[symp] individuals was similar in both analyses, i.e. when only baseline MRI data were considered and when those were combined with baseline RISC and RAVLT measures (see Figure 5.2). The neuroanatomical pattern in both cases consisted of a deficit network, consisting of brain regions that were relatively reduced in HR[ill] compared to the HR[symp] subjects and an excess network, including brain regions of relative increased volume in the HR[ill] compared to the HR[symp] group. Roughly, the neuroanatomical pattern consisted of regions showing relative GM volume reductions in the left fusiform gyrus, the right orbitofrontal and superior medial frontal lobe, left fronto-parietal lobe, right superior temporal lobe and parts of the cerebellum in keeping with other studies in this area (Koutsouleris et al. 2009b, Koutsouleris et al. 2012a).

Interestingly, GM volume increases were observed in HR[ill] subjects compared to the HR[symp] in the orbital frontal lobe, left temporal and inferior parietal lobe regions and the parts of the cerebellum, in line with previous studies (Borgwardt et al. 2007b, Koutsouleris et al. 2015b). Similar networks of relative increases were also reported in other SVM-based studies in high-risk or first-episode cohorts (Koutsouleris et al. 2009b, Koutsouleris et al. 2012a, Mourao-Miranda et al. 2012a). While, this may be counter-intuitive in that the aim is to predict schizophrenia, a condition characterised by generalised and multifocal reductions in GM volume, it partly represents the common use of linear multivariate modelling techniques to produce weights of spatially distributed relative differences which reflect relative changes in index regions and their networks. In addition, however, there is evidence that large brains and constituent parts may predict schizophrenia in populations at high risk (Fusar-Poli et al. 2011).

Both baseline RISC and RAVLT scores were significant in discriminating between the study groups, in keeping with previous studies both in the EHRS literature (Johnstone et al. 2005) and elsewhere (Riecher-Rossler et al. 2006, Lencz et al. 2006,

Koutsouleris et al. 2012b) that have shown that neurocognitive variables are good predictors of transition to schizophrenia and could possibly enhance the predictive power of multivariate models (Lencz et al. 2006, Riecher-Rossler et al. 2009).

Our classification performance using structural MRI data alone is also comparable to findings of previous studies that applied neuroanatomical-based SVM to predict transition in clinical HR cohorts (Koutsouleris et al. 2009b, Koutsouleris et al. 2012a). Koutsouleris and collegues (2009b) built a SVM classifier upon structural MRI data of individuals in early and late at-risk mental state of psychosis subjects and a group of matched HC and evaluated its performance by distinguishing MRI data derived from baseline scans of ARMS subjects who developed schizophrenia (ARMS-T), those who did not (ARMS-NT) and a second matched group of HC. All three-group and pairwise classifiers achieved classification performance above 80%. In the most critical in terms of clinical utility, their ARMS-T vs ARMS-NT classifier achieved an accuracy of 82% (sensitivity, specificity: 83%, 80%), whereas in a follow-up study, the same group (Koutsouleris et al. 2012a) validated their previous analysis by classifying an independent cohort of HC, ARMS-T and ARMS-NT and achieved an improved accuracy in the ARMS-T vs ARMS-NT pairwise classifier (84%). Both studies described ARMS as help-seeking subjects, at imminent risk of psychosis on the basis that they exhibited prodromal symptoms. In contrast, our data sample was acquired from a neuroleptic-naive cohort of subjects at familial HR for the disorder that also presented significantly lower transition rates (nearly 13%) than the ARMS cohorts with transition rates of up to 43% (Koutsouleris et al. 2012a). Moreover, the study presented here is different from the other studies in the field in that it consists of a homogenous sample of individuas who have familial risk rather than samples of subjects with mixed clinical and familial factors. Finally, in this study the aim is to predict later diagnosis of schizophrenia rather than a single psychotic episode as it is the case in clinical-based HR studies.

Certain limitations have to be considered in this study. Firstly, the sample size is modest and therefore our findings should be interpreted with caution. The modest sample size can be, however, explained by the already known difficulties in recruiting subjects in HR studies and the relatively low overall conversion rate (13%)

of our familial HR sample. Therefore, replication of our findings using a larger independent cohort is required in order to further validate our learning model. In addition, it is not clear whether this predictive performance could generalise to other HR cohorts. The HR group studied here was recruited on the basis of familial history, thus representing a genetic risk for the disorder, but also had psychotic symptoms. It is, thus, unclear whether our findings could generalise to HR cohorts, presented to clinical services with psychotic symptoms or other disturbances.

Another limitation pertaining to the generalizability of our results in the context of machine learning methods in general, and of SVM in particular here is the choice of cross-validation. While the leave-one-out cross-validation employed in this study has a low bias, it suffers, however, from high variance. More specifically, the LOO cross-validation underestimates the variance, due to the repeated use of subsets of the data in most training sets. Of course this is an issue for most cross-validation schemes, although other validation techniques such as splitting the original data in half, for training and testing purposes, or choosing a 10-fold cross-validation might produce better estimates of the actual performance of the classifier. However, in light of the small dataset here I reckoned that the LOO-CV would be the most suited choice for the evaluation of the classifier's performance.

Finally, I acknowledge the limitation for having performed a random sampling of the data set. From the original group of 57 HR[symp] subjects, I randomly selected 17 individuals to match and contrast the 17 HR[ill] subjects. I acknowledge that in case of undersampling, removing examples (data instances) from the majority class might cause the classifier to miss out important clues and characteristics pertaining to the majority class. However, studies in this field have shown that class imbalance issues can significantly impact the classifier's performance (Dubey et al. 2014, Duchesnay et al. 2011) and highlighted the importance of having balanced groups of data (Wei & Dunbrack 2013). In particular, Wei and Dunbrack (2013) highlighted the importance of training the classifier using balanced training sets, regardless of having a balanced representation of the groups in the test set. Two other neuroimaging studies in schizophrenia that employ SVM have used similar sampling approaches (Koutsouleris et al. 2015b, Mourao-Miranda et al. 2012a). Specifically, in

Koutsouleris and colleagues (2015b), 33 out of 40 ARMS-NT individuals were selected to contrast the 33 ARMS-T subjects, and the rest 7 ARMS-NT were used to further validate their classifier. In Mourao-Miranda and colleagues (2012a), 91 HC individuals were initially recruited, but in the SVM analysis only 28 of them were selected to contrast the 28 individuals in the continuous and episodic first-episode groups.

5.6 Conclusions

We examined the diagnostic performance of a SVM classifier in predicting transition to schizophrenia using baseline MRI scans and also reported the neuroanatomical pattern that differentiated the HR subjects that developed schizophrenia from a matched group of HR subjects that did not develop the disorder but manifested psychotic symptoms. Additionally, it was shown that the integration of neuroanatomical data with measures of neurocognitive functioning and schizotypal cognition can not only improve predictive performance but can indicate which features contribute to that prediction and hence are in some way discriminative of a pattern that predicts a later diagnosis of schizophrenia. To date, there are few schizophrenia studies that have examined the diagnostic performance of combining data from various sources into a unified learning model. Taking into consideration that clinical, neurocognitive and neuroimaging assessments can individually describe a different aspect of pathology in schizophrenia, I believe that integration of these variables into a single learning framework might provide a clearer view of the patient's status and thus, a stronger insight into disease development and progression.

CHAPTER 6

Individualized prediction of psychosis in subjects with an at-risk mental state

6.1 Introduction

As discussed in Chapter 1, the onset of schizophrenia is usually preceded by a prodromal phase, which is often characterized by sub-threshold psychotic symptoms and a progressive decline in functioning. Individuals presenting with these characteristics are considered at an increased risk for developing schizophrenia and other psychosis-related disorders for clinical reasons and are identified as ultra-high risk (UHR) or at an at-risk mental state (ARMS) for psychosis.

The aim of this chapter is to examine whether our previous findings can be replicated by evaluating the diagnostic performance of our MRI-based classifier in predicting disease conversion in subjects with an ARMS that were drawn from the FePsy (Fruherkennung von Psychosen) study.

6.2 Background

Over the past 20 years, a focus on the early intervention of psychotic disorders has emerged. Initially early intervention strategies were aimed at helping individuals in their first episodes of psychosis (Perkins et al. 2005, Marshall et al. 2005) while over the last few years, early recognition and intervention has moved towards individuals exhibiting prodromal signs of the disorder. Early detection and intervention centers, such as the Personal Assessment and Crisis Evaluation (PACE) in Australia (Yung et al. 1996), the Prevention through Risk Identification, Management and Education (PRIME) in the US (McGlashan et al. 2003), and the Outreach and Support in South London (OASIS) clinic (Fusar-Poli et al. 2013b), have been set up worldwide (Edwards et al. 2005) with the aim of providing case management and provisional treatment for individuals presenting with subthreshold psychotic symptoms.

Converging evidence suggests that early intervention in psychosis may substantially

ameliorate response to treatment and outcome for the illness (McGorry 2002a), and delay (McGorry et al. 2002b) or even prevent disease onset (Amminger et al. 2010, Phillips et al. 2007). However the question of whether and at what stage early intervention is indicated is still a matter of much debate (Riecher-Rossler et al. 2006). A number of ethical considerations and implications arise from treating patients early (Candilis 2003), including the possibility of treating *false positive* cases (people who will not later develop psychosis but are falsely labelled as such), which may result in unnecessary distress and exposure to putative side effects of medication, not to mention the associated stigma for the individuals and their families. The best approach would, therefore, lie in the establishment of an accurate risk assessment system that would quantify the risk of psychosis while reducing the rate of false positives.

As has been discussed before, schizophrenia and other psychosis-related disorders may begin many years before the emergence of frank, clear symptoms, with nonspecific changes, perceptual alterations and often sub-threshold psychotic disturbances. Although individuals that exhibit prodromal symptoms have an increased risk of developing psychosis, not all individuals do so. Thus in order to delineate the prodrome of psychosis from a state of heightened risk the ultra-high-risk (UHR) or the at-risk mental state (ARMS) or clinical high-risk (HR) paradigms have been introduced.

Operationalized criteria for identifying subjects with an ARMS or at clinical HR have been developed. These criteria are based on a combination of trait and state risk factors inferred as increasing psychosis risk and have been used to categorize individuals in the following three subgroups based on the criteria identified by the PACE clinic (Yung et al. 2008, Yung et al. 1998). The first two subgroups specify state risk factors and are defined by the presence of either transient psychotic symptoms, called brief limited intermittent psychotic symptoms (BLIPS) or sub-threshold, attenuated psychotic symptoms (APS). The third subgroup involves trait and state risk factors that are operationally defined by a significant reduction in functioning plus either an affected relative with psychosis or a pre-existing schizotypal personality disorder. Within the concept of *basic symptoms* as defined by

the Bonn group, additional prodromal criteria have also been identified that describe a set of subtle and subjective changes in affect, thinking, emotional and cognitive processing and perceptual disturbances of the self and the world (Huber and Gross 1989, Schultze-Lutter 2009). It should be noted that individuals defined with an ARMS or at clinical HR are all help seeking which effectively means this is part of the criteria.

With regards transition rates to psychosis, significant variation has been observed across sites studying UHR samples, with recent studies reporting a decline in the risk of transition (Yung et al. 2007). Transition rates are an important tool for evaluating the predictive validity of the operationalized clinical HR criteria and thus certifying the need for preventative intervention. A recent meta-analysis of studies employing clinical HR samples estimated that 18% of HR individuals develop psychosis after 6 months of follow-up, reaching 36% after 3 years (Fusar-Poli et al. 2012b) with the age of participants, any received treatment and the diagnostic criteria used being the most influential moderators of transition risk.

Valid and reliable prognostic markers are, thus, needed in order to improve prediction of conversion rates and reduce false positive rates. A series of putative biomarkers have been recently identified, suggesting that the at-risk mental state is characterized by abnormalities in the neurocognitive domain (Lencz et al. 2006, Fusar-Poli et al. 2012a, Koutsouleris et al. 2012b) and alterations at the neuroanatomical (Fusar-Poli et al. 2011, Smieskova et al. 2010) and neurofunctional level (Fusar-Poli et al. 2007). As it has been discussed in previous chapters, conversion to psychosis in ARMS subjects has been associated with reduced grey matter volume in the prefrontal and temporal cortices and other subcortical brain structures (Smieskova et al. 2010).

Recently, multivariate pattern recognition approaches, including SVM methodology, have provided important leads towards the translation of neuroimaging findings into clinical practice, by taking into account inter-regional correlations between brain regions and working at the single-subject level (Orru et al. 2012). These methods may thus provide the means for an individualized risk assessment and prediction of psychosis conversion and possibly deliver increased sensitivity and specificity, both

of which are essential for informing individualized prevention care.

Previous machine learning studies have shown that a neuroanatomical-based prediction of psychosis is possible at the single-subject level (Koutsouleris et al. 2009b), providing diagnostic accuracy up to 84% (Koutsouleris et al. 2012a). The high diagnostic performance of our MRI-based SVM classifier presented in Chapter 5 brings additional evidence to the feasibility of an individualized prediction of psychosis. Here, I aimed to examine whether our SVM approach for the prediction of psychosis conversion could be replicated in a second, independent cohort of subjects at high clinical risk for developing the disorder.

6.3 Materials and Methods

6.3.1 Subjects

The subjects included in this analysis were part of a large prospective, early psychosis study; the FePsy (Fruherkennung von Psychosen) study. Details regarding recruitment and screening procedures have been described in Chapter 4, section 4.1.2 and can also been found in previous studies (Riecher-Rossler et al. 2006, Riecher-Rossler et al. 2007). All aspects of the study were approved by the Ethics Committee of Basel, Switzerland, and written informed consent was obtained for each participant before study inclusion.

Briefly, 37 subjects were identified as having an at-risk mental state (ARMS) for psychosis using a screening procedure based on the Basel Screening Instrument for Psychosis (BSIP), the Brief Psychiatric Rating Scale (BPRS- see Table 6.1) and the Scale for the Assessment of Negative Symptoms (SANS). These assessments were used in order to elicit psychopathology and rate (pre-) psychotic and negative symptoms.

The BPRS consists of 24 items and is one of the most frequently used research instruments for evaluating psychopathological symptoms in patients with

schizophrenia (Velligan et al. 2005). Although there is no widely accepted factorial structure of the BPRS, several authors have proposed a four factorial structure. In the FePsy bibliography, the BPRS total score was used along with four subscales (i.e., Depression/Anxiety, Psychosis/Thought Disturbance, Negative Symptoms, and Activation) derived from the factorial structure of Velligan et al. 2005. (See Appendix IV, Table 1 and Figure 1 for a description of the BPRS subscores and the 4-factor analysis based on Velligan et al. 2005).

Additionally, the SANS assessment consists of 19 items, which are grouped into five domains or factors (Affective flattening, Alogia, Avolition-apathy, Anhedonia-Asociality, and Inattention). In the present study, the SANS global score and the five original subscales were used.

Matched groups of healthy controls and first-episode patients were recruited as well. In short, 22 healthy controls (HC) with no history of any psychiatric disorder and 25 first-episode (FE) individuals, who met operationalized criteria for first-episode psychosis as described in Yung et al. 1998, were recruited. The ARMS, FE and HC individuals did not differ significantly with respect to sex, ethnicity, handedness and current and previous alcohol use (Riecher-Rossler et al. 2007).

Individuals were followed up at monthly intervals during the first year, at 3-month intervals during the second and the third year and annually thereafter until transition to psychosis was established or until the end of the follow-up period (in 2007). In general, all ARMS subjects were followed up for over 4 years during which they were also offered supportive counseling and clinical management.

Transition to psychosis was operationally defined by meeting criteria described in Yung et al. 1998 (Table 6.1) and further determined by a diagnostic interview using the ICD-10 criteria at the time of transition. Follow-up information for 2 ARMS subjects was not available. In this regard, 16 of the 35 ARMS individuals with retained follow-up information made a transition to psychosis (denoted as ARMS-T subjects) and 19 did not convert (ARMS-NT). Seven out of the 35 ARMS participants have received low doses of antipsychotic medication, some time prior to study inclusion (2 participants on olanzapine, 2 chlorprothixene and 3 risperidone),

all for less than 3 weeks.

Table 6.1 ARMS inclusion and transition to psychosis criteria.

ARMS Inclusion criteria

Inclusion into the study was based on the BSIP checklist and required one or more of the following:

1. Attenuated psychotic-like symptoms: at least several times a week and for more than 1 week duration (a score of 2 or 3 on the Brief Psychiatric Rating Scale (BPRS)
2. Brief limited intermittent psychotic symptoms (BLIPS): scores of 4 or above on the hallucination item or 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items of the BPRS, with each symptom lasting less than 1 week before resolving spontaneously
3. Genetic risk: a first or second-degree relative with a psychotic disorder plus at least 2 further risk factors for or indicators of beginning psychosis according to the BSIP screening instrument.

Criteria for transition to psychosis

1. BPRS scores of 4 or above on the hallucination item or scores of 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items
2. Symptoms had to occur daily and persist for more than 1 week.

BSIP, Basel Screening Instrument for Psychosis; BPRS, Brief Psychiatric Rating Scale

Here, I am mainly interested in contrasting the ARMS group that converted to psychosis (ARMS-T) against the ARMS group that did not (ARMS-NT; Table 6.2).

Table 6.2 Socio-demographic and clinical information of the 2 study groups

	Study Groups		
	ARMS-T	ARMS-NT	*P*
Socio-demographic variables			
N	16	19	
Mean age at baseline, y (sd)	26.8 (6.5)	23.9 (6.2)	ns[a]
Sex (male), n (%)	11 (69)	10 (52)	ns[b]
Educational level			ns[b]
<9 y, n (%)	3	8	
9-11 y, n (%)	7	7	
12-13 y, n (%)	5	2	
<13 y, n (%)	1	2	
Mean verbal IQ (MWT-B) (sd)	109.6 (12.6)	107.3(15.4)	ns[a]
Cannabis use at baseline			ns[b]
none	10	11	
rarely	1	1	
Several times/month	0	2	
Several times/week	4	0	
daily	1	5	
Antipsychotics before entry, n (%)	6 (37.5)	1 (5)	<0.05[b]
Antidepressants at baseline, n (%)	7 (44)	5 (26)	ns[b]
Family History			ns[b]
No relative	15	16	
One 1rst degree	1	2	
One 2nd degree	0	1	
Clinical variables			
Mean BPRS total score at baseline (sd)	42.3(10.6)	35.7 (7.1)	ns[c]
Mean SANS global score at baseline (sd)	9.75(5.8)	7.7(4.2)	ns[c]
Mean interval between MRI and disease onset, d (sd)	306.3 (318.3)	na	

ARMS-T: at-risk mental state individuals that later developed psychosis; ARMS-NT: at-risk mental state subjects that did not make a transition. BPRS: the Brief Psychiatric Rating Scale; SANS: the Scale for the Assessment of Negative Symptoms. Verbal IQ Mehrfach-Wortshatztest-B

[a] Student's *T*-test. [b] Fisher's exact test. [c] Mann-Whitney U-test

6.3.2 Image Acquisition and Preprocessing

Subjects were scanned using a Siemens (Erlangen, Germany) Magnetom Vision 1.5 T scanner at the University Hospital Basel. Head movement was minimized by foam padding and velcro straps across the forehead and chin. A three-dimensional volumetric spoiled gradient recalled echo sequence generated 176 contiguous, 1 mm thick sagittal slices. Imaging parameters were: time-to-echo, 4 msec; time-to-repetition, 9.7 msec; flip angle, 12; matrix size, 200 x 256; field of view, 25.6 x 25.6 cm matrix; voxel dimensions, 1.28 x 1 x 1 mm.

Before preprocessing, all structural MRI scans were converted from a sagittal orientation (i.e. slices were recorded from right to left) to an axial orientation (Figure 6.1), using the FSL (FMRIB Software Library; Jenkinson et al. 2012) program.

A

B

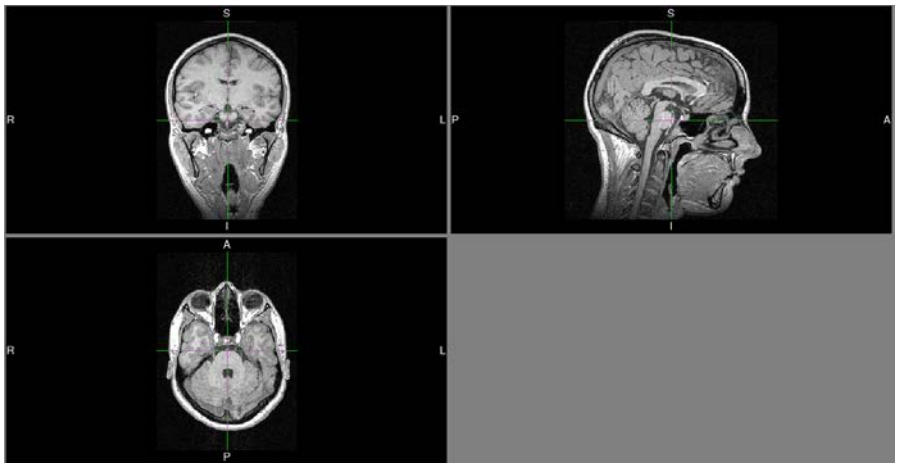


Figure 6.1 FSL view of the orientation of MRI scans. (a) Initially MRI scans had a sagittal orientation (b) and then an axial orientation.

Study-specific templates and customized prior probability maps were constructed using data from all subjects in the Basel study (i.e. 35 ARMS, 22 HC and 25 FE). A detailed description of their creation can be found in Chapter 4, section 4.2.2.

After inspection for gross abnormalities and artefacts, the baseline MRI scans entered a pre-processing pipeline in SPM5 (Wellcome department of Cognitive Neurology, London, UK). A detailed description of the steps followed can be found in Chapter 4, section 4.2.

6.3.3 Multivariate Pattern Classification Analysis

6.3.3.1 Support Vector Machine

As described in detail in previous chapters, a linear SVM classifier was implemented for the analysis of baseline neuroanatomical data derived from the FePsy study. The steps in the pattern classification analysis were exactly the same as the ones described in Chapter 5, section 5.3.4.

6.3.3.2 Feature Extraction

This step is identical to the one described in Chapter 5 section 5.3.4.2. The result of this step is a 116-length vector consisting of GM density values. Again features were scaled to [0 1] before entering as input into the linear classifier.

6.3.3.3 Feature Selection

The rationale and implementation of the feature selection step is identical to what was presented and described in Chapters 5 and 4. Briefly, a recursive feature elimination (RFE) technique was implemented, which was embedded in a nested leave-one-out cross-validation (LOO-CV) in order to increase diagnostic performance of the classifier. A graphical representation of the SVM-RFE and the nested LOO-CV adjusted for the ARMS groups of the FePsy study is given below (Figure 6.2).

A discrimination map was again generated based on the weight coefficients of the features that were selected by the RFE method (Figure 6.3). The discrimination map consists of brain regions that according to the RFE methodology are the most distinctive in the classification task and provides a spatial representation of the decision function in that every feature contributes with a certain weight to this function (or hyperplane). The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=ARMS-T; -1=ARMS-NT group), a positive weight means the weighted average in that region was higher for the ARMS-T group, and a negative weight means the weighted average was higher for ARMS-NT group. Since the SVM classifier is multivariate by nature, it should be noted that all brain regions constituting the decision function contribute to the classification.

