NEUROINFORMATICS AND NEUROIMAGING-BASED SCHIZOPHRENIA MODELING AND DECISION SUPPORT

YANG GUO LIANG

NATIONAL UNIVERSITY OF SINGAPORE

2010

NEUROINFORMATICS AND NEUROIMAGING-BASED SCHIZOPHRENIA MODELING AND DECISION SUPPORT

YANG GUO LIANG

(Msc. CS, National University of Singapore)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor **A/Prof. Poh Kim Leng** (National University of Singapore) for his continuous guidance in decision support theories, modeling technologies and research directions, especially many helpful feedbacks and comments to my work results; to my supervisor **Prof. Wieslaw Lucjan Nowinski** (Biomedical Imaging Lab, Singapore Biomedical Consortium, Agency of Science, Technology and Research, Singapore) for helping me to identify and evaluate the research topics, as well as his continuous encouragement, support and valuable suggestions in many aspects, especially many very detailed and general comments on my thesis. Without their help, this work would not be able to be done in the correct direction.

I would also like to show my appreciation to:

- **Dr. Sim Kang** (Psychiatrist, Institute of Mental Health, Singapore) for helping me in acquiring medical domain knowledge in schizophrenia, and providing medical images and clinical data, comments on standard schizophrenia diagnostic procedures, clinical significance of imaging findings, and invaluable feedback on my results. This work is inspired by a research project in the parietal lobe changes in schizophrenia with passivity, where he is the principal investigator.

- **Dr. Sitoh Yih Yian** (Neuroradiologist, National Neuroscience Institute, Singapore) for explaining to me the imaging protocols and parameters and providing the data about the time and costs involved in the scanning.
- **Dr. Tchoyoson Lim Choie Cheio** (Neuroradiologist, National Neuroscience Institute, Singapore) for helping me in understanding the clinical importance of the relevant brain structures as well as clarifying many expression ambiguities.
- **Dr. Elie Cheniaux** (Psychiatrist, Institute of Psychiatry, Federal University of Rio de Janeiro, Brazil) for his comments on the schizophrenia diagnostic procedures.
- **Dr. Aamer Aziz** (Radiologist, Charles Sturt University, Australia) for helping me in reviewing the thesis.
- **Dr. Li Guo Liang** (Genome Institute of Singapore, Agency of Science, Technology and Research, Singapore) for helping me with acquiring knowledge in Bayesian Networks learning technology and his suggestions in decision support system presentation formats.
- **Mr. Chan Wai Yen** (Institute of Mental Health, Singapore) for helping me in understanding bio-statistics concepts and methods, understanding the meaning of various neurocognitive tests, collecting necessary neuroinformatics data, and verifying the data, as well as many discussions on methods of neuroinformatics data analysis.
- **Dr. Varsha Gupta** (Biomedical Imaging Lab, Singapore Biomedical Consortium, Agency of Science, Technology and Research, Singapore) for her suggestions on decision support system performance measurements.

- **Dr. Liu Ji Min** (Biomedical Imaging Lab, Singapore Biomedical Consortium, Agency of Science, Technology and Research, Singapore) for reviewing the thesis and his many useful suggestions and criticisms.
- **Dr. Bhanu Prakash K. N.** (Biomedical Imaging Lab, Singapore Biomedical Consortium, Agency of Science, Technology and Research, Singapore) for sharing his experience in his PhD work of fetus abnormality modeling using artificial neural networks, and his encouragement to me.
- **Ms. Ow Lai Chun** (National University of Singapore) for her helpful and prompt replies and reactions to all my queries on administrative issues, such as module registration and exemption procedures, research progress reporting, thesis formats and submitting procedure.

Finally I would like to thank my wife **Yang Yi Li** for her hearty support, and great patience and love throughout the whole course of my study; and my son and daughter for bringing me the joys and courage.

Table of Contents

Summary

Purpose: Schizophrenia is a common psychiatric disease of impaired perception or expression of reality. However the etiology of this disease is still not clear after it has been identified for over 100 years, and the current standard schizophrenia diagnostic procedures are based on subjective observations on symptoms. We aimed to discover the relationship between schizophrenia and the objective and quantitative criteria from neuroinformatics data and neuroimaging data, and construct schizophrenia classification models based on this unique combination of data. This novel approach of combining neuroinformatics and neuroimaging for schizophrenia modeling, to our best knowledge, had never been used before by others.

Study Subjects and Methods: With the support from the National Healthcare Group Research Grant (NHG-SIG/05004) and Singapore Bioimaging Consortium Research Grant (SBIC RP C-009/2006), our collaborating hospitals, Institute of Mental Health, Singapore and National Neuroscience Institute, Singapore, recruited 156 study subjects (92 schizophrenia patients, 64 healthy controls). Various types of neuroinformatics data (including demographic data, clinical information, clinical scores, and neurocognitive test results) and neuroimaging data (Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI)) were collected.