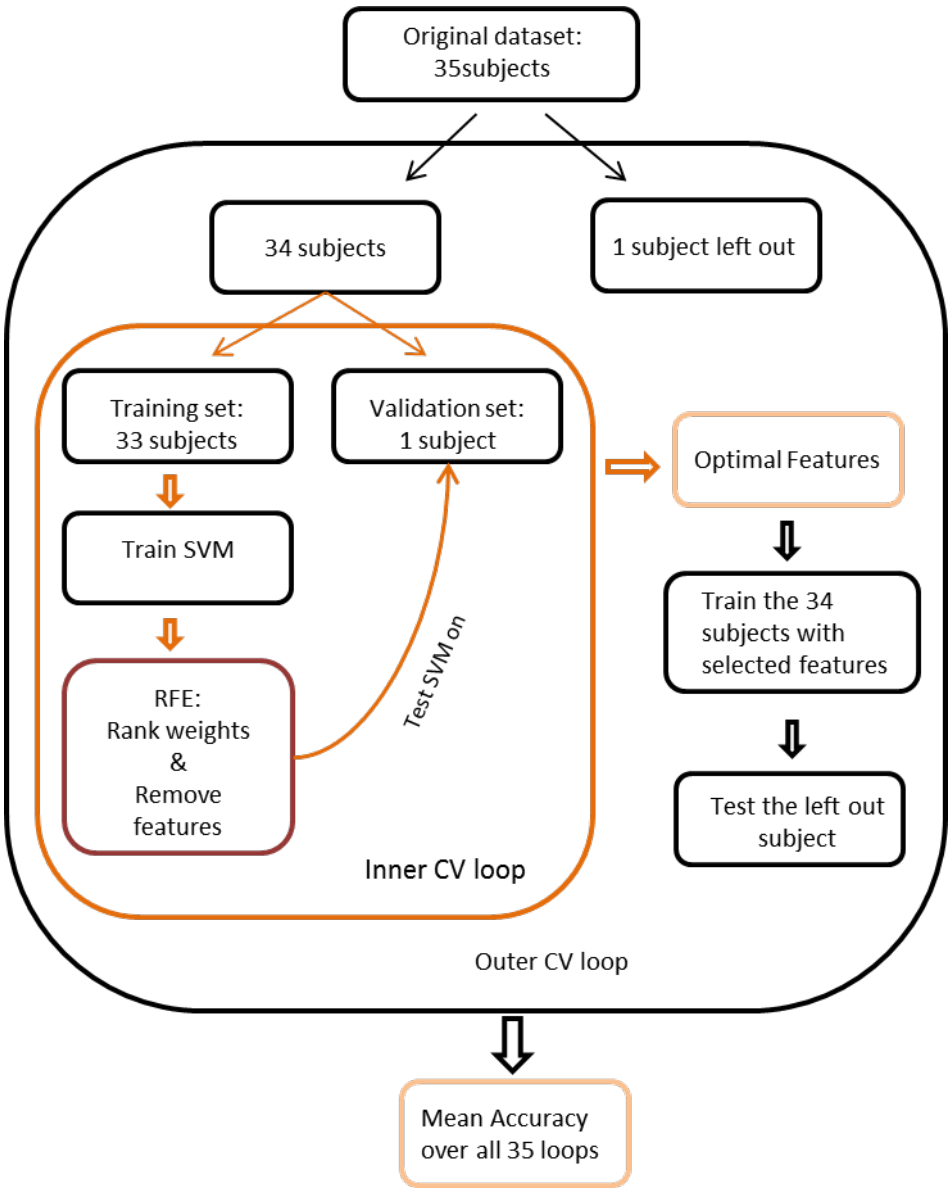


Figure 6.2 Representation of the nested LOO-CV SMV-RFE method. We employed a nested LOO-CV where I repeatedly excluded one subject to comprise the testing set and the remaining subjects were again repeatedly repartitioned in an internal validation loop where one subject was left out for validation and the rest formed the internal training group. In this loop, RFE was repeatedly performed and the mean accuracy on the validation group at each elimination level was recorded until all features were removed. The feature set that produced the maximum accuracy on the validation set was selected and applied to the testing set of the outer testing loop. Finally, mean accuracy was calculated across all outer CV loops.

6.3.3.4 Permutation testing

Again, permutation testing was performed in order to derive a *p* value for the accuracy of our classifier. A detailed description of the permutation testing was given in Chapter 4, section 4.4.5.

6.4 Results

6.4.1. Socio-demographic and clinical findings

The rate of conversion to psychosis was 45.7 % in this ARMS sample of 35 subjects. The mean interval between the baseline scan and disease conversion scan was 306 days (median: 263, range: 25–1137 days). There were no significant differences between converters and non-converters to psychosis with regards to age, gender, educational level, verbal IQ, cannabis use at study entry, baseline global BPRS and SANS scores (Table 6.2). There were also no significant differences on any BPRS and SANS subscales between converters and non-converters, with the exception of the SANS subscale of Asociality/Anhedonia that was significantly different between the two groups (Mann-Whitney U-test= 92.5, *p*-value= 0.048), in keeping with previous findings in the FePsy study literature (Riecher-Rossler et al. 2007), where among the individuals who later developed psychosis, anhedonia and asociality were more frequent than in individuals who did not convert. Additionally, the two groups significantly differed in terms of antipsychotic medication before study entry, with 6 subjects that later developed psychosis (ARMS-T) and 1 subject later categorized as AMRS-NT having been taking neuroleptics some time before study inclusion (Table 6.2). In fact, it could be that the difference in the SANS Asociality/Anhedonia scale reflects the higher use of antipsychotic medication in the ARMS-T group rather than any difference in psychopathology or symptom severity, with the consequent invalidation of the claim that this subscale can serve as a predictor of subsequent transition to psychosis since it might be not an aspect in the disease course itself but

merely a byproduct of medication. Previous studies have specifically linked the clinical ratings of anhedonia to antipsychotic medication through dopamine antagonism (Gard et al. 2008, Erhart et al. 2006).

6.4.2 SVM classification analysis

The application of our LOO-CV SVM-RFE methodology to baseline structural MRI data of the ARMS groups achieved 74% accuracy in predicting conversion to psychosis (Table 6.3). Six out of 16 subjects in the ARMS-T group were wrongly classified as ARMS-NT, while only 3 out of 19 subjects in the ARMS-NT group were incorrectly labeled as ARMS-T (sensitivity/specificity: 63%/84%; PPV/NPV: 77%/73%; permutation test p=0.002).

The likelihood ratio of a positive test result was LR+= 3.95 (Table 6.3), meaning that a positive prognostic test in a given ARMS subject would increase the probability of a subsequent transition to psychosis from 45.7% to 77% (posttest probability=posttest odds/posttest odds+1, posttest odds=pretest odds*LR+).

The misclassified ARMS-NT subjects did not significantly differ from the correctly classified ARMS-NT in any of the socio-demographic or clinical variables (Table 6.4). On the contrary, the misclassified ARMS-T subjects were significantly different from the correctly classified ARMS-T individuals in terms of gender distribution and use of antipsychotic medication before study entry (Table 6.4). This may partly explain the lower sensitivity of the SVM-RFE method, since the ARMS-T group consisted of a more inhomogenous group of individuals with regards to anti-psychotic medication, which in turn might have hindered the identification of a common neuroanatomical signature across subjects in this group. The effect of antipsychotic medication in brain structure is widely acknowledged by the scientific community (Smieskova et al. 2009, Navari and Dazzan 2009), and might have played a major role in the classification of the ARMS subjects here, despite the fact the exposure was before study entry and relatively brief.

In addition, while no other significant differences with regards to the rest of the

demographic and clinical variables (SANS global score and its subscales, BPRS total score and subscales of Psychosis/Thought Disturbance, Negative Symptoms, and Activation) were observed, the scores in the Depression/Anxiety subscale of the BPRS scale were significantly different between the correctly and misclassified ARMS-T subjects (*p*-value <0.05, Mann-Whitney U-test), again possibly suggesting the existence of a cluster of individuals within the ARMS-T group that exhibited more severe psychotic symptoms which may have warranted the use of anti-psychotic medication, which in turn may have impacted on brain structure.

Table 6.3 Classification performance

	TP	TN	FP	FN	Sens (%)	Spec (%)	BAC (%)	FPR (%)	PPV (%)	NPV (%)	LR+/LR-
ARMS-T vs ARMS-NT	10	16	3	6	62.5	84.2	74.2	15.7	77	73	3.9/0.45

The diagnostic performance was evaluated by means of sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), false positive rate (FPR) and positive/negative predictive value (PPV/NPV). LR+ was also calculated as sensitivity/1-specificity and LR- = 1-sensitivity/specificity.

The spatially distributed network that discriminated between the two groups was quite extensive and consisted of GM abnormalities in a spatially distributed network covering all four lobes and the cerebellum. Table 6.5 presents the most discriminating regions in the classification task, namely the brain regions with the highest (absolute) weight value that contributed relatively higher to the decision function. Specifically, the regions that contributed more in the classification of the ARMS-T subjects included the cerebellum, parts of the superior temporal pole bilaterally, the right anterior cingulate cortex, the right superior medial frontal and left orbitofrontal cortex and the insula bilaterally, whereas regions with a higher weighted average for the ARMS-NT group consisted of the right inferior parietal lobe, right medial temporal lobe, the right orbitofrontal cortex and the left pallidum (Table 6.5).

Table 6.4 Misclassification analysis

	ARMS-T → ARMS-NT	ARMS-T → ARMS-T	*P*	ARMS-NT → ARMS-T	ARMS-NT → ARMS-NT	*P*
Socio-demographic variables						
N	6	10		3	16	
Mean age at baseline, y (sd)	29.2(9)	25.4(4.5)	ns[a]	24.8(7.2)	23.8(6.2)	ns[a]
Sex (male), n (%)	2(33)	9(90)	<0.05 [b]	1 (33)	9(56)	ns[b]
Educational level			ns[c]			ns[c]
<9 y, n (%)	1	2		2	6	
9-11 y, n (%)	2	5		1	6	
12-13 y, n (%)	2	3		0	2	
<13 y, n (%)	1	0		0	2	
Cannabis use at baseline			ns[c]			ns[c]
none	4	6		2	9	
rarely	1	0		0	1	
Several times/month	0	0		1	1	
Several times/week	1	3		0	0	
daily	0	1		0	5	
Antipsychotics before entry	5	1	<0.05 [b]	0	1	ns[b]
Anti-depressants at baseline	3	4	ns[b]	2	3	ns[b]
Clinical variables						
Mean BPRS total score at baseline (sd)	45.7(11.5)	40.2(10.2)	ns[c]	38.3(12.9)	37.5(6.2)	ns[c]
Mean BPRS Depression/Anxiety score (sd)	12(3.5)	7.5 (2.3)	<0.05 [c]	7.3(2.1)	7.9 (3)	ns[c]
Mean SANS global score at baseline (sd)	9.7(7.7)	9.8(4.8)	ns[c]	10.3(5)	6.3(4.7)	ns[c]
Mean interval between MRI and disease onset, d (sd)	427.5(483.6)	245.7(215.3)	ns[a]			

[a] Student's *T*-test. [b] Fisher's exact test. [c] Mann-Whitney U-test

A discrimination map showing the spatial pattern by which the groups differ is also illustrated in Figure 6.3. We emphasize that this spatially distributed pattern should not be interpreted as a statistical map, but rather as a spatial representation of the decision boundary.

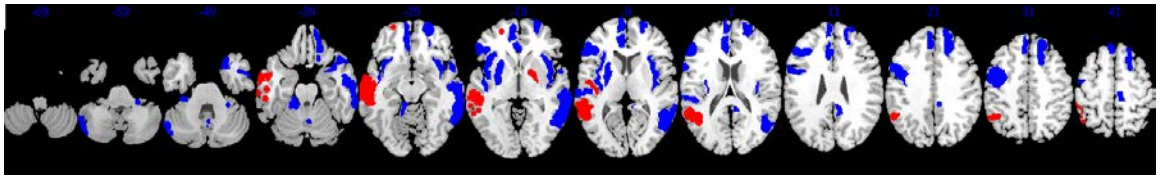


Figure 6.3 Discrimination maps for the classification of ARMS-T vs ARMS-NT. The colours represent the weight of each feature in the classification function (the red scale represents positive weights and the blue scale represents negative weights). The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=ARMS-T; -1=ARMS-NT group), a positive weight means the weighted average in that region was higher for the ARMS-T group, and a negative weight means the weighted average was higher for ARMS-NT group. Note: features correspond to GM volume measures in the AAL-defined brain regions, and not voxels.

Finally, I used the EHRS dataset (the 17 HR[ill] and the 17-matched HR[symp] subjects described in Chapter 5) to train the SVM-RFE method and then tested it using the FePsy data in order to examine the performance of the method when using one HR sample and testing on another. A 65.7% accuracy (sensitivity/specificity: 37.5%/89.5%) was observed, with the method failing to classify above chance levels the ARMS subjects that later developed psychosis. This low sensitivity and classification performance might imply a divergent neuroanatomical pattern between clinical and familial HR cohorts. Additionally, it should be borne in mind that the HR[ill] sample in the EHRS was labelled as such on the basis of a later diagnosis of schizophrenia (according to PSE criteria and later the ICD-10), whereas the ARMS-T subjects in the FePsy study were characterised as making a transition to psychosis, not schizophrenia per se, based on transition criteria of the PACE clinic (see Table

6.1-Yung et al. 1998).

Table 6.5 List of the most discriminative regions for the classification of ARMS-T vs ARMS-NT

Lobe	Region/Hemisphere	**w**
Negative weights		
Cerebellum		
	Cerebellum_Crus2_R	-0.0128
	Cerebellum_3_R	-0.0194
	Cerebellum_4_5_R	-0.0207
	Cerebellum_6_L	-0.0127
	Cerebellum_7b_R	-0.0146
	Cerebellum_10_L	-0.0247
	Vermis_8	-0.0147
Temporal		
	Temporal_Sup_R	-0.0067
	Temporal_Pole_Sup_L	-0.0084
	Temporal_Mid_L	-0.0075
Frontal		
	Frontal_Sup_L	-0.0197
	Frontal_Sup_Orb_L	-0.0166
	Frontal_Mid_R	-0.0089
	Frontal_Inf_Tri_R	-0.0106
	Frontal_Sup_Medial_R	-0.0084
	Frontal_Med_Orb_R	-0.0098
	Precentral_R	-0.0176
Parietal		
	Postcentral_R	-0.0102
	Paracentral_Lobule_R	-0.012
Limbic		
	Cingulum_Ant_R	-0.017
	Cingulum_Post_L	-0.0081
Basal ganglia		
	Putamen_R	-0.0172
Perisylvian		
	Insula_L	-0.0134
	Insula_R	-0.0206
Positive weights		
Temporal		
	Temporal_Mid_R	0.0069
	Heschl_R	0.0069
Frontal		
	Frontal_Sup_Orb_R	0.0101

Parietal		
	Parietal_Inf_R	0.0137
Basal ganglia		
	Pallidum_L	0.0072

Ant, anterior; Crus, crust; Inf, inferior; L, left hemisphere; Mid, middle; Med, medial; Orb, orbital; Post, posterior; R, right hemisphere; Sup, superior; w, weight vector of corresponding features in the classification process. Note: The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=ARMS-T; -1=ARMS-NT group), a positive weight means the weighted average in that voxel was higher for the ARMS-T group, and a negative weight means the weighted average was higher for ARMS-NT group.

6.4.3. Voxel-based morphometry analysis

No significant GM volume differences between the ARMS-T and ARMS-NT subjects were observed when conventional VBM analysis was employed using a family error rate (FWE) of $p<0.05$. At $p<0.001$ uncorrected for multiple comparisons, volumetric differences were observed within the cerebellum, left and right medial and inferior frontal lobe regions and the medial temporal lobe bilaterally. To compare our SVM results with those from the VBM analysis, I lowered the p-value at $p<0.05$ (uncorrected) and observed widespread volumetric between-group differences in the superior and medial temporal lobe bilaterally, the inferior and superior frontal lobe bilaterally, the left superior and inferior parietal lobe, the precentral gyrus bilaterally and parts of the cerebellum.

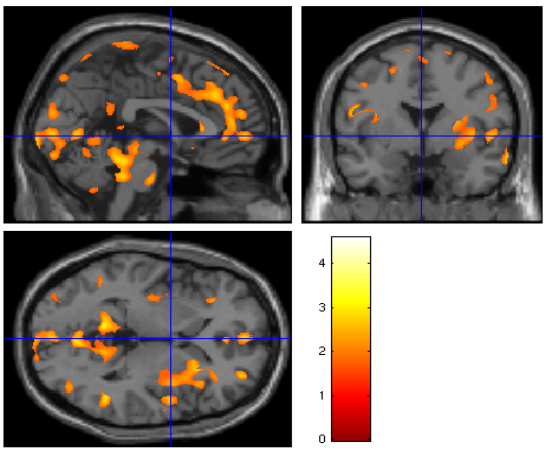


Figure 6.4 Results of the conventional VBM analysis for the smoothed and normalized GM segments (*p*-value<0.05, uncorrected).

6.5 Discussion

The present findings replicate the previous ones in that MRI-based classification methods were able to predict transition to psychosis in subjects at high clinical risk for developing the disorder using neuroanatomical data at study inclusion. The SVM-RFE classifier achieved 74% accuracy in classifying ARMS-T against ARMS-NT subjects.

The neuroanatomical decision function that discriminated the two groups was associated with GM abnormalities relying on a distributed network of regions covering most cortical and sub-cortical brain structures and the cerebellum. Our present findings agree with findings from a recent voxel-based meta-analysis that reported GM volume reductions in subjects that convert to psychosis in the insular and superior temporal lobe cortices (Fusar-Poli et al. 2011) and also with previous VBM findings on the same dataset (Borgardt et al. 2007a).

Few significant between-group GM differences were detected using conventional VBM analysis at a p<0.001 but none was detected when correction for multiple

comparisons was performed. At a lower statistical threshold, $p<0.05$, significant GM volume differences were detected in a network of regions that was largely in agreement with the one identified by the SVM method. However, as stated in previous chapters, an SVM-based classification analysis is advantageous to the conventional VBM analyses in that it can detect subtle neuroanatomical alterations that are not confined to a circumscribed set of few brain regions but rather span across a complex network of inter-correlated brain structures that jointly produce a maximal degree of separation between converters and non-converters. The finding of complex and widespread networks of differentiation that can be achieved through SVM is in line with current disconnection hypotheses of schizophrenia (Friston 1999) that state that the pathology of schizophrenia is associated with highly distributed brain alterations across a number of cortical and subcortical structures.

Despite the fact that the classification accuracy observed in this investigation is significant, it is, however, lower than the accuracy observed in the familial high-risk group (Chapter 5) and that reported in another recent study that similarly implemented a MRI-based SVM classifier from the same ARMS cohort (Koutsouleris et al. 2012a).

In the study conducted by Koutsouleris and co-workers, (2012a), their MRI-based SVM classifier achieved 84.2% accuracy in the critical ARMS-T versus ARMS-NT analysis, with only 2 ARMS-T subjects being misclassified as ARMS-NT and 4 ARMS-NT subjects wrongly labeled as ARMS-T (sensitivity/specificity: 81%/87.5%). However, their implementation of the SVM classifier was completely different and relied upon the construction of nonlinear SVM ensembles that incorporated feature selection, model training and predictive learning wrapped together in a repeated nested cross-validation framework. Ensemble learning approaches are usually selected on the basis that they can achieve higher predictive performance than single classifiers, by combining multiple weak learning models that decide upon the classification of a new instance through majority voting (Polikar 2006). In contrast, our approach is based on a simpler SVM framework that is known to provide better generalization performance whereas more complex models tend to overfit the data in that they provide great training accuracy (small error on the

training set) but often misclassify unknown data instances (higher test error).

Compared to the diagnostic performance of our classifier in the genetic high-risk cohort of the EHRS, the classification performance in the ARMS groups of the FePsy study was notably lower, contrary to what would be expected since the ARMS groups represent help-seeking individuals, most of whom already manifest putative transient and/or sub-threshold psychotic symptoms. Interestingly, 5 out of the 6 ARMS-T subjects that were misclassified received anti-psychotic medication some time before study inclusion (Table 6.4) while the other misclassified ARMS-T subject was prescribed tranquilizers (Lorazepam). Many studies have reported the effect of antipsychotic medication on grey matter volume in the direction of significant regional reductions (Navari and Dazzan 2009, Smieskova et al. 2009), thus possibly suggesting a neuroanatomical heterogeneity expressed with divergent pathophysiological trajectories between subjects receiving and subjects not receiving any anti-psychotic treatment. Additionally, the misclassified ARMS-T subjects had significantly higher scores on the Depression/Anxiety subscale of the BPRS list compared to the correctly classified ARMS-T individuals ($p<0.05$; Table 6.4) while there were no differences between the groups in terms of anti-depressant medication use at baseline (Table 6.4). Depressive symptomatology is a common feature in populations at high clinical risk for psychosis (Hafner et al. 2005). However it remains unclear whether these symptoms are reflected at the neuroanatomical level and thus possibly suggest a depressive sub-syndrome in sub-threshold psychosis or the established state in the same way as positive, negative and disorganized sub-syndromes in schizophrenia could be discerned (Nenadic et al. 2010, Zhang et al. 2015).

Despite the lower diagnostic performance, our MRI-based classifier managed to increase the diagnostic certainty from 45.7% to 77% in case of a positive test result, suggesting that an MRI-based pattern classification system could, with refinement, become a useful part of a multi-step diagnostic procedure that would reliably quantify the risk for conversion to psychosis and inform appropriate care and treatment strategies.

Certain limitations of this study have to be considered. Again the sample size in this

investigation is small. The rate of transition to psychosis amounted to nearly 46%, which is generally in keeping with other clinically at-risk cohorts (Koutsouleris et al. 2009b, Yung et al. 2003, Klosterkotter et al. 2001). However, it is not clear how the classifier would perform if presented with an ARMS cohort with significantly lower conversion rates. Finally, the administration of antipsychotic and antidepressant medication might have confounded our results, despite the fact that any drug treatment was administered before study inclusion.

6.6 Conclusions

In this investigation, previous classification results have been replicated, with current findings suggesting that an early diagnosis of psychosis based on neuroanatomical-based classifier is feasible for cohorts present with sub-threshold disturbances and early signs of psychosis. It remains to be elucidated, however, whether these findings can be generalized to larger cohorts of at-risk samples that are recruited using different assessment criteria and scanned on different scanners and/or using different imaging protocols.

CHAPTER 7

Towards the identification of neuroimaging-based biomarkers for the prediction of psychosis across familial and clinical high-risk cohorts

7.1 Introduction

In the previous chapters, the diagnostic performance of SVM in predicting conversion to schizophrenia and psychosis-related disorders was examined in cohorts at high-risk either due to familial or clinical reasons. However, it remains unclear if it is feasible to detect biomarkers that generalize across research sites, where differences in the recruitment process, MRI scan acquisition and pre-processing might exist.

In this chapter, I attempt a preliminary study of the feasibility of identifying MRI-based biomarkers that can predict transition to psychosis among high-risk individuals recruited at two different research centers, using different assessment criteria and scanned on different scanners.

7.2 Background

Neuroanatomical pattern classification has recently facilitated the identification of imaging biomarkers for the diagnosis and early prediction of various neuropsychiatric disorders such as Alzheimer's disease (Fan et al. 2008b, Kloppel et al. 2008b) and schizophrenia (Davatzikos et al. 2005b, Koutsouleris et al. 2009b), and has pointed in the direction of a translational application of MRI into clinical practice by evaluating distinct patterns of differentiation at the single-subject level.

An important step in the investigation of the clinical utility of imaging biomarkers is to test their generalization performance in independent cohorts (Phillips et al. 2006). This, however, implies a need for large data sets, which is not always possible in single-site, MRI studies. A limited number of studies in schizophrenia have attempted to build classification models based on sets of patients and matched groups of control subjects and then tested their models' performance by classifying independent samples of patients and controls (Kawasaki et al. 2007, Nieuwenhuis et al. 2012, Schnack et al. 2014), and reported good levels of classification accuracy ranging from 70% to 90%. For a detailed review of these studies, please see Chapter

3.

The requirement for large data sets could be satisfied by pooling data across different research sites with the added benefit of both increasing statistical power and possibly enhancing the classifier's performance by virtue of the higher number of samples used for training (Kloppel et al. 2009, Nieuwenhuis et al. 2012, Abdulkadir et al. 2011). An increasing number of multi-center studies aim to combine data acquired on different sites (Mwangi et al. 2012, Koutsouleris et al. 2015b), despite any methodological and between-scanner variability (Stonnington et al. 2008, Moorhead et al. 2009, Suckling et al. 2011, Abdulkadir et al. 2011).

In the first and, to date, only MRI-based cross-center study of prediction of psychosis, Koutsouleris and co-workers combined together two independent cohorts of subjects with an ARMS, recruited and scanned at two different early recognition centers, and examined the prognostic accuracy of their MRI-based SVM ensemble classifier (Koutsouleris et al. 2015b). Their classification system achieved a balanced accuracy of 80.3% in the pooled data set (sensitivity=75.8%, specificity=84.8%) and also enabled a risk staging procedure through additional Kaplan-Meier survival analyses that simultaneously quantified the risk of an ARMS subject in making a transition to psychosis along with an estimation of the time to transition.

In this investigation, our aim is to pool baseline neuroanatomical data from high-risk subjects drawn from the Edinburgh High Risk Study (EHRS) and the Basel FePsy study and examine the feasibility of identifying MRI-based biomarkers across the two high-risk cohorts. No other study to date has attempted to pool genetic and clinical HR cohorts together and examine the existence of a neuroanatomical signature that exists across differently ascertained HR populations.

Differences between the two high risk paradigms begin at the very conceptualization of the underlying risk to psychosis. As described in previous chapters, genetic/familial high-risk cohorts include monozygotic and dizygotic twins discordant for schizophrenia and/or individuals with first- or second-degree relatives

affected with the disease. On the other hand, clinical HR populations are presented with different aspects of sub-threshold positive and negative symptomatology and/or possibly mild cognitive and functioning impairments. The identification of clinical HR individuals presents some variability that is attributed to each early detection center and the corresponding high-risk criteria used (Smieskova et al. 2010). As a result, rates of transition to psychosis vary significantly across sites (Fusar-Poli et al. 2012, Fusar-Poli et al. 2013a), with some reporting as many as 35% of subjects at clinical HR developing psychosis (Pantelis et al. 2003, Yung et al. 2004). It has been estimated that the average transition rate for clinical HR cohorts amounts to 30% whereas only 10% on average of the subjects identified at high-risk for genetic or familial reasons develops psychosis (Smieskova et al. 2013).

Despite their discrepancies, several aspects of familial and clinical HR paradigms overlap, as several individuals identified at a clinical HR have a family history of psychosis and many individuals at familial HR often exhibit pre-psychotic symptoms. In addition, at a structural level clinical and familial HR cohorts seem to share similar volumetric abnormalities (Bois et al. 2015), especially in prefrontal, medial temporal and limbic regions, with individuals showing prodromal psychotic signs exhibiting additional insular and caudate structural deficits (Smieskova et al. 2013). For a detailed discussion of structural MRI studies in HR cohorts see Chapter 2 (section 2.3.2).

In the following section, a comparative presentation of the familial and clinical HR individuals is given.

7.3 Materials and Methods

7.3.1 Subjects

The pooled dataset consists of 69 individuals at high risk for psychosis that were recruited in the prospective studies conducted independently in Edinburgh, at the Department of Psychiatry (Edinburgh High Risk Study, EHRS) and at the University of Basel, Switzerland ('Fruherkennung von Psychosen', FePsy). A detailed description of the studies was given in the three previous chapters. All aspects of the studies were reviewed and approved by the appropriate local ethics committees at both research sites.

Briefly, a familial high-risk paradigm was employed in the EHRS, in that high-risk individuals were identified and included in the study on the basis of having one or more first- or second-degree relatives affected with schizophrenia. All high-risk subjects had no previous history of psychosis or any neuropsychiatric disorder and none of those subjects received anti-psychotic medication at any point during the study or until they fulfilled operational criteria for schizophrenia (when appropriate clinical management was advised and follow-up assessments were discontinued). As described in detail in chapter 4 (sub-section 4.1.1.1), psychopathology in HR subjects in the EHRS was assessed using the PSE criteria (Wing et al. 1974), which allowed the classification of subjects into a 5-scale system based on their PSE subscores (Johnstone et al. 2000). For facilitating future studies in the data set, the HR subjects were stratified based on the presence/absence of (psychotic) symptoms according to the PSE as: fully or partially held psychotic symptoms (scores 2 and 3) HR[symp], absence of psychotic symptoms, (scores 0 and 1) HR[well] or diagnosis of schizophrenia (score 4), HR[ill].

Additionally, all HR subjects underwent baseline and follow-up assessments to evaluate neuropsychological measures (such as verbal learning and memory- RAVL test, executive function, general IQ etc.), schizotypal cognitions via the Rust Inventory of Schizotypal cognitions (RISC) and the Structured Interview for Schizotypy (SIS; Miller et al. 2002b) and elicit psychopathological indices and

transition to schizophrenia again using the PSE. For a detailed review of the clinical and neuropsychological assessments see Johnstone et al. 2002a, Johnstone et al. 2005 and Byrne et al. 2003. (Additionally, a detailed description of the PSE and the RISC items can be found in Chapter 4 (subsection 4.1.1.1) and Chapter 5 (5.3.2) accordingly).

From those HR subjects who provided complete clinical assessments and had a MRI scan, 17 were diagnosed at follow-up with schizophrenia (after an average of 929 days, SD=138) based on the ICD-10. Among the rest, 57 subjects experienced psychotic symptoms (HR[symp]) but these were too transient or mild to warrant a diagnosis of schizophrenia (Johnstone et al. 2000). The rest of the HR subjects remained well, with no symptoms (HR[well]). In this investigation I was mainly interested in the HR[ill] and HR[symp] groups. Again, as seen in Chapter 5, 17 out of the 57 HR[symp] subjects were used to train the classifier, and the remaining 40 HR[symp] were used to further validate the classification performance.

In the FePsy study, ARMS (at-risk mental state) subjects were identified on the basis of the ultra-high risk criteria of the well-established Personal Assessment and Crisis Evaluation (PACE) definitions (Yung et al. 2003). Briefly, ARMS inclusion required (a) Attenuated Positive Symptoms, and/or (b) Brief Limited Intermittent Psychotic Symptoms (both fulfilling specific time criteria) or (c) decline in global functioning combined with family history of psychosis. Additionally prodromal symptoms were assessed with the Brief Psychiatric Rating Scale (BPRS) and the Scale for the Assessment of Negative Symptoms (SANS). For a detailed description of the ARMS inclusion, exclusion and assessment criteria see Chapter 4 (subsection 4.1.2.1) and/or Chapter 6 (subsection 6.3.1 and Table 6.1).

All subjects were followed up for over 4 years and were offered supportive counselling. Transition to psychosis was monitored monthly in the first year and quarterly thereafter and was based on the BPRS severity thresholds (see sections 4.1.2.1 and Table 6.1) and further corroborated using the ICD-10. Based on these criteria, the ARMS groups were subdivided in 16 subjects with transition to psychosis (ARMS-T; diagnosed after an average of 306.3 days, SD= 318.3) and 19 non-conversion individuals (ARMS-NT). Six out of the 16 ARMS-T subjects were

administered low doses of antipsychotic medication for behavioural control before study inclusion (1 participant on olanzapine, 2 Chlorprothixene and 3 risperidone) and also 1 ARMS-NT subject was on olanzapine before study entry.

As it is clear from the descriptions above, the high-risk groups recruited across the two research sites differed in terms of criteria for inclusion and with regards to the assessment of psychopathological indices to elicit psychotic symptomatology and transition to schizophrenia and psychosis-related disorders. A comparative presentation of the PSE and BPRS items to be used as a reference is given in Appendix II (Table 1). Despite the fact that several aspects of psychopathology might overlap between the two assessments tools (such as the depression, anxiety or hallucinations items), severity ratings are appraised differently. For example the PSE scores symptoms as absent, partially/transiently or definitely present. As detailed previously (Johnstone et al. 2000), partial delusions and/or transient hallucinations were regarded as psychotic symptoms and correspond to the clinical high risk literature category of attenuated and/or brief psychotic symptoms. Non-psychotic symptoms are not, however, recorded in the same way between the two assessment tools nor is functional decline assessed in the PSE. Therefore, a direct comparison between the clinical and familial high-risk assessment criteria is not feasible- other than that similar positive psychotic symptomatology was assessed.

To build the MRI-based classifier, 33 converters (17 HR[ill] and 16 ARMS-T) were pooled together to contrast 36 nonconverters (17 HR[symp] and 19 ARMS-NT), while the data of the remaining 40 HR[symp] subjects were used to further validate the classifier (Table 7.1).

Table 7.1 Demographic and clinical details

	EHRS			FePsy			Pooled Data			Test Data
	HR[ill]	HR[symp]	*P*	ARMS-T	ARMS-NT	*P*	HR-T	HR-NT	*P*	HR[symp]
N	17	17		16	19		33	36		40
Mean age at baseline, y (sd)	20.07 (2.37)	20.03 (2.6)	ns	26.8 (6.5)	23.9 (6.2)	ns	23.4 (5.9)	21.9 (5.1)	ns	21.07 (4.59)
Sex (male), n (%)	11 (64.7)	11 (64.7)	ns	11 (69)	10 (52)	ns	22 (67)	21 (60)	ns	14 (35)
Handedness (mixed or left), n (%)	1 (5)	2 (12)	ns	3 (19)	1 (5)	ns	4 (12)	3 (8)	ns	6 (3)
Cannabis use at baseline			ns			ns			ns	
None	8	12		10	11		18	23		26
Occasional	5	4		1	3		6	7		10
Frequent	4	1		5	5		9	6		4
Antipsychotics before entry, n (%)	0	0	ns	6 (37.5)	1 (5)	<0.05	6 (18)	1 (3)	<0.05	0
Family History (1rst: 2nd relative)	17 (9:8)	17 (10:7)	ns	1 (1:0)	3 (2:1)	ns	18 (10:8)	20 (12:8)	ns	40 (27:13)

ARMS-T: subjects at a clinical risk mental state that later developed psychosis; ARMS-NT: at-risk mental state subjects that did not make a transition. HR[ill]: individuals at high familial risk who developed schizophrenia during follow-up period. HR[symp]: individuals at high familial risk who remained well but developed psychotic symptoms during follow-up period. Fisher's exact tests were performed for variables of sex, handedness, cannabis use, antipsychotic medication and family history, and standard T-test was applied to the rest.

7.3.2 Image Acquisition and Preprocessing

Baseline scanning in the EHRS was performed on a 1.0 T Siemens Magnetom scanner (Erlangen, Germany) and for the FePsy study on a 1.5 Siemens Magnetom Vision scanner (Erlangen, Germany). Details of the acquisition protocols can be found in Chapters 5 (section 5.3.3) for the EHRS scans and Chapter 6 (section 6.3.2) for the FePsy images.

The preprocessing pipeline has been extensively described in Chapter 4, section 4.2 and additional information has been given in relevant sections of chapters 5 and 6. The same pre-processing steps (and specific parameters) were applied to all brain images of the EHRS and the FePsy study.

7.3.4 Multivariate Pattern Classification Analysis

This is identical to what was described in Chapter 5 (section 5.3.4) and 6 (section 6.3.3).

7.4 Results

7.4.1 Socio-demographic findings

There were no significant differences between subjects in the study groups, within each research center, in terms of age, sex, handedness, cannabis use and family history (Table 7.1). However, as noted before, the use of antipsychotic medication was significantly different for converters and nonconverters in the Basel, FePsy study and thus in the pooled data set. Additionally, there were no significant ‘center * group’ effects for any of these variables, whereas there was a significant main effect of research center on age (factorial ANOVA; F=21.135, df=2, p=0.0001).

7.4.2 SVM classification analysis

In the pooled dataset (N=69), the MRI-based system correctly classified converters and nonconverters with a balanced accuracy of 82.6% (Table 7.2, sensitivity/specificity: 79%/86%, permutation test $p<0.001$). From the independent set of 40 non-converters, HR[symp] subjects only 9 were wrongly classified as converters (specificity: 77.5, FPR: 22.5%). Thus the classification performance on the entire dataset (N=109) attained a balanced accuracy of 80.7% (sensitivity/specificity: 79%/82%). Given a pre-test probability of 30.3% (equals to transition rate: 33/109), the positive/negative likelihood rations of 4.3/0.26 indicate that an HR person with a positive/negative MRI finding will have a post-test probability of 65%/35% of developing psychosis.

Both the misclassified HR-T and the misclassified HR-NT subjects did not significantly differ from the correctly classified HR-T and HR-NT individuals accordingly, on any of the socio-demographic variables (Table 7.3). Direct comparison on any clinical, neurocognitive and behavioural variables was not, however, possible becaure the EHRS and FePsy data samples did not use the same assessments to record these variables.

Table 7.2 Classification performance

Dataset	TP	TN	FP	FN	Sens (%)	Spec (%)	BAC (%)	FPR (%)	PPV (%)	NPV (%)	LR+/LR-
Pooled	26	31	5	7	78.8	86.1	82.6	13.9	83.8	81.5	5.67/0.25
$HR[symp]_{test}$	0	31	9	0	-	77.5	-	22.5	-	100	-
Overall	26	62	14	7	78.8	81.5	80.7	18.4	65	89.8	4.26/0.26

The diagnostic performance was evaluated by means of sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), false positive rate (FPR) and positive/negative predictive value (PPV/NPV). LR+ was also calculated as sensitivity/1-specificity and LR- = 1-sensitivity/specificity.

The neuroanatomical decision function discriminating between converters and non-converters involved an extensive spatial network of regions covering the superior temporal lobe bilaterally, the left middle temporal lobe, the prefrontal cortex bilaterally covering superior, orbitofrontal and ventromedial lobe structures and extending to inferior frontal and insular structures. These brain regions included the regions that contributed higher in the classification of the HR-T subjects; that is the regions that were more important in classifying HR-T instead of HR-NT subjects. In contrast, the regions contributing more to the classification of the HR-NT group consisted of regions covering the left inferior temporal lobe extending to perisylvian structures (right temporal pole, left insula), the right superior orbital and left medial orbital frontal lobe, the right inferior parietal lobe covering the right precuneus, the left parahippocampal gyrus and the vermal lobule 6 and cerebellar lobule 9. Table 7.4 below presents a list of the most discriminating brain regions in the classification.

Table 7.3 Missclassification analysis

	HR-T → HR-NT	*HR-T → HR-T*	*P*	*HR-NT → HR-T*	*HR-NT → HR-NT*	*P*
Socio-demographic variables						
N	7	26		5	32	
Mean age at baseline, y (sd)	21.3(2.5)	23.9(6.5)	ns[a]	26.6(7.6)	21.3(4.2)	ns[a]
Sex (male), n (%)	5 (71.4)	17 (65.4)	ns[b]	4(80)	17(53.2)	ns[b]
Neuroleptics before entry	1	5	ns[b]	0	1	ns[b]
Cannabis at baseline			ns[b]			ns[b]
none	3	15		2	22	
occassionally	2	4		2	7	
frequent/severe	2	7		1	3	
Family History			ns[b]			ns[b]
no relative	2	13		3	13	
one 1rst degree	0	1		0	3	
one 2nd degree	3	4		1	5	
one 1rst & 2nd degree	2	5		0	9	
two 2nd degree	0	3		1	2	

[a] Student's *T*-test. [b] Fisher's exact test

Table 7.4 List of the most discriminative regions in the classification of the pooled dataset

Lobe	Region/Hemisphere	**w**
Negative weights		
Cerebellum		
	Cerebellum_Crus1_R	-0.5301
	Cerebellum_Crus2_R	-0.5389
	Cerebellum_3_R	-0.1346
	Cerebellum_4_5_R	-0.3237
	Cerebellum_10_R	-0.4523
	Vermis_8	-0.2955
Temporal		
	Temporal_Pole_Sup_L	-0.3212
	Temporal_Pole_Sup_R	-0.1507
	Temporal_Mid_L	-0.2674
Frontal		
	Frontal_Sup_L	-0.7231
	Frontal_Sup_Orb_L	-0.475
	Frontal_Inf_Tri_R	-0.3904
	Frontal_Mid_Orb_R	-0.2946
	Frontal_Sup_Medial_L	-0.1827
Occipital		
	Occipital_Mid_L	-0.6979
Limbic		
	Cingulum_Mid_R	-0.4449
Basal ganglia		
	Caudate_L	-0.259
	Caudate_R	-0.38
Perisylvian		
	Insula_R	-0.5488
Positive weights		
Cerebellum		
	Cerebellum_9_R	0.3593
	Vermis_6	0.2216
Temporal		
	Temporal_Inf_L	0.1709
	Temporal_Pole_Mid_R	0.3635
Frontal		
	Frontal_Sup_Orb_R	0.3211

Table 7.4 Continued

	Frontal_Med_Orb_L	0.653
Parietal		
	Parietal_Inf_R	0.7814
	Precuneus_R	0.254
	Paracentral_Lobule_R	0.4265
Limbic		
	Cingulum_Post_R	0.4616
	ParaHippocampal_L	0.1523
Perisylvian		
	Insula_L	0.3815

Ant, anterior; Crus, crust; Inf, inferior; L, left hemisphere; Mid, middle; Med, medial; Orb, orbital; R, right hemisphere; Sup, superior; w, weight vector of corresponding features in the classification process. Note: The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1=HR-T; -1=HR-NT group), a positive weight means the weighted average in that voxel was higher for the HR-T group, and a negative weight means the weighted average was higher for HR-NT group.

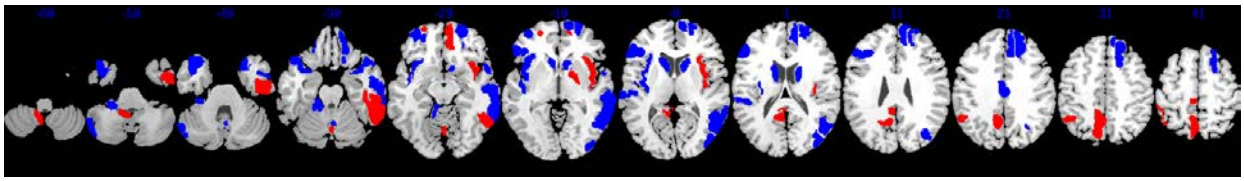


Figure 7.1 Discrimination maps for the classification of the pooled dataset of HR-T vs HR-NT. The colours represent the weight of each feature in the classification function (the red scale represents positive weights and the blue scale represents negative weights). The SVM weight vector is a linear combination or weighted average of the support vectors and defines the decision boundary. The weight vector is therefore a spatial representation of the decision boundary. Every feature contributes with a certain weight to the decision boundary or classification function. Given a positive and a negative class (+1= HR-T; -1=HR-NT group), a positive weight means the weighted average in that region was higher for the HR-T group, and a negative weight means the weighted average was higher for HR-NT group. Note: features correspond to GM volume measures in the AAL-defined brain regions, and not voxels.

7.5 Discussion

This chapter provides a preliminary analysis and evidence for the identification of an MRI-based signature that exists across two independent HR populations that were recruited using different ascertainment criteria and scanned with different scanners and imaging protocols. The predictive power of the classifier was tested using cross-validation and by assigning an independent group of 40 HR subjects that did not develop schizophrenia at follow-up (HR-NT). The generalization performance of the SVM-RFE classifier in the entire dataset attained a balanced accuracy of 81%.

The neuroanatomical pattern that distinguished between HR-T and HR-NT subjects revealed a network of spatially distributed regions in the brain cortex, covering prefrontal, temporal and cerebellar structures and extending to a few subcortical areas such as the caudate nucleus bilaterally that had a higher weighted average for the HR-T grop. These results are in keeping with previous findings that were described in Chapter 5 and 6, and other imaging studies reporting similar GM abnormalities (Fusar-Poli et al. 2011, Smieskova et al. 2013, Koutsouleris et al. 2009b, Koutsouleris et al. 2011, Koutsouleris et al. 2015b). In contrast, regions contributing more to the classification of the HR-NT group included the orbito-frontal, inferior temporal, parietal lobe and parts of the cerebellum, partly in line with previous studies (Borgwardt et al. 2007b, Koutsouleris et al. 2015b).

To date, only one study has attempted to pool data together from subjects that were recruited and scanned at two different early psychosis centers and achieved comparable classification accuracy (Koutsouleris et al. 2015b). In this study, Koutsouleris and colleagues combined together data from 73 ARMS subjects that were recruited in two independent early recognition sites, in Munich and in Basel. Contrary to our case, both research sites employed the same ultra-high-risk inclusion criteria corresponding to the PACE definitions and transition to psychosis was corroborated using similar severity threshold scales, corresponding to criteria by Yung et al. 1998. Participants in both sites were scanned using the same platform,

Siemens Magnetom Vision 1.5T scanners, but different imaging parameters were selected at each site. No calibration of the scanners was performed but a voxel-level adjustment of the group-level means was performed post-hoc in order to correct for inter-scanner differences.

Despite the same rationale for pooling data together, a direct comparison of the present study with the study conducted by Koutsouleris and colleagues is not feasible due to the differences in the recruitment, assessment, scanning and pre-processing procedures of the study samples and the different implementation of the SVM methodology. However, both studies achieved comparable diagnostic accuracies in classifying HR-T against HR-NT subjects and share similar networks of relative GM volume reductions and increases of the neuroanatomical predictor, as discussed above.

There are numerous advantages to pooling data from multiple sites together. The opportunity to accrue larger numbers of subjects allows the introduction of a phenotypic and demographic diversity of the cohort under study along with the possible inclusion of more rare cases of the diseased population. Additionally, the increased number of subjects allows for increased sensitivity and may also ensure enough statistical power to detect more subtle effects or even subgroups within a cohort that might share structural commonalities. All these combined may result in a deeper understanding of the under-study disease.

On the other hand, combining data from multiple research sites may introduce some variability. In the case where the data are, also, acquired from different MRI scanners and imaging protocols, potential scanner-related confounds may be introduced that in turn may affect the integrity and robustness of the results and make their interpretation difficult.

Variations in image intensity and the resulting necessity for standardizing intensity values is not a problem specific to multisite studies but also applies to longitudinal and/or cross-sectional studies where scanner drifts or upgrades may impact on the quality of the scans. Nonuniformity in the imaging sequence and the RF field coils and differences in the positioning of the patient inside the scanner can introduce

variability in image intensity with any given scanner. Intensity inhomogeneity, partial volume effects and variability in image quality (as measured with the signal to noise ratio) can in turn impact on segmentation (Clark et al. 2006, Acosta-Cabronero et al. 2008, Li et al. 2005, Klauschen et al. 2009). However, some of this variation can be attenuated with the use of segmentation processes that do not solely rely on voxel intensities but also incorporate spatial information and correction for bias field inhomogeneity as well. In this analysis, the use of the unified segmentation method in SPM5, which involves fitting spatial priors to the image and bias-field corrections to account for differences in head shape and positioning, ensured the robustness of the segmentation and the consistency of image intensities.

A significant limitation of this preliminary analysis is, however, the potential introduction of confounds that might have occurred as a result of systematic between-site differences, which might have confounded 'actual' disease-related effects. As briefly described above, calibration and/or intensity standardization processes can be performed either before or after scanning in order to effectively account for scanner-related intensity differences.

Briefly, test-retest and calibration studies can ensure the reliability and comparability of multi-center scans by repeatedly scanning a number of participants in all sites and measuring a phantom (van Haren et al. 2003, Jovovich et al. 2006, Tofts 1998). It is not, however, always possible to have a (sub) group of participants available to be scanned in all research sites in order to measure reliability between scans and ensure consistency and quality control.

Post-scanning procedures are, thus, more attractive and include histogram matching or histogram standardization techniques that are designed to bring MRI intensities to a common scale, where similar intensities would have similar tissue meaning. Histogram matching methods try to match the subjects' histogram to a standard (reference) histogram by either matching the intensities in a series of chosen landmarks (Nuyl & Udupa 1999) or by minimizing some information criteria (Jager & Hornegger 2009). Recently, advanced histogram matching techniques that employ spatial correspondence (between the image volumes and the reference image) and nonlinear transformations to account for the 'nonlinear' impact of the MRI scanner

to voxel intensities have also been proposed (Robitaille et al. 2012). (A thorough presentation and comparison of histogram matching techniques is out of the scope of the present thesis and the interested reader is referred to the listed citations for more information.)

In the analysis presented here, the initial intention was to scale the image intensities between the two datasets, the EHRS and FePsy sMRI data, by matching their intensity histograms (that is by finding a correspondence between the study histograms), after pre-processing was completed and right before the smoothed, normalized GM maps entered the SVM analysis. However, the most reasonable thing to do would be to perform any histogram matching or equalization technique over the whole brain in order to account for the variance for all tissue types because from the segmentation step and downwards each preprocessing procedure depends on some point on voxel intensity values. Due to time constraints, however, this was not feasible in the present study and after observing similarities in the intensity histograms of the smoothed, normilized GM segments in the EHRS and FePsy datasets, it was decided to proceed with the data pooling without any scaling or matching. Please see Appendix V for the histograms of the two data sets and a more detailed discussion as to why scaling was not performed.

It is not yet certain if and how scanner-related heterogeneity can impact on the performance of pattern recognition classifiers and specifically SVM. Previous SVM-based studies have reported an improvement in performance when larger numbers of subjects were used for training (Kloppel et al. 2008, Abdulakadir et al. 2011), despite any scanner differences. However, these studies used structural MRI data acquired from different scanners for training and testing purposes separately and did not pool them together. It is, therefore, not clear how scanner-related differences affect the diagnostic performance and the robustness of the results, if data pooling and no correction for scanner effects is employed.

Apart from scanner-related differences, differences in the assessments of psychopathology and transition criteria may have confounded the present analysis. It

is important to note that subjects in the FePsy study had significantly shorter transition intervals than subjects in the EHRS. This is, however, reflective of the at-risk mental state, in which individuals already exhibit putative transient or sub-threshold symptoms and are considered at an imminent risk for conversion. In contrast, HR-T subjects from the EHRS sample had their baseline MRI scan approximately 2.5 years before making a transition to psychosis. This heterogeneity in conversion times, from the baseline MRI scanning until conversion to psychosis was corroborated, might have been reflected at the neuroanatomical level, with divergent pathophysiological processes leading up to psychosis, which may have in turn confounded our results. Additionally, the heterogeneity in terms of psychopathological indices, clinical and neurocognitive measures did not enable the integration of sMRI data with other variables to test the classification performance. Further, one cannot exclude the possibility that the use of anti-psychotic medication in 7 subjects in the pooled data set may have confounded our classification results.

7.6 Conclusions

This preliminary investigation extends previous investigation of the EHRS sample by pooling these data and those of the FePsy study, which involved high-risk subjects, identified employing different high-risk paradigms, recruited using different assessment criteria and examined using different MRI scanners. We provide preliminary evidence for the identification of a sMRI biomarker that exists across independent high-risk cohorts. However, additional work is required in order to clarify the effect of center-related differences on the diagnostic procedure and more importantly to control for effects pertaining to scanner-related differences.

CHAPTER 8

General Discussion

8.1 Introduction

The work described in this thesis has used machine learning in order to examine the diagnostic performance of MRI data, neurocognitive and behavioural variables in predicting transition to psychosis in individuals at high risk for the disorder. The classification method implemented in this thesis was used to classify subjects identified as high-risk for psychosis either due to familial or clinical reasons, derived from the Edinburgh High Risk Study and the Basel, FePsy study accordingly.

As described in Chapter 1, schizophrenia is a highly complex brain disorder, expressed with heterogeneous constellations of symptoms across patients and variable clinical course. Diagnosis of schizophrenia is almost entirely based on self-reported symptomatology and observation of behavioural signs and therefore is subject to the patient's ability to collaborate and the clinical expertise and experience of the clinician. To avoid bias in clinical decision-making, objective biomarkers are pressingly needed in order to aid diagnosis of schizophrenia and support treatment strategies where needed. The potential of a biomarker, however, depends on its predictive capacity at the individual level (Phillips et al. 2006).

It is likely that schizophrenia is associated with neuroanatomical abnormalities that are not limited to a circumscribed set of a few brain regions but rather span over spatially distributed, and possibly intercorrelated, networks. Structural MRI studies have significantly improved our understanding of the underlying pathophysiology of the disorder by systematically studying group-level differences in brain tissue but nevertheless have so far failed at impacting clinical practice. Mass-univariate methods of analyses such as voxel-based morphometry (VBM) examine differences between the experimental groups in localized brain regions at a mean-level. More recently used pattern recognition-based methods on the other hand, look for differences in patterns within the brain that could optimally discriminate between the study groups and furthermore have the potential of making inferences about new subjects at a subject-level. Additionally, by virtue of their multivariate nature, machine learning approaches allow the integration of localized differences over the

whole brain that would otherwise be too weak or too variable to detect if considered on their own.

As described in detail in previous chapters, subjects presented with a family history of schizophrenia and/or sub-threshold, prodromal psychotic symptoms are considered at an elevated risk for developing schizophrenia and psychosis-related disorders. Studies examining subjects at high risk for developing schizophrenia and psychosis-related disorders have stimulated two HR paradigms; i) the clinical and ii) the genetic or familial high risk for psychosis. Magnetic resonance imaging-based studies of high-risk for psychosis cohorts can serve to delineate the extent to which neuroanatomical alterations form part of an underlying risk and vulnerability for the development of psychosis or are exclusively associated with the illness itself. To further disentangle the issue of whether (and which) specific volumetric abnormalities are associated with transition to psychosis, researchers have begun to study those HR subjects that develop psychosis against those HR individuals that do not. Identifying the neuroanatomical alterations that could distinguish at baseline those HR subjects that will later develop psychosis from those who will not would have significant implications in further elucidating the pathophysiology and aetiology of the disorder and possibly providing an opportunity for early intervention with the aim of either relieving the disorder's symptom burden or even prevent its onset.

This thesis, therefore, aimed:

1) To identify neuroanatomical-based markers of transition to schizophrenia in subjects at high familial risk drawn from the Edinburgh High Risk Study, using an implementation of Support Vector Machine that includes subject classification and also feature selection. In addition to baseline structural MRI data, other neurocognitive and behavioural variables were included in the analysis with the aim of testing the classification performance and also replicating previous findings in the EHRS literature (Johnstone et al. 2005)

2) To examine whether the implemented classifier would be able to generalize to clinical high-risk cohorts, using neuroanatomical data from the FePsy study.

3) And finally, in a strictly exploratory and preliminary analysis, to test the diagnostic performance of our classifier by pooling structural MRI data of the EHRS and the FePsy study together, despite the differences in the recruitment, assessment and imaging procedures.

To do that, data from high-risk subjects that later developed psychosis were contrasted to data from high-risk subjects that did not make the transition. There were no specific hypotheses regarding the number or which brain regions would constitute the pattern based on which the study groups would be optimally discriminated against, although it was expected that previous findings, especially in the EHRS literature might be replicated. However, based on the multivariate nature of the SVM method, I expected that a host of brain structures across cortical and subcortical brain areas would comprise the decision function in each case. Additionally, it was hypothesized that the diagnostic ability of the classification method would be increased when baseline clinical and behavioural variables were combined with neuroanatomical data of the EHRS groups.

Since each of the previous chapters contains a separate discussion section, the aim of this chapter is to offer a more general discussion integrating the main results from the previous chapters, discussing limitations of the current findings, and suggesting possible steps and ideas for future work.

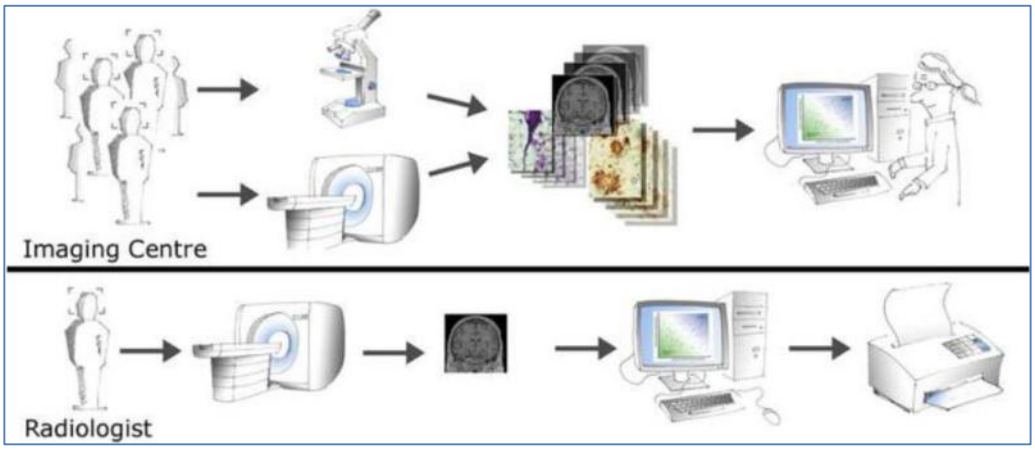


Figure 8.1 Potential clinical setting for a pattern recognition-based system. Neuroimaging data are acquired in a specialized clinic and can then be combined with other variables. The classifier is trained using these features. Gold standard diagnosis can be established using i.e. post-mortem histological examination. The trained classifier could then, be used at another research/clinical site and applied directly to a new patient. Classification results for this patient could then be fed to the clinicial for appropriate clinical management. (Figure adjusted from Klöppel et al. 2012)

8.2 Summary of main results

In this thesis, it was shown that it is feasible to make an early diagnosis of psychosis, and primarily schizophrenia, in subjects at high clinical or familial risk for the disorder, by using baseline neuroanatomical information. The findings presented here suggest that high-risk subjects who will subsequently develop psychosis may be already distinguished at baseline, above chance levels and before illness onset, from those high-risk individuals that will not make the transition. The diagnostic performance of the SVM classifier was mainly tested using cross-validation and also by classifying an independent cohort of high-risk subjects drawn from the EHRS

study that did not develop schizophrenia, yet showed psychotic symptoms at follow-up (Results sections of Chapters 5 and 7). Table 8.1 provides a comparison of the classification results from the experimental Chapters 5, 6 and 7 of this thesis.

The high-dimensional discriminative morphological map comparing high risk subjects with later transition to psychosis and high-risk subjects without illness transition revealed a complex pattern of regional volumetric abnormalities affecting grey matter across most lobar brain structures and the cerebellum. Although, the discriminating decision maps were different for each analysis of the EHRS, the FePsy and the pooled datasets, there were however common brain regions that comprised the discrimination maps in all cases-analyses. The brain regions that contributed the highest in discriminating between the study groups in all cases included superior and inferior frontal and superior temporal lobar structures, parts of the cingulate gyrus and the cerebellum, all which have been consistently reported as important in the pathophysiology of schizophrenia.

Table 8.1 Comparison of classification results from Chapters 5, 6 and 7.

	sMRI data alone		sMRI+ clinical data	
	N/N	BAC (%)	N/N	BAC (%)
HR[ill] vs HR[symp]	17/17	88.2	17/17	94.1
ARMS-T vs ARMS-NT	16/19	74.2	-	-
HR-T vs HR-NT	33/36	82.6	-	-

The number of subjects in each diagnostic group (N/N) and the balanced accuracy are shown for the classifications based on baseline structural MRI data alone, and the combination of baseline MRI with clinical variables (just for subjects in the EHRS). Note: HR[ill] and HR[symp] refer to subjects from the EHRS, ARMS-T and ARMS-NT refer to subjects drawn from the FePsy study and HR-T, HR-NT refer to the subjects from the pooled dataset, in accordance with Chapters 5, 6 and 7.

Despite the very encouraging results presented in this thesis, it is my personal opinion that neuroimaging-based pattern recognition methods have a long way to go before making an impact in the current diagnostic and prognostic clinical context of schizophrenia. I believe that the research and imaging community is not yet in place to speak about imaging-based biomarkers in schizophrenia research but rather about neuroanatomical patterns that can differentiate one diagnostic group (in this case schizophrenia, or converters to schizophrenia) from another (healthy controls or non-converters). Pattern recognition-based methods implemented to date are still rather crude to accommodate the idea of identifying neuroanatomical biomarkers, that warrant large effect sizes and therefore larger data samples, replication to independent cohorts, consistency in classification performances across cohorts, data samples that are representative of routine clinical cases (and therefore eliminate the associated selection bias), a meaningful and widely acknowledged interpretation of the classification results (i.e. classifiers weights) and effective ways and methods of dealing with confounding variables such as medication effects or comorbidity. These issues have yet to be examined and resolved, and could be probably best dealt with in large muti-site, longitudinal studies (please see section 8.4 for a more detailed discussion of the challenges and future directions of imaging-based machine learning)

The ultimate goal for imaging-based machine learning is, however, to develop tools that could aid clinicians in everyday routine care and management. That is, machine learning-based systems, like the one depicted in Figure 8.1, that will be able to combine together various imaging, clinical and neuropsychological data and provide results for an individual that could be fed back quickly to the clinician in a way that would be helpful for clinical management. The major challenge for neuroimaging-based pattern classification in significantly impacting clinical practice is, however, in the early diagnosis of individuals at an imminent risk of developing schizophrenia and other psychosis-related disorders and in refining prognostic distinctions; that is questions of whether an individual will develop schizophrenia over bipolar disorder or whether he/she will develop unipolar depression over bipolar disorder and ultimately whether the patient will benefit from one treatment over another. These are the clinical questions that could bear added benefit in clinical routine and would

likely help clinicians in a way that will move the field of clinical psychiatry forward, alongside the majority of other medical specialties where biological and clinical tests are routinely used to guide clinical decision-making and care.

8.3 Strengths and Limitations

One of the main strengths of the present thesis is the ability to include prospective studies of both familial and clinical high-risk cohorts. Additionally, the high-risk subjects from the EHRS cohort were anti-psychotic naïve at the time of study inclusion, which means that our results were completely unconfounded by medication effects.

In this thesis, Support Vector Machine was chosen to implement the classification task. Among the machine learning algorithms, Support Vector Machine is one of the most frequently used multivariate approaches, mainly on account of its good classification performance and ability to deal effectively with high-dimensional data. The latter characteristic is due to the fact that in determining the optimal separating hyperplane the SVM only takes into account data samples that are closely located to it (these samples are called support vectors) in the feature space. In that way, the SVM classifier inherently focuses on subtle between-group morphological differences and not on gross differences that might be more easily identifiable. Additionally, as all pattern-classification methods do, the SVM enabled the prediction of illness onset at an individual level, which has added potential for clinical translation.

In order to identify the most informative features in the classification process and possibly increase the diagnostic performance of the classifier, recursive feature elimination (RFE) was implemented in a nested cross-validation scheme. The RFE algorithm is a well-validated technique that iteratively removes the most redundant and least informative features while identifying the most discriminative ones. A

significant number of MRI-based pattern recognition studies have used RFE for the identification of the most discriminative features in the feature space and/or for improving the generalization performance (Ecker et al. 2010, Zacharaki et al. 2009, Marquand et al. 2011, Mourao-Miranda et al. 2012b, De Martino et al. 2008, Fan et al. 2007).

Another important strength of the current thesis is the validation of the classifier using data from an independent high-risk cohort; a step that is crucial in the development and verification of imaging-based biomarkers in clinical settings (Fu & Costafreda 2013). Pooling together data from different research sites is additionally beneficial in examining whether prediction is robust enough to differences in prevalence, recruitment and image acquisition strategies.

As stated before, taken together this thesis leads to the conclusion that machine learning can significantly impact the identification of MRI-based biomarkers for the early prediction of psychosis in high-risk cohorts. That is, even at the early stages of psychosis, subjects who later develop schizophrenia demonstrate already at baseline distinct patterns of grey matter volume differences that differentiate them from high-risk subjects that do not develop the disorder. However, many challenges and limitations remain. The limitations encountered here arise either as a result of limitations inherent in the datasets or limitations attributed to the chosen preprocessing or analysis methodologies.

The relatively modest sample size, especially in the cases where subjects from the EHRS and the FePsy study were considered separately (Chapters 5 and 6 accordingly), is one of the main limitations of this study, which is also observed in most neuroimaging-based studies in this field. As a result of the limited sample size and the recruitment criteria followed, the study cohorts included only a limited representation of the whole spectrum of clinical and psychopathological manifestations of the high-risk state, which is not always the case in standard clinical practice. Comorbid substance abuse and comorbidities with other disorders such as depression or anxiety were part of the exclusion criteria, which is not a realistic mirroring of the cases encountered in health services.

Both in the EHRS and the FePsy study, more males than females developed schizophrenia, possibly confounding our classification results. The construction of different classifiers for males and females, as presented in the work of Fan et al. 2007, was not feasible due to the already limited number of cases in the HR-T groups. The finding of the higher male morbidity may be, however, attributed to the higher incidence of schizophrenia in males than females (Aleman et al. 2003).

Antipsychotics and other medications have been administered to some high-risk subjects in the FePsy study, some time prior to MRI scanning, and therefore one cannot rule out the possibility that medication effects may have contributed to the observed classification results (Chapter 6). Interestingly enough, most of the subjects that received antipsychotic medication were misclassified by the classifier possibly suggesting a divergent pathophysiology at the neuroanatomical level.

Another potential limitation is the use of a 1.0T scanner in the EHRS samples that compared to most widely used 1.5T scanners (as in the FePsy study) might have failed to provide enough spatial detail to characterize subtle structural abnormalities.

Regarding the classification task, one limitation of the SVM method and other machine learning methods in general, is that the features comprising the optimal decision function (that is often illustrated as a weighted brain map) are not easily interpretable, mainly due to their multivariate nature. As stated before, the brain regions constituting the discrimination map all contribute to the classification function (with a certain weight that indicates the degree of the contribution) and it is not possible to make inferences about localized alterations.

Additionally, while the leave-one-out cross-validation is a well-validated and highly used technique in evaluating the performance of the classifier due to its almost unbiased estimation of the true error rate, it underestimates variance due to the repeated use of any subset of data in most training sets. The ideal scenario of splitting the dataset into non-overlapping sets of training data and test data instances could be realized in large, multisite cohort studies.

Finally, one of the most important limitations with the data pooling is the

introduction of systematic scanner-related differences that may have confounded both the preprocessing of the baseline scans and downstream from there to the classification task. Future studies would be required that would account for scanner-related differences between sites and tackle the issue of intensity standardization at the initial stage, before any preprocessing, by matching histograms for every tissue type (i.e. grey matter, white matter, cerebrospinal fluid etc.).

8.4 Future work

As seen and discussed in depth in previous chapters of this thesis, brain imaging and more specifically magnetic resonance imaging has contributed much in our current understanding of the pathophysiology of schizophrenia and psychosis-related disorders. However, it has so far made little impact in routine clinical practice where a MRI scan is usually ordered in order to exclude organic diseases, such as a brain tumour or brain haemorrhage and not to diagnose or inform targeted theurapeutic plans. Therefore, one of the basic aims for clinical psychiatry to move into the future is to develop objective, biologically based markers (*biomarkers*) that could map psychiatric disorders to brain structure and function. A careful investigation of the feasibility and main challenges in translating imaging research findings into routine clinical practice is necessary in order to pave the way for the development of clinically useful biomarkers for establishing diagnosis and illness course or treatment outcome in schizophrenia.

In order for a diagnostic biomarker to be valid, it needs to attain certain levels of sensitivity and specificity. A perfect biomarker, with 100% sensitivity and 100% specificity, would detect only true positives and no false negatives and thus would accurately reflect the prevalence of schizophrenia in the true population. However, in practice such diagnostic levels are not always possible and thus the norm is to develop biomarkers that can minimize the error for false positives and false negatives as much as possible. Addtionally, the lack of consistency in the classification

performances observed in current diagnostic studies of schizophrenia and lack of replication studies that would confirm previous classification results hinder the identification of diagnostic biomarkers.

The development of clinically relevant biomarkers for shcizophrenia is also hindered by the lack of a diagnostic gold standard for the disease. In traditional medicine, a biological test, such as a blood test, often suffices for a definitive pathological diagnosis. For Alzheimer's disease, neuropathological confirmation can be achieved by post-mortem histopathological examination. On the contrary, there is no definite examination to confirm diagnosis of schizophrenia that leads directly back to brain structure (or function), and currently the closest one can get to a 'gold-standard' diagnosis for schizophrenia is the DSM/ICD criteria, which are empirically based and often limited by the patient's willingness to cooperate and the clinical expertise and acumen of the clinician conducting the interview.

Additionally, linking back to the importance of specificity in the development of a biomarker, most imaging-based studies have examined and contrasted schizophrenia patients to perfectly healthy, normal participants (see Chapter 3). In clinical practice, however, it is rarely the case that a clinician would wonder about whether to categorize (or label) an individual as a patient or as a healthy control and more often the question raised involves whether this patient has schizophrenia, or bipolar disorder etc. This is not to say that imaging findings contrasting schizophrenia patients to healthy controls were conducted in vain since these have provided the research community with invaluable information and insight into the pathophysiology of the disease, but to highlight the importance of including more pragmatic scenarios that could have an added benefit to clinical care.

The differential diagnosis of schizophrenia from other psychosis-spectrum disorders such as bipolar disorder and mood disorders such as depression presents a major clinical challenge. Recent structural and functional MRI-based multivariate pattern recognition studies have achieved very promising classification performances in distinguishing schizophrenia patients from patients with bipolar disorder (Costafreda et al. 2009, Costafreda et al. 2011, Calhoun et al. 2008, Schnack et al. 2014, Bansal et al. 2012) or major depression (Koutsouleris et al. 2015a, Bansal et al. 2012).

However, future studies will be needed that would comprise of larger cohorts and possibly integrate various imaging modalities in order to unravel the neuroanatomical and neurophysiological correlates that differentiate schizophrenia from other psychosis- or mood-related disorders (Lawrie et al. 2011). This research direction will hopefully shed light into the question of whether these nosological entities are characterized by distinct and/or non-overlapping pathopshysiological abnormalities and help identify patterns of brain structure that are unique to schizophrenia. Unsupervised or semi-supervised machine learning methods could aid in that direction by grouping together (into clusters) subsets of patients that share similar neuroanatomical and/or clinical profiles and then possibly exploit supervised pattern recognition to examine the diagnostic differentiation between the clusters.

Imaging findings need to be replicated in independent study cohorts in order to ensure the validity of imaging-based biomarkers in their way towards incorporation to routine clinical care. Currently, most researchers tried to develop diagnostic tools using one cohort of patients at a time while only a few have tested their classifiers using independent cohorts (Nieuwenhuis et al. 2012, Schnack et al. 2014, Kawasaki et al. 2007) with classification performance being lower on the independent sets, and ranging from 70% to 80% accuracy. Despite little effort being done, machine-learning techniques could really contribute to the effort of replicating imaging findings across research sites because different teams could train classifiers using their own data sets and then once the classification functions are found, they could be shared among collaborating research sites in the context of creating a classifiers' repository.

Additionally, incorporating more 'pragmatic' cases for both patient and healthy control cohorts is another major clinical challenge. Apart from the limiting case of contrasting schizophrenia patients with healthy control individuals presented above, it should be noted that most control participants recruited include perfectly healthy individuals ('hyper-normal' cases) while excluding any individual presenting with any substance abuse or even family history of psychiatric conditions. Similarly, schizophrenia cohorts often exclude patients with any comorbid psychiatric disorder, or substance abuse, which is not the norm for cases of schizophrenia presenting in

everyday clinical practice. The need for more realistic representation of schizophrenia and healthy participants is in alignment with the need for larger data samples; a well-known hindrance in most research studies in terms of recruiting procedures and findings' robustness. Multi-site studies could address the need for large data samples and additionally ensure the sufficient phenotypic and demographic representation of the under-study population. Prospective, longitudinal multi-site studies for the early diagnosis and management of early and first episode psychosis, such as the PSYSCAN and PRONIA projects have recently been launched with the aim of developing multimodal biomarkers for the prediction of onset, course and outcome of psychosis. However, multi-site initiatives warrant further calibration and standardization efforts in order to ensure the compatibility and reproducibility of imaging data across sites. Standardization procedures in the pre-processing and/or analysis pipelines as the ones suggested by the ENIGMA consortium could be used in order to reduce variability in findings. Test-retest reliability studies and novel histogram matching techniques could be used to account for site differences and guarantee the quality and standardization between follow-up scanning.

To further understand which structural brain alterations take place before psychosis onset and/or are responsible for the transition from a state of heightened risk to psychosis, further longitudinal studies are needed that could perhaps involve follow-up scanning before the onset of psychosis and while the subject still meets HR criteria. This might clarify the nature and extent of longitudinal changes that predate the onset of illness that are possibly not confounded by effects of antipsychotic medication and illness-specific alterations. Additionally, longitudinal studies of HR subjects with low probability of conversion to psychosis (Smieskova et al. 2012) and/or those HR subjects without transition (Addington et al. 2011) could reveal protective, resilience factors that would be clinically relevant in early intervention approaches. In the same context, longitudinal imaging–based studies in healthy control populations would help the research community in mapping normal, brain developmental trajectories in an effort to elucidate normal structural and functional variability before characterizing deviant, illness-specific trajectories.

Within the context of technical challenges, probabilistic machine learning might also be a promising tool in neuroimaging-based schizophrenia research. More specifically, probabilistic classification approaches such as Gausian Process classifier (GCP) (Rasmussen & Williams, 2006) and Relevance vector machine, can be used to quantify a degree of uncertainty in the prediction and could thus be applied in the context of predicting transition to psychosis or future clinical outcome, indicating for example a percentage of confidence for classification into one group or another (e.g. 75% risk transition to schizophrenia and 25% not making a transition.). There are several examples of studies that have used probabilistic classifiers employing structural MRI data (Tognin et al. 2013), neurophysiological imaging data (either PET or fMRI) (Marquand et al. 2010, Phillips et al. 2011, Mourao-Miranda et al. 2012b) and a multimodal approach where structural and functional imaging data together with genotype information are combined in the same probabilistic learning framework (Young et al. 2013).

It is a priority that future studies also address the challenge and opportunity of fusing neuroimaging data from various imaging modalities, along with genetic and clinical information, that seem likely to interact in determining the development and outcome in schizophrenia (Lawrie et al. 2011). It would be reasonable to assume that the introduction of neurocognitive and other clinical measures could possibly enhance diagnostic power of the classifier. Just as a clinician takes a detailed report of symptoms and other clinical measures to diagnose a patient with schizophrenia, so might the integration of symptom severity measures and other neurocognitive scores, along with MRI scans aid to the classification process. Early studies have already shown that classification performance might well be improved (Yang et al. 2010, Sui et al. 2012), as in Karageorgiou et al. (2011) where the authors observed a 92% accuracy in classifying recent-onset schizophrenia when structural MRI data and neuropsychological variables (NP) were combined than when employing either quantitative measure alone (86.7% when only NP data were used and 70.7% with sMRI data alone). Other neuroimaging technologies such as arterial spin labeling (ASL) perfusion MRI and diffusion tensor imaging (DTI) have shown very promising leads in unraveling the neurobiological substrate of several psychiatric and neurological disorders (Pinkham et al. 2011, Sussmann et al. 2009, van Essen et al.

2012), and might as well be combined with MRI methods in schizophrenia research. The interpretability of such data is not, however, necessarily straightforward and, as a general rule, each additional diagnostic variable increases sensitivity at the expense of specificity. It is overdue, though, that combined features, such as symptoms, duration of illness, genomics and proteomics along with various brain imaging modalities are incorporated into imaging and other evaluations in clinical research studies, with the scope of making more reliable and objective judgements about the diagnosis of schizophrenia and to classifying patients into more homogenous subgroups (Lawrie et al. 2008).

In conclusion, despite the fact that the medical and research community is currently far away from utilizing automated neuroimaging-based classification methods to diagnose schizophrenia and possibly predict its onset, clinical outcome and treatment response, in the long run MRI-based machine learning could significantly impact everyday clinical practice. Indeed, these methods could add value to current diagnostic procedures and aid clinicians in a complementary fashion in order to reach diagnosis and possibly inform appropriate treatment strategies after evaluating the risk of transition to psychosis in subjects at an increased risk. Clinical decisions based on imaging data could, thus facilitate the stratification of patient care and potentially reduce health-care costs. Further studies are, however, necessary that would involve large sample sizes, possibly using multicenter, longitudinal designs that could additionally examine the classification performance of combined data types, such as multi-modal imaging data, clinical, neurocognitive and genetic variables.

REFERENCES

A

Abdulkadir A, Mortamet B, Vemuri P, Jack CR, Krueger G, Klöppel S. (2011) Alzheimer's Disease Neuroimaging Initiative. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. *Neuroimage*, 58(3):785-792.

Acosta-Cabronero J, Williams GB, Pereira JMS, Pengas G, Nestor PJ. (2008) The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. *Neuroimage*, 39(4):1654-1665.

Addington D, Addington J, Patten S. (1996) Gender and affect in schizophrenia. *Can J Psychiatry* ,41(5):265-268.

Addington J, Cornblatt BA, Cadenhead KS, Cannon TD, McGlashan TH, Perkins DO, Seidman LJ, Tsuang MT, Walker EF, Woods SW, Heinssen R. (2011) At Clinical High Risk for Psychosis: Outcome for Nonconverters. *The American journal of psychiatry*, 168(8):800-805. doi:10.1176/appi.ajp.2011.10081191.

Akbani R, Kwek S, Japkowicz N. (2004) Applying support vector machines to imbalanced datasets, in Proceedings of the 15th European Conference on Machine Learning, 39–50.

Aleman A, Kahn RS, Selten J-P. (2003) Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch. Gen. Psychiatry*, 60(6):565-571.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th edn (DSM-V). (2013) Washington, DC: American Psychiatric Press.

Amminger GP, Schäfer MR, Papageorgiou K, Klier CM, Cotton SM, Harrigan SM, Mackinnon A, McGorry PD, Berger GE. (2010) Long-chain omega-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. *Arch. Gen. Psychiatry*, 67(2):146-154.

Anderson A, Dinov ID, Sherin JE, Quintana J, Yuille AL, Cohen MS. (2010) Classification of spatially unaligned fMRI scans. *Neuroimage*, 49(3):2509-2519.

Andreasen NC The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and theoretical foundations. (1989) *Br J Psychiatry Suppl*, 7:49 –58.

Andreasen NC, Arndt S, Alliger R, Miller D, Flaum M. (1995) Symptoms of schizophrenia. Methods, meanings, and mechanisms. *Arch. Gen. Psychiatry*, 52(5):341-351.

Andreasen NC, Carpenter WT, Kane JM, Lasser RA, Marder SR, Weinberger DR. (2005) Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry*, 162(3):441-449. doi:10.1176/appi.ajp.162.3.441.

Arseneault L, Cannon M, Witton J, Murray RM. (2004) Causal association between cannabis and psychosis: examination of the evidence. *Br J Psychiatry*, 184:110-117.

Ashburner J, Friston KJ. (1997) Multimodal image coregistration and partitioning – A unified framework. *NeuroImage*, 6:209–217.

Ashburner J, Friston KJ. (2000) Voxel-based morphometry--the methods. *Neuroimage*, 11(6 Pt 1):805-821. doi:10.1006/nimg.2000.0582.

Ashburner J, Friston KJ. (2005) Unified segmentation. *Neuroimage*, 26(3):839-851. doi:10.1016/j.neuroimage.2005.02.018.

Ashtari M. (1994) Increase in caudate nuclei volumes of first-episode schizophrenic patients taking antipsychotic drugs. *Am J Psychiatry*, 151(10):1430-1436.

B

Baiano M, David A, Versace A, Churchill R, Balestrieri M, Brambilla P. (2007) Anterior cingulate volumes in schizophrenia: a systematic review and a meta-analysis of MRI studies. *Schizophr. Res*, 93(1-3):1-12. doi:10.1016/j.schres.2007.02.012.

Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M, Peteron BS. (2012) Anatomical Brain Images Alone Can Accurately Diagnose Chronic Neuropsychiatric Illnesses. PLoS ONE 7(12): e50698. doi:10.1371/journal.pone.0050698

Barnes TRE, Mutsatsa SH, Hutton SB, Watt HC, Joyce EM. (2006) Comorbid substance use and age at onset of schizophrenia. *Br J Psychiatry*, 188:237-242. doi:10.1192/bjp.bp.104.007237.

Bennett CM, Wolford GL, and Miller MB. (2009). The principled control of false positives in neuroimaging *Soc Cogn Affect Neurosci 4 (4): 417-422*

Bleuler E. (1911) Dementia Praecox: Or the Group of Schizophrenias. New York: International Universities Press.

Bois C, Whalley HC, McIntosh AM, Lawrie SM. (2015) Structural magnetic resonance imaging markers of susceptibility and transition to schizophrenia: a review of familial and clinical high risk population studies. *J. Psychopharmacol,* 29(2):144-154. doi:10.1177/0269881114541015.

Bora E, Fornito A, Radua J, Walterfang M, Seal M, Wood SJ, Yücel M, Velakoulis D, Pantelis C. (2011) Neuroanatomical abnormalities in schizophrenia: a multimodal voxelwise meta-analysis and meta-regression analysis. *Schizophr. Res,*127(1-3):46-57. doi:10.1016/j.schres.2010.12.020.

Borgwardt SJ, McGuire PK, Aston J, Berger G, Dazzan P, Gschwandtner U, Pflüger M, D'Souza M, Radue E-W, Riecher-Rössler A. (2007a) Structural brain abnormalities in individuals with an at-risk mental state who later develop psychosis. Br J Psychiatry Suppl, 51:s69-75. doi:10.1192/bjp.191.51.s69.

Borgwardt SJ, Riecher-Rössler A, Dazzan P, Chitnis X, Aston J, Drewe M, Gschwandtner U, Haller S, Pflüger M, Rechsteiner E, D'Souza M, Stieglitz R-D, Radü E-W, McGuire PK. (2007b) Regional gray matter volume abnormalities in the at risk mental state. Biol. Psychiatry, 61(10):1148-1156. doi:10.1016/j.biopsych.2006.08.009.

Borgwardt SJ, McGuire PK, Aston J, Gschwandtner U, Pflüger MO, Stieglitz R-D, Radue E-W, Riecher-Rössler A. (2008) Reductions in frontal, temporal and parietal volume associated with the onset of psychosis. *Schizophrenia Research*, 106(2–3):108-114. doi:10.1016/j.schres.2008.08.007.

Borgwardt S, Koutsouleris N, Aston J, Studerus E, Smieskova R, Riecher-Rössler A, Meisenzahl EM. (2013) Distinguishing prodromal from first-episode psychosis using neuroanatomical single-subject pattern recognition. *Schizophr Bull*, 39(5):1105-1114. doi:10.1093/schbul/sbs095.

Borgwardt S, McGuire P, Fusar-Poli P. (2011) Gray matters!--mapping the transition to psychosis. *Schizophr. Res,* 133(1-3):63-67. doi:10.1016/j.schres.2011.08.021.

Borgwardt S, Fusar-Poli P. (2012) Third-generation neuroimaging in early schizophrenia: translating research evidence into clinical utility. *Br J Psychiatry*, 200(4):270-272. doi:10.1192/bjp.bp.111.103234.

Boser, B.E., Guyon, I.M., Vapnik, V.N. (1992) A Training Algorithm for Optimal Margin Classifiers, in: Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory. ACM Press, pp. 144–152.

Boydell J., Murray R. (2003) Urbanization, migration and risk of schizophrenia. In: Murray RM, Jones PB, Susser E, van Os J, Cannon M. The Epidemiology of Schizophrenia. Cambridge: Cambridge University Press, 49–67.

Bray S, Chang C, Hoeft F. (2009) Applications of multivariate pattern classification analyses in developmental neuroimaging of healthy and clinical populations. *Front Hum Neurosci*, 3:32. doi:10.3389/neuro.09.032.2009.

Brett M, Penny W and Kiebel S. (2003) An Introduction to Random Field Theory. *SPM Note.*

Brewer WJ, Francey SM, Wood SJ, Jackson HJ, Pantelis C, Phillips LJ, Yung AR, Anderson VA, McGorry PD. (2005) Memory impairments identified in people at ultra-high risk for psychosis who later develop first-episode psychosis. *Am J Psychiatry,* 162(1):71-78. doi:10.1176/appi.ajp.162.1.71.

Brewer WJ, Wood SJ, Phillips LJ, Francey SM, Pantelis C, Yung AR, Cornblatt B, McGorry PD. (2006) Generalized and specific cognitive performance in clinical high-risk cohorts: a review highlighting potential vulnerability markers for psychosis. *Schizophr Bull*, 32(3):538-555. doi:10.1093/schbul/sbj077.

Buchanan RW, Carpenter WT. (1994) Domains of psychopathology: an approach to the reduction of heterogeneity in schizophrenia. *J. Nerv. Ment. Dis,* 182(4):193-204.

Burges CJC. (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121–167.

Byrne M, Hodges A, Grant E, Owens DC, Johnstone EC. (1999) Neuropsychological

assessment of young people at high genetic risk for developing schizophrenia compared with controls: preliminary findings of the Edinburgh High Risk Study (EHRS). *Psychol Med*, 29(5):1161-1173.

Byrne M, Clafferty BA, Cosway R, Grant E, Hodges A, Whalley HC, Lawrie SM, Owens DGC, Johnstone EC. (2003) Neuropsychology, genetic liability, and psychotic symptoms in those at high risk of schizophrenia. *J Abnorm Psychol*, 112(1):38-48.

Byrne M, Agerbo E, Bennedsen B, Eaton WW, Mortensen PB. (2007) Obstetric conditions and risk of first admission with schizophrenia: a Danish national register based study. *Schizophr. Res*, 97(1-3):51-59. doi:10.1016/j.schres.2007.07.018.

C

Cahn W, Hulshoff Pol HE, Lems E, van Haren N, Schnack HG, van der Linden JA, Schothorst PF, van Engeland H, Kahn RS. (2002) Brain volume changes in first-episode schizophrenia: a 1-year follow-up study. *Arch. Gen. Psychiatry*, 59(11):1002-1010.

Calhoun VD, Adali T, Kiehl KA, Astur R, Pekar JJ, Pearlson GD. (2006) A method for multitask fMRI data fusion applied to schizophrenia. *Hum Brain Mapp*, 27(7):598-610. doi:10.1002/hbm.20204.

Calhoun VD, Maciejewski PK, Pearlson GD, Kiehl KA. (2008) Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Hum Brain Mapp*, 29(11):1265-1275. doi:10.1002/hbm.20463.

Candilis PJ. (2003) Early intervention in schizophrenia: three frameworks for guiding ethical inquiry. *Psychopharmacology*, 171(1):75-80. doi:10.1007/s00213-003-1412-3.

Cannon TD, Mednick SA. (1993) The schizophrenia high-risk project in Copenhagen: three decades of progress. *Acta Psychiatr Scand Suppl*, 370:33-47.

Cannon TD. (2005) Clinical and genetic high-risk strategies in understanding vulnerability to psychosis. *Schizophr. Res*, 79(1):35-44. doi:10.1016/j.schres.2005.06.014.

Cantor-Graae E, Selten J-P. (2005) Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry*, 162(1):12-24. doi:10.1176/appi.ajp.162.1.12.

Carney CP, Jones L, Woolson RF. (2006) Medical comorbidity in women and men with schizophrenia: a population-based controlled study. *J Gen Intern Med*, 21(11):1133-1137. doi:10.1111/j.1525-1497.2006.00563.x.

Castle DJ, Murray RM. (1991) The neurodevelopmental basis of sex differences in schizophrenia. *Psychol Med*, 21(3):565-575.

Castle D, Sham P, Murray R. (1998) Differences in distribution of ages of onset in males and females with schizophrenia. *Schizophr. Res*, 33(3):179-183.

Castro E, Martínez-Ramón M, Pearlson G, Sui J, Calhoun VD. (2011) Characterization of groups using composite kernels and multi-source fMRI analysis data: application to schizophrenia. *Neuroimage*, 58(2):526-536. doi:10.1016/j.neuroimage.2011.06.044.

Chakos MH, Lieberman JA, Bilder RM, Borenstein M, Lerner G, Bogerts B, Wu H, Kinon B, Ashtari M. (1994) Increase in caudate nuclei volumes of first-episode schizophrenic patients taking antipsychotic drugs. *Am J Psychiatry*, 151(10):1430-1436.

Chakos MH, Lieberman JA, Alvir J, Bilder R, Ashtari M. (1995) Caudate nuclei volumes in schizophrenic patients treated with typical antipsychotics or clozapine. *Lancet*, 345(8947):456-457.

Chang CC, Lin CJ. 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27:1–27:27.

Chua SE, Lam IWS, Tai K-S, Cheung C, Tang W-N, Chen EYH, Lee PWH, Chan F-L, Lieh-Mak F, McKenna PJ. (2003) Brain morphological abnormality in schizophrenia is independent of country of origin. *Acta Psychiatr Scand*, 108(4):269-275.

Ciompi L. (1980) The natural history of schizophrenia in the long term. *Br J Psychiatry*, 136:413-420.

Clark KA, Woods RP, Rottenber DA, Toga AW, Mazziotta JC. (2006) Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *NeuroImage*, 29:185–202.

Compton, M. T., & Walker, E. F. (2009). Physical manifestations of neurodevelopmental disruption: are minor physical anomalies part of the syndrome of schizophrenia? *Schizophrenia Bulletin*, *35*(2), 425–436.

Cooper D, Barker V, Radua J, Fusar-Poli P, Lawrie SM. (2014) Multimodal voxel-based meta-analysis of structural and functional magnetic resonance imaging studies in those at elevated genetic risk of developing schizophrenia. *Psychiatry Res*, 221(1):69-77. doi:10.1016/j.pscychresns.2013.07.008.

Cornblatt B, Lencz T, Obuchowski M. (2002) The schizophrenia prodrome: treatment and high-risk perspectives. *Schizophr. Res*, 54(1-2):177-186.

Costafreda SG, Fu CHY, Picchioni M, Kane F, McDonald C, Prata DP, Kalidindi S, Walshe M, Curtis V, Bramon E, Kravariti E, Marshall N, Toulopoulou T, Barker GJ, David AS, Brammer MJ, Murray RM, McGuire PK. (2009) Increased inferior frontal activation during word generation: a marker of genetic risk for schizophrenia but not bipolar disorder? *Hum Brain Mapp*, 30(10):3287-3298. doi:10.1002/hbm.20749.

Costafreda SG, Fu CHY, Picchioni M, Toulopoulou T, McDonald C, Kravariti E, Walshe M, Prata D, Murray RM, McGuire PK. (2011) Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry*, 11:18. doi:10.1186/1471-244X-11-18.

Cosway R, Byrne M, Clafferty R, et al. (2000) Neuropsychological change in young people at high risk for schizophrenia: results from the first two neuropsychological assessments of the Edinburgh High Risk Study. *Psychol Med*, 30(5):1111-1121.

Cox DD, Savoy RL. (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2 Pt 1):261-270.

Crespo-Facorro B, Nopoulos PC, Chemerinski E, Kim J-J, Andreasen NC, Magnotta V. (2004) Temporal pole morphology and psychopathology in males with schizophrenia. *Psychiatry Res*, 132(2):107-115. doi:10.1016/j.pscychresns.2004.09.002.

D

Davatzikos C. (2004) Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage,* 23(1):17-20. doi:10.1016/j.neuroimage.2004.05.010.

Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughead JW, Gur RC, Langleben DD. (2005a) Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage***,** 28(3):663-668. doi:10.1016/j.neuroimage.2005.08.009.

Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y, Hughett P, Turetsky BI, Gur RE. (2005b) Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Arch. Gen. Psychiatry***,** 62(11):1218-1227. doi:10.1001/archpsyc.62.11.1218.

Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging,* 32(12):2322.e19-27. doi:10.1016/j.neurobiolaging.2010.05.023.

Dazzan P, Soulsby B, Mechelli A, Wood SJ, Velakoulis D, Phillips LJ, Yung AR, Chitnis X, Lin A, Murray RM, McGorry PD, McGuire PK, Pantelis C. (2012) Volumetric abnormalities predating the onset of schizophrenia and affective psychoses: an MRI study in subjects at ultrahigh risk of psychosis. *Schizophr Bull,* 38(5):1083-1091. doi:10.1093/schbul/sbr035.

DeLisi LE, Hoff AL, Schwartz JE, Shields GW, Halthore SN, Gupta SM, Henn FA, Anand AK. (1991) Brain morphology in first-episode schizophrenic-like psychotic patients: a quantitative magnetic resonance imaging study. *Biol. Psychiatry,* 29(2):159-175.

DeLisi LE. (1992) The significance of age of onset for schizophrenia. *Schizophr Bull,* 18(2):209-215.

DeLisi LE, Sakuma M, Maurizio AM, Relja M, Hoff AL. (2004) Cerebral ventricular change over the first 10 years after the onset of schizophrenia. *Psychiatry Res,* 130(1):57-70. doi:10.1016/j.pscychresns.2003.08.004.

DeLisi LE, Szulc KU, Bertisch HC, Majcher M, Brown K. (2006) Understanding structural brain changes in schizophrenia. *Dialogues Clin Neurosci,* 8(1):71-78.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, 43(1):44-58. doi:10.1016/j.neuroimage.2008.06.037.

Demirci O, Calhoun VD. (2009) Functional Magnetic Resonance Imaging - Implications for Detection of Schizophrenia. *Eur Neurol Rev*, 4(2):103-106.

Dixon L. (1999) Dual diagnosis of substance abuse in schizophrenia: prevalence and impact on outcomes. *Schizophr. Res*, 35 Suppl:S93-100.

Dubey R, Zhou J, Wang Y, Thompson PM, Ye J. (2014) Analysis of sampling techniques for imbalanced data: an N=648 ADNI Study. *NeuroImage*. 87:220-241. doi:10.1016/j.neuroimage.2013.10.005.

Duchesnay E, Cachia A, Boddaert N, Chabane N, Mangin JF, Martinot JL, Brunelle F, Zilcovicius M. (2011) Feature selection and classification of imbalanced datasets: application to PET images of children with autistic spectrum disorders. *Neuroimage*, 57(3): 1003-14.

E

Eack SM, Newhill CE. (2007) Psychiatric Symptoms and Quality of Life in Schizophrenia: A Meta-Analysis. *Schizophr Bull*, 33(5):1225-1237. doi:10.1093/schbul/sbl071.

Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG, MRC AIMS Consortium. (2010) Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *Neuroimage*, 49(1):44-56. doi:10.1016/j.neuroimage.2009.08.024.

Edwards J, Harris MG, Bapat S. (2005) Developing services for first-episode psychosis and the critical period. *Br J Psychiatry Suppl*, 48:s91-97. doi:10.1192/bjp.187.48.s91.

Ellison-Wright I, Glahn DC, Laird AR, Thelen SM, Bullmore E. (2008) The anatomy

of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. *Am J Psychiatry*, 165(8):1015-1023. doi:10.1176/appi.ajp.2008.07101562.

Erlenmeyer-Kimling L, Adamo UH, Rock D, Roberts SA, Bassett AS, Squires-Wheeler E, Cornblatt BA, Endicott J, Pape S, Gottesman II. (1997) The New York High-Risk Project. Prevalence and comorbidity of axis I disorders in offspring of schizophrenic parents at 25-year follow-up. *Arch. Gen. Psychiatry*, 54(12):1096-1102.

Erlenmeyer-Kimling L, Rock D, Roberts SA, Janal M, Kestenbaum C, Cornblatt B, Adamo UH, Gottesman II. (2000) Attention, memory, and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High-Risk Project. *Am J Psychiatry*, 157(9):1416-1422.

F

Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. (2007) COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Trans Med Imaging*, 26(1):93-105. doi:10.1109/TMI.2006.886812.

Fan Y, Gur RE, Gur RC, Wu X, Shen D, Calkins ME, Davatzikos C. (2008a) Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study. *Biol. Psychiatry*, 63(1):118-124. doi:10.1016/j.biopsych.2007.03.015.

Fan Y, Resnick SM, Wu X, Davatzikos C. (2008b) Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage*, 41(2):277-285. doi:10.1016/j.neuroimage.2008c.02.043.

Fan Y, Liu Y, Wu H, Hao Y, Liu H, Liu Z, Jiang T. (2011) Discriminant analysis of functional connectivity patterns on Grassmann manifold. *Neuroimage*, 56(4):2058-2067. doi:10.1016/j.neuroimage.2011.03.051.

Flaum M, O'Leary DS, Swayze VW, Miller DD, Arndt S, Andreasen NC. (1995) Symptom dimensions and brain morphology in schizophrenia and related psychotic disorders. *J Psychiatr Res*, 29(4):261-276.

Folnegović Z, Folnegović-Smalc V. (1994) Schizophrenia in Croatia: age of onset differences between males and females. *Schizophr. Res,* 14(1):83-91.

Folsom DP, Lebowitz BD, Lindamer LA, Palmer BW, Patterson TL, Jeste DV. (2006) Schizophrenia in late life: emerging issues . *Dialogues in Clinical Neuroscience*, 8(1):45-52.

Fornito A, Yücel M, Dean B, Wood SJ, Pantelis C. (2009) Anatomical abnormalities of the anterior cingulate cortex in schizophrenia: bridging the gap between neuroimaging and neuropathology. *Schizophr Bull*, 35(5):973-993. doi:10.1093/schbul/sbn025.

Friston KJ, Frith CD, Liddle PF, and Frackowiak RSJ. (1991) Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, 11:690:699.

Friston KJ, Jezzard P and Turner R. (1994) Analysis of Functional MRI Time-Series Human *Brain Mappin*.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. (1995) Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Human Brain Imaging,* 189-210.

Friston KJ. (1999) Schizophrenia and the disconnection hypothesis. *Acta Psychiatr Scand Suppl*, 395: 68-79.

Fu CHY, Mourao-Miranda J, Costafreda SG, Khanna A, Marquand AF, Williams SCR, Brammer MJ. (2008) Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biol. Psychiatry*, 63(7):656-662. doi:10.1016/j.biopsych.2007.08.020.

Fu CHY, Costafreda SG. (2013) Neuroimaging-based biomarkers in psychiatry: clinical opportunities of a paradigm shift. *Can J Psychiatry*, 58(9):499-508.

Fuller, R.L.M., Schultz, S.K. and Andreasen, N.C. (2003) The Symptoms of Schizophrenia, in Schizophrenia, Second Edition (eds S. R. Hirsch and D. R. Weinberger), Blackwell Science Ltd, Oxford, UK. doi: 10.1002/9780470987353.ch3

Fusar-Poli P, Perez J, Broome M, Borgwardt S, Placentino A, Caverzasi E, Cortesi

M, Veggiotti P, Politi P, Barale F, McGuire P. (2007) Neurofunctional correlates of vulnerability to psychosis: a systematic review and meta-analysis. *Neurosci Biobehav Rev*, 31(4):465-484. doi:10.1016/j.neubiorev.2006.11.006.

Fusar-Poli, P., Bhattacharyya, S., Allen, P., Crippa, J. A., Borgwardt, S., Martin-Santos, R., McGuire, P. (2010). Effect of image analysis software on neurofunctional activation during processing of emotional human faces. *Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia*, *17*(3), 311–314.

Fusar-Poli P, Borgwardt S, Crescini A, Deste G, Kempton MJ, Lawrie S, Mc Guire P, Sacchetti E. (2011) Neuroanatomy of vulnerability to psychosis: a voxel-based meta-analysis. *Neurosci Biobehav Rev*, 35(5):1175-1185. doi:10.1016/j.neubiorev.2010.12.005.

Fusar-Poli P, Yung AR. (2012) Should attenuated psychosis syndrome be included in DSM-5? *Lancet*, 379(9816):591-592. doi:10.1016/S0140-6736(11)61507-9.

Fusar-Poli P, Deste G, Smieskova R, Barlati S, Yung AR, Howes O, Stieglitz RD, Vita A, McGuire P, Borgwardt S. (2012a) Cognitive functioning in prodromal psychosis: a meta-analysis. *Arch Gen Psychiatry*, 69(6):562-571. doi:10.1001/archgenpsychiatry.2011.1592.

Fusar-Poli P, Bonoldi I, Yung AR, Borgwardt S, Kempton MJ, Valmaggia L, Barale F, Caverzasi E, McGuire P. (2012b) Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. *Arch. Gen. Psychiatry*, 69(3):220-229. doi:10.1001/archgenpsychiatry.2011.1472.

Fusar-Poli P, Borgwardt S, Bechdolf A, Addington J, Riecher-Rössler A, Schultze-Lutter F, Keshavan M, Wood S, Ruhrmann S, Seidman LJ, Valmaggia L, Cannon T, Velthorst E, De Haan L, Cornblatt B, Bonoldi I, Birchwood M, McGlashan T, Carpenter W, McGorry P, Klosterkötter J, McGuire P, Yung A. (2013a) The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry*, 70(1):107-120. doi:10.1001/jamapsychiatry.2013.269.

Fusar-Poli P, Byrne M, Badger S, Valmaggia LR, McGuire PK. (2013) Outreach and support in south London (OASIS), 2001-2011: ten years of early diagnosis and treatment for young individuals at high clinical risk for psychosis. *Eur. Psychiatry*, 28(5):315-326. doi:10.1016/j.eurpsy.2012.08.002.

G

Geddes JR, Lawrie SM. (1995) Obstetric complications and schizophrenia: a meta-analysis. *Br J Psychiatry,* 167(6):786-793.

Gilmore JH, Murray RM. (2006) Prenatal and perinatal factors. In: Lieberman JA, Stroup TS, Perkins DO, editors. The American Psychiatric Publishing Textbook of Schizophrenia. Washington, DC: American Psychiatric Publishing Inc.

Glahn DC, Ragland JD, Abramoff A, Barrett J, Laird AR, Bearden CE, Velligan DI. (2005) Beyond hypofrontality: a quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum Brain Mapp,* 25(1):60-69. doi:10.1002/hbm.20138.

Goldstein JM, Tsuang MT, Faraone SV. (1989) Gender and schizophrenia: implications for understanding the heterogeneity of the illness. *Psychiatry Res*, 28(3):243-253.

Goldstein JM, Cohen LS, Horton NJ, Lee H, Andersen S, Tohen M, Crawford A-MK, Tollefson G. (2002) Sex differences in clinical response to olanzapine compared with haloperidol. *Psychiatry Res*, 110(1):27-37.

Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage,* 14(1 Pt 1):21-36. doi:10.1006/nimg.2001.0786.

Gottesman II, McGuffin P, Farmer AE. (1987) Clinical genetics as clues to the "real" genetics of schizophrenia (a decade of modest gains while playing for time). *Schizophr Bull*, 13(1):23-47.

Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N. (2012) Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry,* 3:53. doi:10.3389/fpsyt.2012.00053.

Grube BS, Bilder RM, Goldman RS. (1998) Meta-analysis of symptom factors in schizophrenia. *Schizophr. Res,* 31(2-3):113-120.

Gschwandtner U, Aston J, Borgwardt S, Drewe M, Feinendegen C, Lacher D,

Lanzarone A, Stieglitz RD, Riecher-Rossler A. (2003) Neuropsychological and neurophysiological findings in individuals suspected to be at risk for schizophrenia: preliminary results from the Basel early detection of psychosis study - Früherkennung von Psychosen (FEPSY). *Acta Psychiatr Scand*, (2):152-155.

Guyon I, Weston J, Barnhill S, Vapnik V. (2002) Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn*, 46(1-3):389–422. doi:10.1023/A:1012487302797.

Guoyn I., Elisseeff A. (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research 3*, 1157-1182

Gur RE, Petty RG, Turetsky BI, Gur RC. (1996) Schizophrenia throughout life: sex differences in severity and profile of symptoms. *Schizophr. Res.* 21(1):1-12.

Gur RE, Cowell P, Turetsky BI, Gallacher F, Cannon T, Bilker W, Gur RC. (1998) A follow-up magnetic resonance imaging study of schizophrenia. Relationship of neuroanatomical changes to clinical and neurobehavioral measures. *Arch. Gen. Psychiatry*, 55(2):145-152.

Gur RE, Cowell PE, Latshaw A, Turetsky BI, Grossman RI, Arnold SE, Bilker WB, Gur RC. (2000) Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. *Arch. Gen. Psychiatry*, 57(8):761-768.

Gur RE, McEvoy J, Perkins D, Hamer RM, Gu H, Tohen M, HGDH Study Group. (2005) Antipsychotic drug effects on brain morphology in first-episode psychosis. *Arch. Gen. Psychiatry*, 62(4):361-370. doi:10.1001/archpsyc.62.4.361.

Gur RE, Keshavan MS, Lawrie SM. (2007) Deconstructing psychosis with human brain imaging. *Schizophr Bull*, 33(4):921-931. doi:10.1093/schbul/sbm045.

Gureje O. (1991) Gender and schizophrenia: age at onset and sociodemographic attributes. *Acta Psychiatr Scand*, 83(5):402-405.

H

Häfner H, Löffler W, Maurer K, Hambrecht M, an der Heiden W. (1999) Depression, negative symptoms, social stagnation and social decline in the early course of schizophrenia. *Acta Psychiatr Scand*, 100(2):105-118.

Häfner H. (2000) Onset and early course as determinants of the further course of schizophrenia. *Acta Psychiatr Scand Suppl*, 407 :44-48.

Häfner H, Maurer K, Trendler G, an der Heiden W, Schmidt M,Konnecke R. (20050 Schizophrenia and depression: challenging the paradigm of two separate diseases–a controlled study of schizophrenia,depression and healthy controls. *Schizophr. Res,* 77:11–24.

Haijma, S. V., Van Haren, N., Cahn, W., Koolschijn, P. C. M. P., Hulshoff Pol, H. E., & Kahn, R. S. (2013). Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. *Schizophrenia Bulletin*, *39*(5), 1129–1138.

Haro JM, Novick D, Suarez D, Alonso J, Lépine JP, Ratcliffe M, SOHO Study Group. (2006) Remission and relapse in the outpatient care of schizophrenia: three-year results from the Schizophrenia Outpatient Health Outcomes study. *J Clin Psychopharmacol,* 26(6):571-578. doi:10.1097/01.jcp.0000246215.49271.b8.

Harrison PJ, Weinberger DR. (2005) Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry,* 10(1):40-68; image 5. doi:10.1038/sj.mp.4001558.

Hashemi R, Bradley W, Lisanti C. 2004 MRI: The Basics. Lippincott Williams & Wilkins.

Hastie T, Tibshirani R, Friedman J. (2001) The Elements of Statistical Learning: Data Mining, Inference, and Prediction: Springer.

Hawton K, Sutton L, Haw C, Sinclair J, Deeks JJ. (2005) Schizophrenia and suicide: systematic review of risk factors. *Br J Psychiatry,* 187:9-20. doi:10.1192/bjp.187.1.9.

Haynes J-D, Rees G. (2006) Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci,* 7(7):523-534. doi:10.1038/nrn1931.

He H and Garcia EA. (2009) Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data engineering. Vol 21. No9.

Heinrichs RW, Zakzanis KK. (1998) Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, 12(3):426-445.

Hickling FW, McKenzie K, Mullen R, Murray R. (1999) A Jamaican psychiatrist evaluates diagnoses at a London psychiatric hospital. *Br J Psychiatry*, 175:283-285.

HMSO, 1991. Classification of Occupations of the Registrar General. HMSO, London.

Ho B-C, Andreasen NC, Nopoulos P, Arndt S, Magnotta V, Flaum M. (2003) Progressive structural brain abnormalities and their relationship to clinical outcome: a longitudinal magnetic resonance imaging study early in schizophrenia. *Arch. Gen. Psychiatry*, 60(6):585-594. doi:10.1001/archpsyc.60.6.585.

Hodges A, Byrne M, Grant E, Johnstone E. (1999) People at risk of schizophrenia. Sample characteristics of the first 100 cases in the Edinburgh High-Risk Study. *Br J Psychiatry*, 174:547-553.

Honea R, Crow TJ, Passingham D, Mackay CE. (2005) Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*,162(12):2233-2245. doi:10.1176/appi.ajp.162.12.2233.

Hsu C.C. (2003) A Practical Guide to Support Vector Classification Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin.

Hsu C., Chang C., and Lin C. (2010) A practical guide to support vector classifcation. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Huber G, Gross G. (1989). The concept of basic symptoms in schizophrenic and schizoaffective psychoses. *Recenti Prog Med*,80:646-652.

Huettel SA, Song AW, McCarthy G. (2004) Functional magnetic resonance imaging: *Sinauer Associates Sunderland*, Mass.

Hulshoff Pol HE, Schnack HG, Mandl RC, van Haren NE, Koning H, Collins DL, Evans AC, Kahn RS. (2001) Focal gray matter density changes in schizophrenia. *Arch. Gen. Psychiatry*, 58(12):1118-1125.

Hulshoff Pol HE, Kahn RS. (2008) What happens after the first episode? A review of progressive brain changes in chronically ill patients with schizophrenia. *Schizophr Bull,* ; 34(2):354-366. doi:10.1093/schbul/sbm168.

Hulshoff Pol HE, Kahn RS. (2014) Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage*, 84:299-306. doi:10.1016/j.neuroimage.2013.08.053.

Hunter R, Barry S. (2012) Negative symptoms and psychosocial functioning in schizophrenia: neglected but important targets for treatment. *Eur. Psychiatry*, 27(6):432-436. doi:10.1016/j.eurpsy.2011.02.015.

I

Ismail, B., Cantor-Graae, E., & McNeil, T. F. (1998). Minor physical anomalies in schizophrenic patients and their siblings. *The American Journal of Psychiatry*, *155*(12), 1695–1702.

J

Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A. (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol Med Monogr Suppl*, 20:1-97.

Jablensky A. (1997) The 100-year epidemiology of schizophrenia. *Schizophr. Res*, 28(2-3):111-125.

Jafri MJ, Calhoun VD. (2006) Functional classification of schizophrenia using feed forward neural networks. *Conf Proc IEEE Eng Med Biol Soc,* Suppl:6631-6634. doi:10.1109/IEMBS.2006.260906.

Jager F, Hornegger J. (2009) Nonrigid Registration of Joint Histograms for Intensity Standardization in Magnetic Resonance Imaging. *IEEE Trans Med Imag*. 28(1):137–150.

Japkowicz, N. The Class Imbalance Problem: Significance and Strategies. In Proceedings of the 2000 International Conference on Artificial Intelligence: Special

Track on Inductive Learning, Las Vegas, Nevada.

Japkowicz N, Stephen S. (2002) The class imbalance problem: a systematic study. Journal of Intelligent Data Analysis, 5 (6):429-449.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. (2012) FSL. *Neuroimage,* 62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015.

Job DE, Whalley HC, McConnell S, Glabus M, Johnstone EC, Lawrie SM. (**2002**) Structural gray matter differences between first-episode schizophrenics and normal controls using voxel-based morphometry. *Neuroimage,* 17(2):880-889.

Job DE, Whalley HC, McConnell S, Glabus M, Johnstone EC, Lawrie SM. (2003) Voxel-based morphometry of grey matter densities in subjects at high risk of schizophrenia. *Schizophr. Res.* 64(1):1-13.

Job DE, Whalley HC, Johnstone EC, Lawrie SM. (2005) Grey matter changes over time in high risk subjects developing schizophrenia. *Neuroimage,* 25(4):1023-1030. doi:10.1016/j.neuroimage.2005.01.006.

Job DE, Whalley HC, McIntosh AM, Owens DGC, Johnstone EC, Lawrie SM. (2006) Grey matter changes can improve the prediction of schizophrenia in subjects at high risk. *BMC Med*, 4:29. doi:10.1186/1741-7015-4-29.

Johnstone EC, Crow TJ, Frith CD, Husband J, Kreel L. (1976) Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *Lancet,* 2(7992):924-926.

Johnstone EC, Abukmeil SS, Byrne M, Clafferty R, Grant E, Hodges A, Lawrie SM, Owens DG. (2000) Edinburgh high risk study--findings after four years: demographic, attainment and psychopathological issues. *Schizophr. Res,* 46(1):1-15.

Johnstone EC, Lawrie SM, Cosway R. (2002a) What does the Edinburgh high-risk study tell us about schizophrenia? *Am. J. Med. Genet,* 114(8):906-912. doi:10.1002/ajmg.b.10304.

Johnstone EC, Cosway R, Lawrie SM. (2002b) Distinguishing characteristics of subjects with good and poor early outcome in the Edinburgh High-Risk Study. *Br J Psychiatry Suppl*, 43:s26-29.

Johnstone EC, Ebmeier KP, Miller P, Owens DGC, Lawrie SM. (2005) Predicting schizophrenia: findings from the Edinburgh High-Risk Study. *BJP*, 186(1):18-25. doi:10.1192/bjp.186.1.18.

Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, Macfall J, Fischl B, Dale A. (2006) Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*, 30(2):436-443. doi:10.1016/j.neuroimage.2005.09.046.

K

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

Karageorgiou E, Schulz SC, Gollub RL, Andreasen NC, Ho B-C, Lauriello J, Calhoun VD, Bockholt HJ, Sponheim SR, Georgopoulos AP. (2011) Neuropsychological testing and structural magnetic resonance imaging as diagnostic biomarkers early in the course of schizophrenia and related psychoses. *Neuroinformatics*, 9(4):321-333. doi:10.1007/s12021-010-9094-6.

Kasai K, Shenton ME, Salisbury DF, Hirayasu Y, Lee C-U, Ciszewski AA, Yurgelun-Todd D, Kikinis R, Jolesz FA, McCarley RW. (2003) Progressive decrease of left superior temporal gyrus gray matter volume in patients with first-episode schizophrenia. *Am J Psychiatry*, 160(1):156-164.

Kasparek T, Thomaz CE, Sato JR, Schwarz D, Janousova E, Marecek R, Prikryl R, Vanicek J, Fujita A, Ceskova E. (2011) Maximum-uncertainty linear discrimination analysis of first-episode schizophrenia subjects. *Psychiatry Res*, 191(3):174-181. doi:10.1016/j.pscychresns.2010.09.016.

Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou S-Y, Nakamura K, Matsui M, Sumiyoshi T, Seto H, Kurachi M. (2007) Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *Neuroimage*, 34(1):235-242. doi:10.1016/j.neuroimage.2006.08.018.

Kendler KS, Lieberman JA, Walsh D. (1989) The Structured Interview for Schizotypy (SIS): a preliminary report. *Schizophr Bull*, 15(4):559-571.

Kernighan B. W., Ritchie D. M. (1978). *The C Programming Language* (1st ed.). Englewood Cliffs, NJ: Prentice Hall. ISBN 0-13-110163-3.

Keshavan MS, Bagwell WW, Haas GL, Sweeney JA, Schooler NR, Pettegrew JW. (1994) Changes in caudate volume with neuroleptic treatment. *Lancet*, 344(8934):1434.

Keshavan MS, Gilbert AR, Diwadkar VA. Neurodevelopmental theories. In: Lieberman JA, Stroup TS, Perkins DO, editors. The American Psychiatric Publishing Textbook ofSchizophrenia. Washington, DC: American Psychiatric Publishing Inc.; 2006. p. 69-84.

Khodayari-Rostamabad A, Hasey GM, Maccrimmon DJ, Reilly JP, de Bruin H. (2010) A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. *Clin Neurophysiol*, 121(12):1998-2006. doi:10.1016/j.clinph.2010.05.009.

Klauschen F, Goldman A, Barra V, Meyer-Lindenberg A, Lundervold A. (2009) Evaluation of automated brain MR image segmentation and volumetry methods. *Hum Brain Mapp*, 30(4):1310-1327. doi:10.1002/hbm.20599.

Klöppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, Mader I, Mitchell LA, Patel AC, Roberts CC, Fox NC, Jack CR, Ashburner J, Frackowiak RSJ. (2008a) Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain*, 131(Pt 11):2969-2974. doi:10.1093/brain/awn239.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR, Ashburner J, Frackowiak RSJ. (2008b) Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(Pt 3):681-689. doi:10.1093/brain/awm319.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Ashburner J, Frackowiak, RSJ. (2009) A plea for confidence intervals and consideration of generalizability in diagnostic studies. *Brain*, 132(4):e102. doi:10.1093/brain/awn091.

Klöppel S, Abdulkadir A, Jack CR, Koutsouleris N, Mourão-Miranda J, Vemuri P. (2012) Diagnostic neuroimaging across diseases. *Neuroimage*, 61(2):457-463. doi:10.1016/j.neuroimage.2011.11.002.

Klosterkötter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. (2001) Diagnosing schizophrenia in the initial prodromal phase. *Arch. Gen. Psychiatry*, 58(2):158-164.

Kohavi R. (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection International Joint Conference on Articial Intelligence (IJCAI).

Kohavi R., Provost F. (1998) Glossary of terms, Machine Learning, Vol. 30, No. 2/3, pp. 271-274.

Kopelman A, Andreasen NC, Nopoulos P. (2005) Morphology of the anterior cingulate gyrus in patients with schizophrenia: relationship to typical neuroleptic exposure. *Am J Psychiatry*, 162(10):1872-1878. doi:10.1176/appi.ajp.162.10.1872.

Koutsouleris N, Gaser C, Jäger M, Bottlender R, Frodl T, Holzinger S, Schmitt GJE, Zetzsche T, Burgermeister B, Scheuerecker J, Born C, Reiser M, Möller H-J, Meisenzahl EM. (2008) Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. *Neuroimage*, 39(4):1600-1612. doi:10.1016/j.neuroimage.2007.10.029.

Koutsouleris N, Schmitt GJE, Gaser C, Bottlender R, Scheuerecker J, McGuire P, Burgermeister B, Born C, Reiser M, Möller H-J, Meisenzahl EM. (2009) Neuroanatomical correlates of different vulnerability states for psychosis and their clinical outcomes. *Br J Psychiatry*, 195(3):218-226. doi:10.1192/bjp.bp.108.052068.

Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Möller H-J, Gaser C. (2009) Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch. Gen. Psychiatry*, 66(7):700-712. doi:10.1001/archgenpsychiatry.2009.62.

Koutsouleris N, Gaser C, Bottlender R, Davatzikos C, Decker P, Jäger M, Schmitt G, Reiser M, Möller H-J, Meisenzahl EM. (2010) Use of neuroanatomical pattern regression to predict the structural brain dynamics of vulnerability and transition to psychosis. *Schizophr. Res*, 123(2-3):175-187. doi:10.1016/j.schres.2010.08.032.

Koutsouleris N, Borgwardt S, Meisenzahl EM, Bottlender R, Möller H-J, Riecher-Rössler A. (2012a) Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study. *Schizophr Bull*, 38(6):1234-1246. doi:10.1093/schbul/sbr145.

Koutsouleris N, Davatzikos C, Bottlender R, Patschurek-Kliche J, Scheuerecker J, Decker O, Gaser C, Moller HJ, Meisenzahl EM. (2012b) Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull*, 38(6):1200-1215. doi:10.1093/schbul/sbr037.

Koutsouleris N, Meisenzahl EM, Borgwardt S, Riecher-Rössler A, Frodl T, Kambeitz J, Köhler Y, Falkai P, Möller H-J, Reiser M, Davatzikos C. (2015a) Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*, doi:10.1093/brain/awv111.

Koutsouleris N, Riecher-Rössler A, Meisenzahl EM, Smieskova R, Studerus E, Kambeitz-Ilankovic L, von Saldern S, Cabral C, Reiser M, Falkai P, Borgwardt S. (2015b) Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. *Schizophr Bull*, 41(2):471-482. doi:10.1093/schbul/sbu078.

Kraepelin E. 1919 Dementia Praecox and Paraphrenia, edited by Robertson G., Krieger, New York, 1971.

Kubicki M, Shenton ME, Salisbury DF, Hirayasu Y, Kasai K, Kikinis R, Jolesz FA, McCarley RW. (2002) Voxel-based morphometric analysis of gray matter in first episode schizophrenia. *Neuroimage,* 17(4):1711-1719.

L

LaConte S, Strother S, Cherkassky V, Anderson J, Hu X. (2005) Support vector machines for temporal classification of block design fMRI data. *Neuroimage*, 26(2):317-329. doi:10.1016/j.neuroimage.2005.01.048.

Lawrie SM, Abukmeil SS. (1998) Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. *Br J Psychiatry*, 172:110-120.

Lawrie SM, Whalley H, Kestelman JN, Abukmeil SS, Byrne M, Hodges A, Rimmington JE, Best JJ, Owens DG, Johnstone EC. (1999) Magnetic resonance imaging of brain in people at high risk of developing schizophrenia. *Lancet*, 353(9146):30-33. doi:10.1016/S0140-6736(98)06244-8.

Lawrie SM, Byrne M, Miller P, Hodges A, Clafferty RA, Cunningham Owens DG, Johnstone EC. (2001a) Neurodevelopmental indices and the development of psychotic symptoms in subjects at high risk of schizophrenia. *Br J Psychiatry*, 178:524-530.

Lawrie SM, Whalley HC, Abukmeil SS, Kestelman JN, Donnelly L, Miller P, Best JJ, Owens DG, Johnstone EC. (2001b) Brain structure, genetic liability, and psychotic symptoms in subjects at high risk of developing schizophrenia. *Biol. Psychiatry*, 49(10):811-823.

Lawrie SM, Buechel C, Whalley HC, Frith CD, Friston KJ, Johnstone EC. (2002) Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. *Biol. Psychiatry*, 51(12):1008-1011.

Lawrie SM, McIntosh AM, Hall J, Owens DGC, Johnstone EC. (2008) Brain structure and function changes during the development of schizophrenia: the evidence from studies of subjects at increased genetic risk. *Schizophr Bull* 34(2):330-340. doi:10.1093/schbul/sbm158.

Lawrie SM, Olabi B, Hall J, McIntosh AM. (2011) Do we have any solid evidence of clinical utility about the pathophysiology of schizophrenia? *World Psychiatry*, 10(1):19-31.

Lazar N. (2008) The Statistical Analysis of Functional MRI Data. *Statistics for Biology and Health*, Springer.

Lemm S., Blankertz B., Dickhaus T., Muller K.R. (2011) Introduction to machine learning for brain imaging. *Neuroimage*, 56(2):387-399. doi:10.1016/j.neuroimage.2010.11.004.

Lencz T, Smith CW, McLaughlin D, Auther A, Nakayama E, Hovey L, Cornblatt BA. (2006) Generalized and specific neurocognitive deficits in prodromal schizophrenia. *Biol. Psychiatry*, 59(9):863-871. doi:10.1016/j.biopsych.2005.09.005.

Leucht S, Burkard T, Henderson J, Maj M, Sartorius N. (2007) Physical illness and schizophrenia: a review of the literature. *Acta Psychiatr Scand*, 116(5):317-333. doi:10.1111/j.1600-0447.2007.01095.x.

Leung A, Chue P. (2000) Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr Scand Suppl*, 401:3-38.

Li X, Li L, Lu H, Liang Z. (2005) Partial volume segmentation of brain magnetic resonance images based on maximum a posteriori probability. *Med Phys*, 32(7):2337-2345.

Lieberman JA. (1999) Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biol. Psychiatry*, 46(6):729-739.

Lieberman J, Chakos M, Wu H, Alvir J, Hoffman E, Robinson D, Bilder R. (2001) Longitudinal study of brain morphology in first episode schizophrenia. *Biol. Psychiatry*, 49(6):487-499.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*(4), 423–428.

Lieberman JA, Tollefson GD, Charles C, Zipursky R, Sharma T, Kahn RS, Keefe RSE, Green AI, Lin A, Reniers RLEP, Wood SJ. (2013) Clinical staging in severe mental disorder: evidence from neurocognition and neuroimaging. *Br J Psychiatry Suppl*, 54:s11-17. doi:10.1192/bjp.bp.112.119156.

Liu, M., Zhang, D., Shen, D. and the Alzheimer's Disease Neuroimaging Initiative. (2012) Ensemble sparse classification of Alzheimer's disease. *Neuroimage* (60), 1106-1116.

Lukoff D, Nuechterlein KH, Ventura J. (1986) Manual for the expanded brief psychiatric rating scale. *Schizophr Bull*, 12:594–602.

Lymer GKS, Job DE, William T, Moorhead J, McIntosh AM, Owens DGC, Johnstone EC, Lawrie SM. (2006) Brain-behaviour relationships in people at high genetic risk of schizophrenia. *Neuroimage*, 33(1):275-285. doi:10.1016/j.neuroimage.2006.06.031.

M

Madsen AL, Karle A, Rubin P, Cortsen M, Andersen HS, Hemmingsen R. (1999) Progressive atrophy of the frontal lobes in first-episode schizophrenia: interaction with clinical course and neuroleptic treatment. *Acta Psychiatr Scand*, 100(5):367-

374.

Mäki P, Veijola J, Jones PB, Murray GK, Koponen H, Tienari P, Miettunen J, Tanskanen P, Wahlberg K-E, Koskinen J, Lauronen E, Isohanni M. (2005) Predictors of schizophrenia--a review. *Br. Med. Bull*, 73-74:1-15. doi:10.1093/bmb/ldh046.

Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, Susser ES. (2001) Advancing paternal age and the risk of schizophrenia. *Arch. Gen. Psychiatry*, 58(4):361-367.

Marquand AF, De Simoni S, O'Daly OG, Williams SC, Mourão-Miranda J, Mehta MA. (2011) Pattern Classification of Working Memory Networks Reveals Differential Effects of Methylphenidate, Atomoxetine, and Placebo in Healthy Volunteers. *Neuropsychopharmacology*, 36(6):1237-1247. doi:10.1038/npp.2011.9.

Marshall M, Lockwood A. (2004) Early Intervention for psychosis. *Cochrane Database Syst Rev* (2):CD004718. doi:10.1002/14651858.CD004718.

Marshall M, Lewis S, Lockwood A, Drake R, Jones P, Croudace T. (2005) Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Arch. Gen. Psychiatry*, 62(9):975-983. doi:10.1001/archpsyc.62.9.975.

Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourão-Miranda J. (2010) Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage*, 49(3):2178-2189. doi:10.1016/j.neuroimage.2009.10.072.

Mathalon DH, Sullivan EV, Lim KO, Pfefferbaum A. (2001) Progressive brain volume changes and the clinical course of schizophrenia in men: a longitudinal magnetic resonance imaging study. *Arch. Gen. Psychiatry*, 58(2):148-157.

McCormick L, Decker L, Nopoulos P, Ho B-C, Andreasen N. (2005) Effects of atypical and typical neuroleptics on anterior cingulate volume in schizophrenia. *Schizophr Res*, 80(1):73-84. doi:10.1016/j.schres.2005.06.022.

McGlashan TH. (1999) Duration of untreated psychosis in first-episode schizophrenia: marker or determinant of course? *Biol Psychiatry*, 46(7):899-907. doi:10.1016/S0006-3223(99)00084-0.

McGlashan TH, Zipursky RB, Perkins D, Addington J, Miller TJ, Woods SW, Hawkins KA, Hoffman R, Lindborg S, Tohen M, Breier A. (2003) The PRIME North America randomized double-blind clinical trial of olanzapine versus placebo in patients at risk of being prodromally symptomatic for psychosis. I. Study rationale and design. *Schizophr. Res,* 61(1):7-18.

McGorry PD. (2002a) The recognition and optimal management of early psychosis: an evidence-based reform. *World Psychiatry*, 1(2):76-83.

McGorry PD, Yung AR, Phillips LJ, Yuen HP, Francey S, Cosgrave EM, Germano D, Bravin J, McDonald T, Blair A, Adlard S, Jackson H. (2002b) Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms. *Arch. Gen. Psychiatry*, 59(10):921-928.

McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. (2004) A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med*, 2:13. doi:10.1186/1741-7015-2-13.

McGrath J, Saha S, Chant D, Welham J. (2008) Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev*. 3067- 76

McGuffin P, Farmer A, Harvey I. (1991) A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch. Gen. Psychiatry*, 48(8):764-770.

McGuffin P, Owen MJ, Farmer AE. (1995) Genetic basis of schizophrenia. *Lancet*, 346(8976):678-682.

McIntosh AM, Whalley HC, McKirdy J, Hall J, Sussmann JED, Shankar P, Johnstone EC, Lawrie SM. (2008) Prefrontal function and activation in bipolar disorder and schizophrenia. *Am J Psychiatry*, 165(3):378-384. doi:10.1176/appi.ajp.2007.07020365.

Mechelli A., Price C.J., Friston K.J., Ashburner J (2005). Voxel-based morphometry of the human brain: methods and applications. *Curr Med Imaging Reviews*;1:105–113

Mechelli A, Riecher-Rössler A, Meisenzahl EM, Tognin S, Wood SJ, Borgwardt SJ,

Koutsouleris N, Yung AR, Stone JM, Phillips LJ, McGorry PD, Valli I, Velakoulis D, Woolley J, Pantelis C, McGuire P. (2011) Neuroanatomical abnormalities that predate the onset of psychosis: a multicenter study. *Arch. Gen. Psychiatry*, 68(5):489-495. doi:10.1001/archgenpsychiatry.2011.42.

Meisenzahl EM, Koutsouleris N, Bottlender R, Scheuerecker J, Jäger M, Teipel SJ, Holzinger S, Frodl T, Preuss U, Schmitt G, Burgermeister B, Reiser M, Born C, Möller H-J. (2008a) Structural brain alterations at different stages of schizophrenia: a voxel-based morphometric study. *Schizophr. Res*, 104(1-3):44-60. doi:10.1016/j.schres.2008.06.023.

Meisenzahl EM, Koutsouleris N, Gaser C, Bottlender R, Schmitt GK, McGuire P, Decker P, Burgermeister B, Born C, Reiser M, Moller HJ. (2008b) Structural brain alterations in subjects at high-risk of psychosis: a voxel-based morphometric study. *Schizophr Res*, 102(1-3):150-162. doi:10.1016/j.schres.2008.02.023.

Miller P, Byrne M, Hodges A, Lawrie SM, Owens DGC, Johnstone EC. (2002a) Schizotypal components in people at high risk of developing schizophrenia: early findings from the Edinburgh High-Risk Study. *BJP*, 180(2):179-184. doi:10.1192/bjp.180.2.179.

Miller PM, Lawrie SM, Byrne M, Cosway R, Johnstone EC. (2002b) Self-rated schizotypal cognitions, psychotic symptoms and the onset of schizophrenia in young people at high risk of schizophrenia. *Acta Psychiatrica Scandinavica*, 105(5):341–345. doi:10.1034/j.1600-0447.2002.1o175.x.

Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC. (2009) Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch. Gen. Psychiatry*, 66(8):811-822. doi:10.1001/archgenpsychiatry.2009.91.

Meyer-Lindenberg A. (2010) From maps to mechanisms through neuroimaging of schizophrenia. *Nature*, 468(7321):194-202. doi:10.1038/nature09569.

Mirsky AF, Kugelmass S, Ingraham LJ, Frenkel E, Nathan M. (1995) Overview and summary: twenty-five-year follow-up of high-risk children. *Schizophr Bull*, 21, 227-39.

Mitchell T.M., Hutchinson R., Niculescu R.S., Pereira F., Wang X. (2004) Learning to Decode Cognitive States from Brain Images, *Machine Learning*, 57, 145-175.

Moore THM, Zammit S, Lingford-Hughes A, Barnes TRE, Jones PB, Burke M, Lewis G. (2007) Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet*, 370(9584):319-328. doi:10.1016/S0140-6736(07)61162-3.

Moorhead TWJ, Gountouna V-E, Job DE, McIntosh AM, Romaniuk L, Lymer GKS, Whalley HC, Waiter GD, Brennan D, Ahearn TS, Cavanagh J, Condon B, Steele JD, Wardlaw JM, Lawrie SM. (2009) Prospective multi-centre Voxel Based Morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. *BMC Med Imaging*, 9:8. doi:10.1186/1471-2342-9-8.

Moorhead TWJ, Stanfield AC, McKechanie AG, Dauvermann MR, Johnstone EC, Lawrie SM, Cunningham Owens DG. (2013) Longitudinal gray matter change in young people who are at enhanced risk of schizophrenia due to intellectual impairment. *Biol. Psychiatry*, 73(10):985-992. doi:10.1016/j.biopsych.2012.12.011.

Morcom AM, Fletcher PC. (2007) Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage*, 37(4):1073-1082. doi:10.1016/j.neuroimage.2006.09.013.

Morgan VA, Castle DJ, Jablensky AV. (2008) Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) Disorders. *Aust N Z J Psychiatry*, 42(1):74-82. doi:10.1080/00048670701732699.

Mourão-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M. (2005) Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage*, 28(4):980-995. doi:10.1016/j.neuroimage.2005.06.070.

Mourão-Miranda J, Hardoon DR, Hahn T, Marquand AF, Williams SCR, Shawe-Taylor J, Brammer M. (2011) Patient classification as an outlier detection problem: an application of the One-Class Support Vector Machine. *Neuroimage*, 58(3):793-804. doi:10.1016/j.neuroimage.2011.06.042.

Mourao-Miranda J, Reinders A a. TS, Rocha-Rego V, Lappin J, Rondina J, Morgan C, Morgan KD, Fearon P, Jones PB, Doody GA, Murray RM, Kapur S, Dazzan P. (2012a) Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol Med*, 42(5):1037-1047. doi:10.1017/S0033291711002005.

Mourão-Miranda J, Oliveira L, Ladouceur CD, Marquand A, Brammer M, Birmaher B, Axelson D, Phillips ML. (2012b) Pattern recognition and functional neuroimaging help to discriminate healthy adolescents at risk for mood disorders from low risk adolescents. *PLoS ONE*, 7(2):e29482. doi:10.1371/journal.pone.0029482

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. (2005) Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement*, 1(1):55-66. doi:10.1016/j.jalz.2005.06.003.

Murray RM, Lewis SW. (1987) Is schizophrenia a neurodevelopmental disorder? *British Medical Journal (Clinical research ed)*, 295(6600):681-682.

Murthy GV, Janakiramaiah N, Gangadhar BN, Subbakrishna DK. (1998) Sex difference in age at onset of schizophrenia: discrepant findings from India. *Acta Psychiatr Scand*, 97(5):321-325.

Mwangi B, Ebmeier KP, Matthews K, Steele JD. (2012) Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain*, 135(Pt 5):1508-1521. doi:10.1093/brain/aws084.

N

Nakamura M, Salisbury DF, Hirayasu Y, Bouix S, Pohl KM, Yoshida T, Koo M-S, Shenton ME, McCarley RW. (2007) Neocortical gray matter volume in first-episode schizophrenia and first-episode affective psychosis: a cross-sectional and longitudinal MRI study. *Biol. Psychiatry*, 62(7):773-783. doi:10.1016/j.biopsych.2007.03.030.

Navari S, Dazzan P. (2009) Do antipsychotic drugs affect brain structure? A systematic and critical review of MRI findings. *Psychol Med*, 39(11):1763-1777. doi:10.1017/S0033291709005315.

Nelson MD, Saykin AJ, Flashman LA, Riordan HJ. (1998) Hippocampal volume reduction in schizophrenia as assessed by magnetic resonance imaging: a meta-analytic study. *Arch Gen Psychiatry*, 55(5):433-440.

Nenadic I, Sauer H, Gaser C. (2010) Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. *Neuroimage*,

49(2):1153-1160. doi:10.1016/j.neuroimage.2009.10.014.

Nieman DH, Ruhrmann S, Dragt S, Soen F, Van Tricht MJ, Koelman JH, Bour LJ, Velthorst E, Becker HE, Weise M, Linszen DH, de Haan L. (2014) Psychosis prediction: stratification of risk estimation with information-processing and premorbid functioning variables. *Schizophr Bull*, 40(6):1482-1490. doi:10.1093/schbul/sbt145.

Nieuwenhuis M, van Haren NEM, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG. (2012) Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage*, 61(3):606-612. doi:10.1016/j.neuroimage.2012.03.079.

Nyul LG, Udupa JK. (1999) On Standardizing the MR Image Intensity Scale. *Mag Res in Medicine*, 42(6):1072–1081.

O

O'Connor M, Harris JM, McIntosh AM, Owens DGC, Lawrie SM, Johnstone EC. (2009) Specific cognitive deficits in a group at genetic high risk of schizophrenia. *Psychol Med*, 39(10):1649-1655. doi:10.1017/S0033291709005303.

Ogawa S., Lee T.M., Kay A.R., Tank D.W. (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci*, 9868-9872.

Olabi B, Ellison-Wright I, McIntosh AM, Wood SJ, Bullmore E, Lawrie SM. (2011) Are there progressive brain changes in schizophrenia? A meta-analysis of structural magnetic resonance imaging studies. *Biol. Psychiatry*, 70(1):88-96. doi:10.1016/j.biopsych.2011.01.032.

Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. (2012) Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*, 36(4):1140-1152. doi:10.1016/j.neubiorev.2012.01.004.

Owens DGC, Johnstone EC. ()2006 Precursors and prodromata of schizophrenia: findings from the Edinburgh High Risk Study and their literature context. *Psychol Med*, 36(11):1501-1514. doi:10.1017/S0033291706008221.

P

Pantelis C, Velakoulis D, McGorry PD, Wood SJ, Suckling J, Phillips LJ, Yung AR, Bullmore ET, Brewer W, Soulsby B, Desmond P, McGuire PK. (2003) Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. *Lancet*, 361(9354):281-288. doi:10.1016/S0140-6736(03)12323-9.

Pantelis, C., Yücel, M., Wood, S. J., Velakoulis, D., Sun, D., Berger, G., McGorry, P. D. (2005). Structural brain imaging evidence for multiple pathological processes at different stages of brain development in schizophrenia. *Schizophrenia Bulletin, 31*(3), 672–696. http://doi.org/10.1093/schbul/sbi034

Pereira F, Mitchell T, Botvinick M. (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1 Suppl):S199-209. doi:10.1016/j.neuroimage.2008.11.007.

Perkins DO, Gu H, Boteva K, Lieberman JA. (2005) Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry*, 162(10):1785-1804. doi:10.1176/appi.ajp.162.10.1785.

Pettersson-Yeo W, Allen P, Benetti S, McGuire P, Mechelli A. (2011) Dysconnectivity in schizophrenia: where are we now? *Neurosci Biobehav Rev*, 35(5):1110-1124. doi:10.1016/j.neubiorev.2010.11.004.

Pettersson-Yeo W, Benetti S, Marquand AF, Dell'acqua F, Williams SC, Allen P, prate D, McGuire P, Mechelli A. (2013) Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. *Psychol Med*, 43(12):2547-2562. doi:10.1017/S003329171300024X.

Phillips LJ, Yung AR, McGorry PD. (2000) Identification of young people at risk of psychosis: validation of Personal Assessment and Crisis Evaluation Clinic intake criteria. *Aust N Z J Psychiatry*, 34 Suppl:S164-169.

Phillips LJ, Velakoulis D, Pantelis C, Wood S, Yuen HP, Yung AR, Desmond P, Brewer W, McGorry PD. (2002) Non-reduction in hippocampal volume is associated with higher risk of psychosis. *Schizophr Res*, 58(2-3):145-158.

Phillips KA, Van Bebber S, Issa AM. (2006) Diagnostics and biomarker development: priming the pipeline. *Nat Rev Drug Discov*, 5(6):463-469. doi:10.1038/nrd2033.

Phillips LJ, McGorry PD, Yuen HP, Ward J, Donovan K, Kelly D, Francey SM, Yung AR. (2007) Medium term follow-up of a randomized controlled trial of interventions for young people at ultra high risk of psychosis. *Schizophr. Res*, 96(1-3):25-33. doi:10.1016/j.schres.2007.05.018.

Phillips CL, Bruno M-A, Maquet P, Boly M, Noirhomme Q, Schnakers C, Vanhaudenhuyse A, Bonjean M, Hustinx R, Moonen G, Luxen A, Laureys S. (2011) "Relevance vector machine" consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. *Neuroimage*, 56(2):797-808. doi:10.1016/j.neuroimage.2010.05.083.

Pinkham A, Loughead J, Ruparel K, Wu W-C, Overton E, Gur R, Gur R. (2011) Resting quantitative cerebral blood flow in schizophrenia measured by pulsed arterial spin labeling perfusion MRI. *Psychiatry Res*, 194(1):64-72. doi:10.1016/j.pscychresns.2011.06.013.

Polikar, R. (2006) Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine, 6*(3), 21–45. doi:10.1109/MCAS.2006.1688199

Pukrop R, Schultze-Lutter F, Ruhrmann S, Brokhaus-Dumke A, Tendolar, Bechdolf A, Matuschek E, Klosterkotter J. (2006) Neurocognitive functioning in subjects at risk for a first episode of psychosis compared with first- and multiple-episode schizophrenia. *J Clin Exp Neuropsychol*, 28(8):1388-1407. doi:10.1080/13803390500434425.

Pukrop R, Ruhrmann S, Schultze-Lutter F, Bechdolf A, Brockhaus-Dumke A, Klosterkötter J. (2007) Neurocognitive indicators for a conversion to psychosis: comparison of patients in a potentially initial prodromal state who did or did not convert to a psychosis. *Schizophr Res*, 92(1-3):116-125. doi:10.1016/j.schres.2007.01.020.

R

Rakotomamonjy A. (2003)Variable selection using svm-based criteria. Journal of Machine Learning Research 31357-1370

Rasmussen CE, Williams CKI. (2006) Gaussian Processes for Machine Learning. MIT Press.

Rey A. (1964) L'examen clinique en psychologie. Presses Universitaires de France, Paris.

Riecher-Rössler A, Häfner H. (2000) Gender aspects in schizophrenia: bridging the border between social and biological psychiatry. *Acta Psychiatr Scand Suppl*, (407):58-62.

Riecher-Rössler A, Gschwandtner U, Borgwardt S, Aston J, Pflüger M, Rössler W. (2006) Early detection and treatment of schizophrenia: how early? *Acta Psychiatr Scand Suppl*, 429:73-80. doi:10.1111/j.1600-0447.2005.00722.x.

Riecher-Rössler A, Gschwandtner U, Aston J, Borgwardt S, Drewe M, Fuhr P, Pflüger M, Radü W, Schindler C, Stieglitz R-D. (2007) The Basel early-detection-of-psychosis (FEPSY)-study--design and preliminary results. *Acta Psychiatr Scand*, 115(2):114-125. doi:10.1111/j.1600-0447.2006.00854.x.

Riecher-Rössler A, Aston J, Ventura J, Merlo M, Borgwardt S, Gschwandtner U, Stieglitz R-D. (2008) [The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity]. *Fortschr Neurol Psychiatr*, 76(4):207-216. doi:10.1055/s-2008-1038155.

Riecher-Rössler A, Pflueger MO, Aston J, Borgwardt SJ, Brewer WJ, Gschwandtner U, Stieglitz R-D. (2009) Efficacy of using cognitive status in predicting psychosis: a 7-year follow-up. *Biol. Psychiatry*, 66(11):1023-1030. doi:10.1016/j.biopsych.2009.07.020.

Riecher A, Maurer K, Loffler W, Fatkenhaurer B, An Der Heiden W, Stromgren E, Hafner H. (1990) Gender differences in age at onset and course of schizophrenic disorders. A contribution to the understanding of the disease? in Search for the Causes of Schizophrenia pp.14-33, Springer.

Ripke, S., B. M. Neale, A. Corvin, J. T. Walters, K. Farh, P. A. Holmans, P. Lee, et al. (2014). “Biological Insights From 108 Schizophrenia-Associated Genetic Loci.”

Nature 511 (7510): 421- 427. doi:10.1038/nature13595. http://dx.doi.org/10.1038/nature13595.

Robitaille N, Mouiha A, Crépeault B, Valdivia F, Duchesne S, The Alzheimer's Disease Neuroimaging Initiative. (2012) Tissue-Based MRI Intensity Standardization: Application to Multicentric Datasets. *International Journal of Biomedical Imaging*, 347120. doi:10.1155/2012/347120.

Ruhrmann S, Schultze-Lutter F, Salokangas RK, Heinimaa M, Linszen D, Dingemans P, Birchwood M, Patterson P, Juckel G, Heunz A, Morrison A, Lewis S, von Reventlow HG, Klostekotter J. (2010) Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European Prediction of Psychosis Study. *Arch Gen Psychiatry*, 67:241–251**.**

Russell AT. (1994) The clinical presentation of childhood-onset schizophrenia. *Schizophr Bull*, 20(4):631-646.

Rust J. (1988) The Rust Inventory of Schizotypal Cognitions (RISC). *Schizophr Bull*, 14(2):317-322.

S

Salmond, C. H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D. G., & Friston, K. J. (2002). Distributional assumptions in voxel-based morphometry. *NeuroImage*, *17*(2), 1027–1030.

Scheepers FE, de Wied CC, Hulshoff Pol HE, van de Flier W, van der Linden JA, Kahn RS. (2001) The effect of clozapine on caudate nucleus volume in schizophrenic patients previously treated with typical antipsychotics. *Neuropsychopharmacology*, 24(1):47-54. doi:10.1016/S0893-133X(00)00172-X.

Schenkel LS, Silverstein SM. (2004) Dimensions of premorbid functioning in schizophrenia: a review of neuromotor, cognitive, social, and behavioral domains. *Genet Soc Gen Psychol Monogr*, 130(3):241-270. doi:10.3200/MONO.130.3.241-272.

Schnack HG, Nieuwenhuis M, van Haren NEM, Abramovic L, Scheewe TW, Brouwer RM, Hulshoff Pol HE, Kahn RS. (2014) Can structural MRI aid in clinical

classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage*, 84:299-306. doi:10.1016/j.neuroimage.2013.08.053.

Schneider K. (1959) Clinical Psychopathology. Grune and Stratton, New York. Translated by Hamilton, MW.

Schultze-Lutter F. (2009) Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. *Schizophr Bull*, 35(1):5-8. doi:10.1093/schbul/sbn139.

Seidman LJ, Giuliano AJ, Smith CW, Stone WS, Glatt SJ, Meyer E, Faraone SV, Tsuang MT, Cornblatt B. (2006) Neuropsychological functioning in adolescents and young adults at genetic risk for schizophrenia and affective psychoses: results from the Harvard and Hillside Adolescent High Risk Studies. *Schizophr Bull*, 32(3):507-524. doi:10.1093/schbul/sbj078.

Shen H, Wang L, Liu Y, Hu D. (2010) Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *Neuroimage*, 49(4):3110-3121. doi:10.1016/j.neuroimage.2009.11.011.

Shenton ME, Wible CG, McCarley R. (1997) A review of magnetic resonance imaging of brain abnormalities in schizophrenia. In: *Brain Imaging in Clinical Psychiatry*, New York: Marcel Dekker, Inc., pp 297–380.

Shenton ME, Dickey CC, Frumin M, McCarley RW. (2001) A review of MRI findings in schizophrenia. *Schizophr Res* 49(1-2):1-52.

Sigmundsson T, Suckling J, Maier M, Williams S, Bullmore E, Greenwood K, Fukuda R, Ron M, Toone B. (2001) Structural abnormalities in frontal, temporal, and limbic regions and interconnecting white matter tracts in schizophrenic patients with prominent negative symptoms. *Am J Psychiatry*, 158(2):234-243.

Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J, Radue EW, McGuire PK, Riecher-Rössler A, Borgwardt SJ. (2009) The effects of antipsychotics on the brain: what have we learnt from structural imaging of schizophrenia?--a systematic review. *Curr. Pharm. Des*, 15(22):2535-2549.

Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J, Radue EW, McGuire PK, Riecher-Rössler A, Borgwardt SJ. (2010) Neuroimaging predictors of

transition to psychosis--a systematic review and meta-analysis. *Neurosci Biobehav Rev*, 34(8):1207-1222. doi:10.1016/j.neubiorev.2010.01.016.

Smieskova R, Fusar-Poli P, Riecher-Rossler A, Borgwardt S. (2012). Neuroimaging and resilience factos-staging of the at-risk mental state? Current pharmaceutical design 18 (4), 416-421

Smieskova R, Marmy J, Schmidt A, Bendfeldt K, Riecher-Rössler A, Walter M, Lang UE, Borgwardt S. (2013) Do subjects at clinical high risk for psychosis differ from those with a genetic high risk? --A systematic review of structural and functional brain abnormalities. *Curr. Med. Chem*, 20(3):467-481.

Steen RG, Mull C, McClure R, Hamer RM, Lieberman JA. (2006) Brain volume in first-episode schizophrenia: systematic review and meta-analysis of magnetic resonance imaging studies. *Br J Psychiatry*, 188:510-518. doi:10.1192/bjp.188.6.510.

Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack CR, Chen K, Ashburner J, Frackowiak RSJ. (2008) Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage*, 39(3):1180-1185. doi:10.1016/j.neuroimage.2007.09.066.

Struyf J, Dobrin S, Page D. (2008) Combining gene expression, demographic and clinical data in modeling disease: a case study of bipolar disorder and schizophrenia. *BMC Genomics*, 9(1):531.

Suckling J, Barnes A, Job D, Brenan D, Lymer K, Dazzan P, Marques TR, MacKay C, McKie S, Williams SR, Williams SCR, Lawrie S, Deakin B (2010) Power calculations for multicenter imaging studies controlled by the false discovery rate. *Hum Brain Mapp*, 31(8):1183-1195. doi:10.1002/hbm.20927.

Sui J, Adali T, Yu Q, Chen J, Calhoun VD. (2012) A review of multivariate methods for multimodal fusion of brain imaging data. *J. Neurosci. Methods*, 204(1):68-81. doi:10.1016/j.jneumeth.2011.10.031.

Sullivan PF, Kendler KS, Neale MC. (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry*, 60(12):1187-1192. doi:10.1001/archpsyc.60.12.1187.

Sun D, van Erp TGM, Thompson PM, Bearden CE, Daley M, Kushan L, Hardt ME, Nuechterlein KH, Toga AW, Cannon TD. (2009a) Elucidating a magnetic resonance

imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol. Psychiatry*, 66(11):1055-1060. doi:10.1016/j.biopsych.2009.07.019.

Sun, J., Maller, J. J., Guo, L., & Fitzgerald, P. B. (2009b). Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. *Brain Research Reviews, 61*(1), 14–32.

Sussmann JE, Lymer GKS, McKirdy J, Moorhead TWJ, Muñoz Maniega S, Job D, Hall J, Bastin ME, Johnstone EC, Lawrie SM, McIntosh AM. (2009) White matter abnormalities in bipolar disorder and schizophrenia detected using diffusion tensor magnetic resonance imaging. *Bipolar Disord*, 11(1):11-18. doi:10.1111/j.1399-5618.2008.00646.x.

Suzuki M, Nohara S, Hagino H, Kurokawa K, Yotsutsuji T, Kawasaki Y, Takahashi T, Matsui M, Watanabe N, Seto H, Kurachi M. (2002) Regional changes in brain gray and white matter in patients with schizophrenia demonstrated with voxel-based analysis of MRI. *Schizophr. Res.*, 55(1-2):41-54.

T

Tandon R, Nasrallah HA, Keshavan MS. (2009) Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophr Res*., 110(1-3):1-23. doi:10.1016/j.schres.2009.03.005.

Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, Malaspina D, Owen MJ, Schultz S, Tsuang M, Van Os J, Carpenter W. (2013) Definition and description of schizophrenia in the DSM-5. *Schizophr. Res.*, 150(1):3-10. doi:10.1016/j.schres.2013.05.028.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior, 8*(2), 153–182. http://doi.org/10.1007/s11682-013-9269-5

Tijms BM, Seriès P, Willshaw DJ, Lawrie SM. (2012) Similarity-based extraction of individual networks from gray matter MRI scans. *Cereb. Cortex*, 22(7):1530-1541. doi:10.1093/cercor/bhr221.

Tofts PS. (1998) Standardisation and optimisation of magnetic resonance techniques for multicentre studies. *J. Neurol. Neurosurg. Psychiatr.*, 64 Suppl 1:S37-43.

Tognin S, Pettersson-Yeo W, Valli I, Hutton C, Woolley J, Allen P, McGuire P, Mechelli A. (2013) Using structural neuroimaging to make quantitative predictions of symptom progression in individuals at ultra-high risk for psychosis. *Front Psychiatry*, 4:187. doi:10.3389/fpsyt.2013.00187.

Torrey EF, Miller J, Rawlings R, Yolken RH. (1997) Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. *Schizophr. Res.*, 28(1):1-38.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F., Etard O., Delcroix N., mazoyer B., Jolio M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273-289. doi:10.1006/nimg.2001.0978.

U

Usall J, Suarez D, Haro JM, SOHO Study Group. (2007) Gender differences in response to antipsychotic treatment in outpatients with schizophrenia. *Psychiatry Res*, 153(3):225-231. doi:10.1016/j.psychres.2006.09.016.

V

van Erp, T. G. M., Hibar, D. P., Rasmussen, J. M., Glahn, D. C., Pearlson, G. D., Andreassen, O. A., Turner, J. A. (2015). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular Psychiatry*. http://doi.org/10.1038/mp.2015.63

Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser MF, Harel N, Heath AC, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen SE, Prior F, Schlaggar BL, Smith SM, Snyder AZ, Xu J, Yacoub E, WU-Minn HCP Consortium. (2012) The Human Connectome Project: a data acquisition perspective. *Neuroimage*, 62(4):2222-2231. doi:10.1016/j.neuroimage.2012.02.018.

Van Haren NEM, Cahn W, Hulshoff Pol HE, Schnack HG, Caspers E, Lemstra A, Sitskoorn MM, Wiersma D, van den Bosch RJ, Dingemans PM, Schene AH, Kahn RS. (2003) Brain volumes as predictor of outcome in recent-onset schizophrenia: a multi-center MRI study. *Schizophr. Res*, 64(1):41-52.

Van Haren NEM, Picchioni MM, McDonald C, Marshall N, Davis N, Ribchester T, Hulshoff Pol HE, Sharma T, Sham P, Kahn RS, Murray R. (2004) A controlled study of brain structure in monozygotic twins concordant and discordant for schizophrenia. *Biol. Psychiatry*, 56(6):454-461. doi:10.1016/j.biopsych.2004.06.033.

Van Haren NEM, Hulshoff Pol HE, Schnack HG, Cahn W, Mandl RCW, Collins DL, Evans AC, Kahn RS. (2007) Focal gray matter changes in schizophrenia across the course of the illness: a 5-year follow-up study. *Neuropsychopharmacology*, 32(10):2057-2066. doi:10.1038/sj.npp.1301347.

Van Haren NEM, Hulshoff Pol HE, Schnack HG, Cahn W, Brans R, Carati I, Rais M, Kahn RS. (2008a) Progressive brain volume loss in schizophrenia over the course of the illness: evidence of maturational abnormalities in early adulthood. *Biol. Psychiatry*, 63(1):106-113. doi:10.1016/j.biopsych.2007.01.004.

Van Haren NEM, Cahn W, Hulshoff Pol HE, Kahn RS. (2008b) Schizophrenia as a progressive brain disease. *Eur. Psychiatry*, 23(4):245-254. doi:10.1016/j.eurpsy.2007.10.013.

Van Hulse J, Khoshgoftaar TM, Napolitano A. (2007) Experimental Perspectives on Learning from Imbalanced Data Proceedings of the 24th International Conference on Machine Learning. ACM, Corvalis, Oregon , pp. 935–942

Vapnik, W. N., and Chervonenkis, A. Y. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*. 17, 264-280.

Vapnik, VN (1995) *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Velakoulis D, Wood SJ, Wong MTH, McGorry PD, Yung A, Phillips L, Smith D, Brewer W, Proffitt T, Desmond P, Pantelis C. (2006) Hippocampal and amygdala volumes according to psychosis stage and diagnosis: a magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. *Arch. Gen. Psychiatry*, 63(2):139-149. doi:10.1001/archpsyc.63.2.139.

Velligan D, Prihoda T, Dennehy E, Biggs M, Shores-Wilson K, Crismon ML, Rush AJ, Miller A, Suppes T, Trivedi M, Kashner TM, Witte B, Toprac M, Carmody T, Chiles J, Shon S. (2005) Brief psychiatric rating scale expanded version: How do new items affect factor structure? *Psychiatry Res*,135(3):217-228. doi:10.1016/j.psychres.2005.05.001.

Venkataraman A, Whitford TJ, Westin C-F, Golland P, Kubicki M. (2012) Whole brain resting state functional connectivity abnormalities in schizophrenia. *Schizophr. Res,* 139(1-3):7-12. doi:10.1016/j.schres.2012.04.021.

Ventura J, Lukoff D, Nuechterlein KH, Liberman RP, Green M, Shaner A. (1993) Training and quality assurance with the brief psychiatric rating scale:__The Drift Busters__; Appendix 1 The Brief Psychiatric Rating Scale (expanded version). *Int J Meth Psychiatric Res,* 3:221–224.

Vita A, De Peri L, Silenzi C, Dieci M. (2006) Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. *Schizophr. Res.*, 82(1):75-88. doi:10.1016/j.schres.2005.11.004.

W

Walker E., Shapiro D., Esterberg M., Tritman H. (2010) Neurodevelopment and Schizophrenia: Broadening the Focus. *Current Directions in Psychological Science*, 19:204

Warner-Schmidt JL, Duman RS. Hippocampal neurogenesis: opposing effects of stress and antidepressant treatment. *Hippocampus* 2006;16(3):239-249. doi:10.1002/hipo.20156.

Watson DR, Anderson JME, Bai F, Barrett SL, McGinnity TM, Mulholland CC, Rushe TM, Cooper SJ. (2012) A voxel based morphometry study investigating brain structural changes in first episode psychosis. *Behav. Brain Res,* 227(1):91-99. doi:10.1016/j.bbr.2011.10.034.

Wei Q, Dunbrack RL. (2013) The Role of Balanced Training and testing data Sets for Binary Classifiers in Bioinformatics. *Plos ONE,* 8(7)

Weinberger DR. (1987) Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry*, 44(7):660-669.

Weinberger, D. R. (1996). On the plausibility of “the neurodevelopmental hypothesis” of schizophrenia. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *14*(3 Suppl), 1S–11S.

Weiss, GM (2004) Mining with Rarity: A Unifying Framework. *SIGKDD Explor. Newsl*, *6*(1), 7–19. doi:10.1145/1007730.1007734

Whalley HC, Simonotto E, Moorhead W, McIntosh A, Marshall I, Ebmeier KP, Owens DGC, Goddard NH, Johnstone EC, Lawrie SM. (2006) Functional imaging as a predictor of schizophrenia. *Biol. Psychiatry*, 60(5):454-462. doi:10.1016/j.biopsych.2005.11.013.

Whalley HC, Simonotto E, Marshall I, Owens DGC, Goddard NH, Johnstone EC, Lawrie SM. (2005) Functional disconnectivity in subjects at high genetic risk of schizophrenia. *Brain*, 128(Pt 9):2097-2108. doi:10.1093/brain/awh556.

Whyte M-C, Brett C, Harrison LK, Byrne M, Miller P, Lawrie SM, Johnstone EC. (2006) Neuropsychological performance over time in people at high risk of developing schizophrenia and controls. *Biol. Psychiatry*, 59(8):730-739. doi:10.1016/j.biopsych.2005.08.028.

Wilke M, Kaufmann C, Grabner A, Pütz B, Wetter TC, Auer DP. (2001) Gray matter-changes and correlates of disease severity in schizophrenia: a statistical parametric mapping study. *Neuroimage*, 13(5):814-824. doi:10.1006/nimg.2001.0751.

Wing JK, Cooper JE, Sartorius N (1974): Measurement and Classification of Psychiatric Symptoms, an Instruction Manual for the PSE and CATEGO Programme. London: Cambridge University Press.

Winklbaur B, Ebner N, Sachs G, Thau K, Fischer G. (2006) Substance abuse in patients with schizophrenia. *Dialogues in Clinical Neuroscience*, 8(1):37-43.

Wood SJ, Velakoulis D, Smith DJ, Bond D, Stuart GW, McGorry PD, Brewer WJ, Bridle N, Eritaia J, Desmond P, Singh B, Copolov D, Pantelis C. (2001) A longitudinal study of hippocampal volume in first episode psychosis and chronic schizophrenia. *Schizophr. Res.*, 52(1-2):37-46.

Wood SJ, Pantelis C, Proffitt T, Phillips LJ, Stuart GW, Buchanan JA, Mahony K,

Brewer W, Smith DJ, McGorry PD. (2003) Spatial working memory ability is a marker of risk-for-psychosis. *Psychol Med*., 33(7):1239-1247.

Woods BT. (1998) Is schizophrenia a progressive neurodevelopmental disorder? Toward a unitary pathogenetic mechanism. *Am J Psychiatry*, 155(12):1661-1670.

World Health Organization. International classification of diseases, 10th edn. Geneva: WHO, 1994.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*(1), 58–73.

Worsley, K. J., Taylor, J. E., Tomaiuolo, F., & Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage*, *23 Suppl 1*, S189–195.

Wright IC, Rabe-Hesketh S, Woodruff PW, David AS, Murray RM, Bullmore ET. (2000) Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatry*., 157(1):16-25.

Y

Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. (2010) A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia. *Front Hum Neurosci*, 4:192. doi:10.3389/fnhum.2010.00192.

Yoon U, Lee J-M, Im K, Shin Y-W, Cho BH, Kim IY, Kwon JS, Kim SI. (2007) Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *Neuroimage*, 34(4):1405-1415. doi:10.1016/j.neuroimage.2006.11.021.

Yoon JH, Nguyen DV, McVay LM, Deramo P, Minzenberg MJ, Ragland JD, Niendham T, Solomon M, Carter CS. (2012) Automated classification of fMRI during cognitive control identifies more severely disorganized subjects with schizophrenia. *Schizophr. Res*, 135(1-3):28-33. doi:10.1016/j.schres.2012.01.001.

Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S, Alzheimer's Disease Neuroimaging Initiative. (2013) Accurate multimodal probabilistic

prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin*, 2:735-745. doi:10.1016/j.nicl.2013.05.004.

Yung AR, McGorry PD, McFarlane CA, Jackson HJ, Patton GC, Rakkar A. (1996) Monitoring and care of young people at incipient risk of psychosis. *Schizophr Bull*, 22(2):283-303.

Yung AR, Phillips LJ, McGorry PD, McFarlane CA, Francey S, Harrigan S, Patton GC, Jackson HJ. (1998) Prediction of psychosis. A step towards indicated prevention of schizophrenia. *Br J Psychiatry Suppl*, 172(33):14-20.

Yung AR, Phillips LJ, Yuen HP, Francey SM, McFarlane CA, Hallgren M, McGorry PD. (2003) Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophr. Res.*, 60(1):21-32.

Yung AR, Phillips LJ, Yuen HP, McGorry PD. (2004) Risk factors for psychosis in an ultra high-risk group: psychopathology and clinical features. *Schizophr. Res.*, 67(2-3):131-142. doi:10.1016/S0920-9964(03)00192-0.

Yung AR, Yuen HP, Berger G, Francey S, Hung T-C, Nelson B, Phillips L, McGorry P. (2007) Declining transition rate in ultra high risk (prodromal) services: dilution or reduction of risk? *Schizophr Bull*, 33(3):673-681. doi:10.1093/schbul/sbm015.

Yung AR, Nelson B, Stanford C, Simmons MB, Cosgrave EM, Killackey E, Phillips LJ, Bechdolf A, Buckby J, McGorry PD. (2008) Validation of "prodromal" criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. *Schizophr. Res.*, 105(1-3):10-17. doi:10.1016/j.schres.2008.07.012.

Z

Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, Davatzikos C. (2009) Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med*, 62(6):1609-1618. doi:10.1002/mrm.22147.

Zakzanis KK, Poulin P, Hansen KT, Jolic D. (2000) Searching the schizophrenic brain for temporal lobe deficits: a systematic review and meta-analysis. *Psychol*

Med., 30(3):491-504.

Zanetti MV, Schaufelberger MS, Doshi J, Ou Y, Ferreira LK, Menezes PR, Scazufca M, Davatzikos C, Busatto GF. (2013) Neuroanatomical pattern classification in a population-based sample of first-episode schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, 43:116-125. doi:10.1016/j.pnpbp.2012.12.005.

Zhang T, Koutsouleris N, Meisenzahl E, Davatzikos C. (2015) Heterogeneity of structural brain changes in subtypes of schizophrenia revealed using magnetic resonance imaging pattern analysis. *Schizophr Bull*, 41(1):74-84. doi:10.1093/schbul/sbu136.

Zipursky RB, Lambe EK, Kapur S, Mikulis DJ. (1998) Cerebral gray matter volume deficits in first episode psychosis. *Arch. Gen. Psychiatry*, 55(6):540-546.

Appendices

Appendix I

This section presents a list of first-author paper and abstract conferences publications.

E. Zarogianni et al. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. *Neuroimage Clinical* . 2013;3:279-289.

E. Zarogianni et al. Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. *Neuroimage*. *Under review*

E. Zarogianni et al. Early diagnosis in the at-risk mental state for psychosis: application to the FePsy study. *In preparation*

Zarogianni E, Moorhead TWJ, Strokey AJ, Lawrie SM. Improved individualised prediction of schizophrenia in subjects at high genetic risk, based on neuroanatomical and clinical data. Presented at the Schizophrenia International Research Society Biennial Meeting in 2014.

Zarogianni E, Moorhead TWJ, Strokey AJ, Lawrie SM. Combining neuroanatomical and clinical data to improve individualized early diagnosis of schizophrenia in subjects at high familial risk. Conference Abstract in IEEE, presented at the Pattern Recognition in Neuroimaging in 2014

Appendix II

Derivation of the dual form of SVM

The primal formulation of the optimization problem is given by:

Minimize $\frac{1}{2}\|\vec{w}\|^2$ subject to $y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0 \quad \forall_i$, for i=1, …, N

Define Lagrangian: $\Lambda_p(\vec{w}, b, \vec{a}) = \frac{1}{2}\sum_{i=1}^{n} w_i^{\,2} - \sum_{i=1}^{N} a_i \left(y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \right)$

We need to minimize this Lagrangian with respect to $\vec{w}$ and b and simultaneously require that the derivative with respect to $\vec{a}$ vanishes , all subject to the constraints that $a_i \geq 0$.

If we set the derivatives with respect to $\vec{w}$ and b to 0, we obtain:

- $\frac{\partial \Lambda_p(\vec{w}, b, \vec{a})}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} a_i y_i = 0$
- $\frac{\partial \Lambda_p(\vec{w}, b, \vec{a})}{\partial \vec{w}} = 0 \Rightarrow \vec{w} = \sum_{i=1}^{N} a_i y_i \vec{x}_i$

We substitute the above into the equation $\Lambda_p(\vec{w}, b, \vec{a})$ and obtain the dual formulation of linear SVM:

$$\Lambda_D(\vec{a}) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j a_i a_j \vec{x}_i^{\,T} \cdot \vec{x}_j$$

We seek to maximize the above Lagrangian with respect to $\vec{a}$, subject to the constraints $a_i \geq 0$ and $\sum_{i=1}^{N} a_i y_i = 0$.

Appendix III

- The F-score (Fisher score) is a simple and effective criterion to measure the discrimination between a feature and the label. Based on statistic characteristics, it is independent of the classifiers. Following Chen and Lin (2006), a variant of F-score is used. Given training instances $\vec{x}_i, i = 1,...,l$, the F-score of the jth feature is defined as:

$$F(j) = \frac{(\vec{x}_j^{(+)} - \vec{x}_j)^2 + (\vec{x}_j^{(-)} - \vec{x}_j)^2}{\frac{1}{n_+ - 1}\sum_{i=1}^{n_+}(x_{i,j}^{(+)} - \vec{x}_j^{(+)})^2 + \frac{1}{n_- - 1}\sum_{i=1}^{n_-}(x_{i,j}^{(-)} - \vec{x}_j^{(-)})^2}$$

Where n_+ and n_- are the number of positive and negative instances, respectively; $\vec{x}_j$, $\vec{x}_j^{(+)}$, $\vec{x}_j^{(-)}$ are the average of the jth feature of the whole, positive-labeled and negative-labeled data sets; $x_{i,j}^{(+)} \diagup x_{i,j}^{(-)}$ is the jth feature of the ith positive/negative instance. The numerator denotes the inter-class variance, while the denominator is the sum of thye variance within each class. A larger F-score indicated that the feature is more discriminative.

Although, the F-score is easy to implement, its deficiency, however, lies in that it cannot reveal any mutual information between features and only considers each feature separately.

- the Pearson correlation coefficient, whose use in feature selection is closely related to that of the t-Test (Guyon and Elisseef 2003), measures the relevance of each feature to the classification task. The larger the absolute

value of the Pearson correlation coefficient, the more relevant the feature to the classification. Given a location, u, in the template space, the Pearson correlation coefficient between a feature, $f^i(u)$, of tissue i and class label y is defined as:

$$\rho^i(u) = \frac{\sum_j (f_j^i(u) - \overline{f^i(u)})(y_j - \bar{y})}{\sqrt{\sum_j (f_j^i(u) - \overline{f^i(u)})^2 \sum_j (y_j - \bar{y})^2}}$$

where *j* denotes the *j*th sample in the training dataset. Thus, $f_j^i(u)$ is a morphological feature of tissue *i* in the location *u* of jth sample, and $\overline{f^i(u)}$ is the mean of $f_j^i(u)$ over all samples. Similarly, y_j is a class label (controls -1 or patients +1) of the *j*th sample, and $\bar{y}$ is the mean of y_j over all samples.

Appendix IV

Table 1. A comparative presentation of the PSE and BPRS subscales

PSE sub-categories	**BPRS subcategories**
-Situational Anxiety	-Somatic concern
-Nervous Tension	-Anxiety
-Depression	-Depression
-Mania	-Suicidality
-Overactivity	-Guilt
-Incoherence/Disorganization	-Hostility
-Changed Perception	-Elevated mood
-Hallucinations	-Grandiosity
-Thought Disturbance	-Suspiciousness
-Delusional Construction	-Hallucinations
-Outside Control	-Unusual thought content
-Negative Symptoms	-Bizarre behavior
	-Self-neglect
	-Disorientation
	-Conceptual disorganization
	-Blunted affect
	-Emotional withdrawal
	-Motor retardation
	-Tension
	-Uncooperativeness
	-Excitement
	-Distractibility
	-Motor hyperactivity
	-Mannerism and posturing

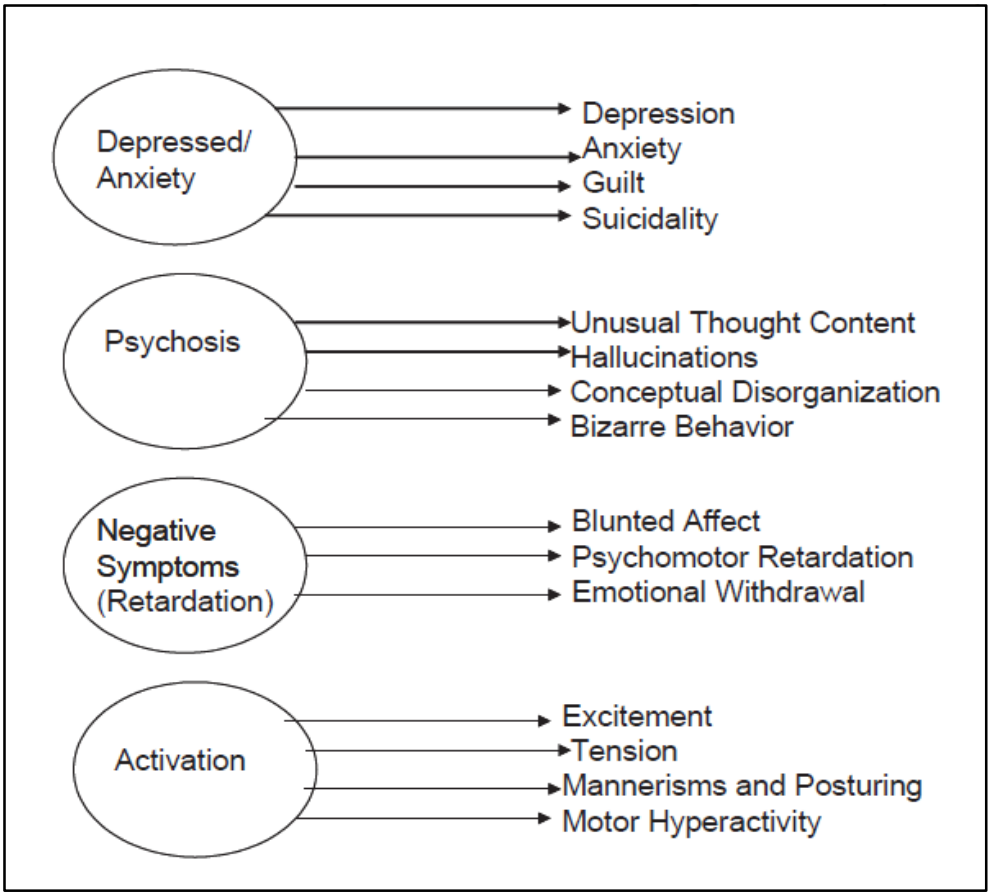


Figure 1. Factorial analysis of the BPRS subscores and the derived four-scale model. Figure adapted from Velligan et al. 2005

Appendix V

Here, I present the histograms, generated after concatenating all voxel intensity values of the smoothed GM segments of all scans from the EHRS and the FePsy data set.

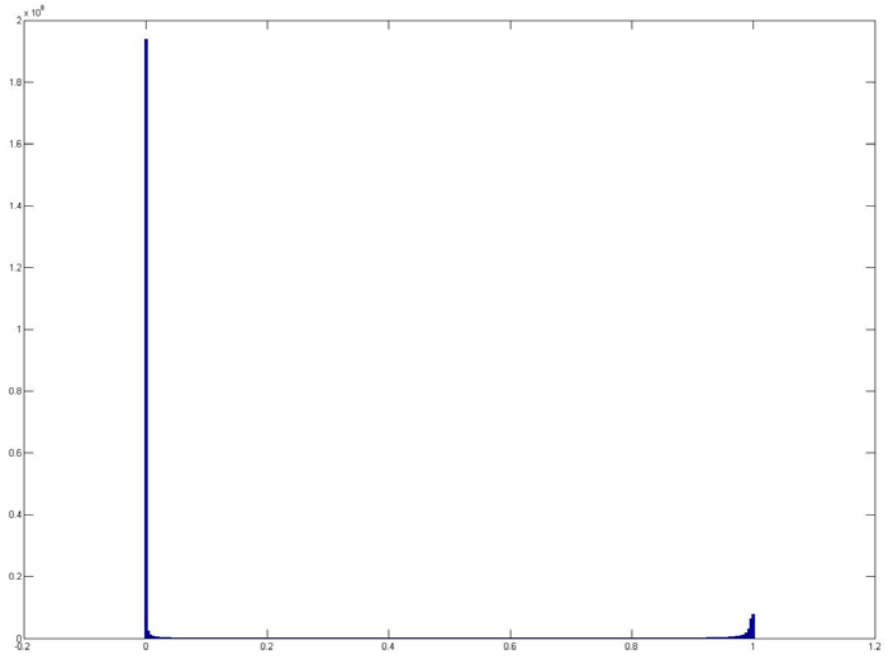


Figure 1. Histogram of the EHRS data set

It should be noted that values in the x-axis correspond to probabilities of the voxels being GM, and the y-axis represent the number of voxels having the relevant probabilities. The only noticeable difference between the histograms of the EHRS and the FePsy scans (Figure 1 and 2 accordingly) lies in the assignment of higher probability values in the EHRS data set relative to the FePsy one (the far-right side of the distributions). However, this can not be attributed to any difference in the preprocessing stage, since the pre-processing pipeline was exactly the same for the

two data sets but can be reasonably due to a much more accurate segmentation of the scans in the EHRS set.

In this context, given the similarity of the distributions between the EHRS and the FePsy data sets, it was safe to assume that pooling together the data from the two data sets without performing any type of scaling or histogram matching would work fine and would not introduce much variability in the context of scanner-related differences.

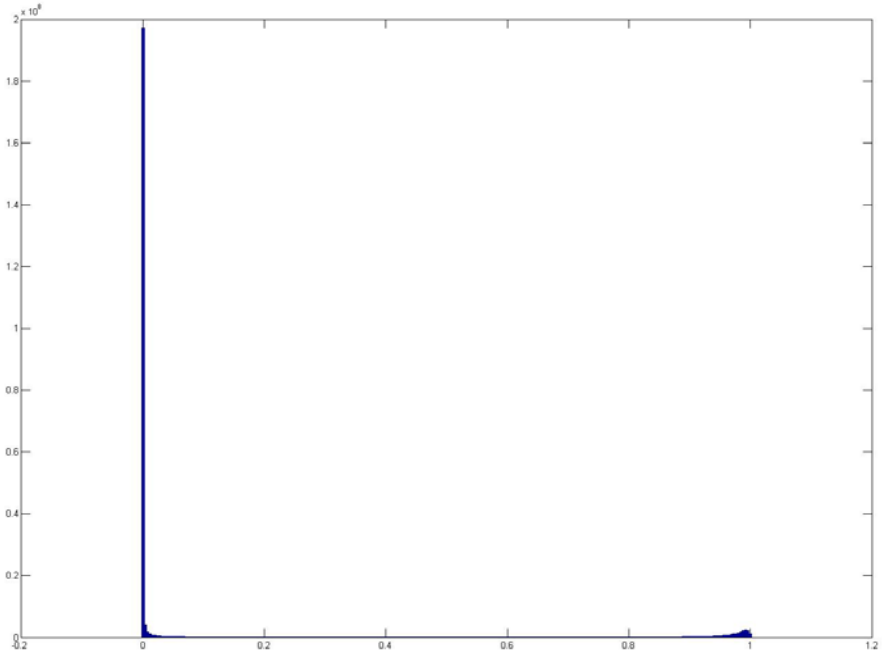


Figure 2. Histogram of the FePsy data set

It should also be noted here, that even in the case where the distributions between the data sets were different, it would be most reasonable to perform some type of scaling or matching before any pre-processing is done and try to match all tissue types, that is GM, WM and CSF, and not just GM. However, this endeavour is quite complex and could be a project on its own and that is why it was decided not to attempt this and stick to pooling all of the available data without performing any scaling whatsoever.