A subset of study subjects consisting of 84 cases (59 patients and 25 controls) was used as training dataset for modeling. Significant features were selected from over 300 data items. Bayesian Network learning technologies were applied to construct various Bayesian Network models for the classification of schizophrenia patients and normal controls using the selected features. The 10-fold cross-validation method was used for internal model validation. Limited external validation was also performed using the test dataset.

Results: The following eight factors were chosen by the feature selection process: 1) Family history of psychiatric diseases, 2) Raven's Progressive Matrices (RPM) test result (RPM raw score), 3) Wechsler Adult Intelligence Scale (WAIS) test result (Digit Span backward score), 4) Wisconsin Card Sorting Test (WCST) result (Perseverative Responses raw scores), 5-8) Mean Fractional Anisotropy (FA) values in four brain structures from neuroimaging results: cingulate gyrus, left subcallosal gyrus, left thalamus: lateral dorsal nucleus, and right thalamus: anterior nucleus.

The classification accuracies of models built on clinical information (family history) plus various combinations of neurocognitive tests (but no neuroimaging features) ranged from 75% to 85.7%. On the other hand, the accuracy of the model on neuroimaging features alone was 77.4%, and the accuracy of model on clinical information and neuroimaging features (but no neurocognitive test) was 84.5%. Models built on clinical information and neuroimaging features plus various combinations of neurocognitive test further increased accuracy to 85.7%-89.3%.

The most comprehensive model consisted of all eight significant factors. The accuracy of this model, 89.3%, was the highest among all models.

Contributions: By applying the first ever Talairach brain atlas based FA image quantification method developed at Biomedical Imaging Lab, Agency for Science, Technology and Research, Singapore, we placed a large amount of Region of Interests (144 ROIs for 48 brain structures) on brain images, and quantified their image features (mean and standard deviation of FA values) automatically, which was usually difficult for manual methods. This method made studies involving large amount of patients/controls more consistent and feasible than the manual processing. The quantified image features have been used in further model constructions and decision support.

We found that schizophrenia was highly related to a person's family history of psychiatric disease, deficit in eductive and reproductive functions, deficit in verbal working memory, undue perseverative responses (which is caused by frontal lobe deficit), reduced neural connectivity in the cingulate gyrus (which is associated with attention function), the subcallosal gyrus (which is associated with the left and right prefrontal interhemispheric communication), and the thalamus lateral dorsal nucleus and anterior nucleus (which are associated with somatosensory and visuo-spatial function and modulation of alertness).

We demonstrated the first ever schizophrenia classification models based on objective and quantitative criteria including neurocognitive tests and neuroimaging. These models quantified the relationships between schizophrenia and the relative

factors, which helped us to achieve a better understanding and management of the disease.

Based on our schizophrenia classification models, we made two Decision Support Flow Charts to choose suitable tests by using different strategies: the highest accuracy gain, and the highest cost effectiveness. These flow charts could help clinicians to choose the best further tests in order to achieve a higher diagnostic accuracy with or without cost consideration.

We also developed decision support system software for schizophrenia diagnosis. This software could classify a person as either a schizophrenia patient or healthy (together with probability distribution), using the given clinical information, and the neurocognitive and neuroimaging test results. It could also provide suggestions on what further tests should be done in order to improve the diagnosis accuracy.

The methodology (modeling using neuroinformatics and neuroimaging) we developed in this study has the potential to be applied to other diseases with informatics and imaging data.

Conclusions: Schizophrenia classification models can be constructed using objective and quantitative criteria from neuroinformatics and neuroimaging data. The classification accuracy of the most comprehensive model consisting of all eight significant features is 89.3%. These models reveal the quantitative relationship between schizophrenia and various intermediate phenotypes (as assessed by neurocognitive tests) and brain abnormalities (as assessed by

neuroimaging). A decision support system based on these models can provide additional evidence to clinicians and augment the current schizophrenia diagnostic procedures, which may help to improve the diagnosis accuracy.

The approach described in this thesis for the schizophrenia modeling and decision support can also be applied to other mental sickness such as schizoaffective disorder, bipolar disorder or unipolar depression, where neurocognitive tests and neuroimaging test are used.

Despite our data uniqueness, our models and decision support system are still tentative and limited due to the relatively small sample size and types of data. Even for the most comprehensive model including all eight features, there is a noticeable false positive rate (normal control classified as patient) of 20%. Further refinements need to be considered by recruiting more study subjects, using more extensive clinical and biological information (such as genetic data).

Keywords: neuroimaging, neuroinformatics, neurocognitive test, schizophrenia, decision support, Bayesian Network, classification model, MRI, DTI

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
AC	Anterior Commissure
BIF	Bayesian Interchange Format
CCTCC	Cortico-Cerebellar-Thalamic-Cortical Circuit
COMT	Catechol-O-methyl Transferase
CNS	Central Nervous System
CPT	Continuous Performance Task (or Test)
CT	Computer Tomography
DAG	Directed Acyclic Graph
DISC1	Disrupted-in-Schizophrenia 1
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTI	Diffusion Tensor Imaging
DTNBP1	Dystrobrevin-Binding Protein 1
DWI	Diffusion Weighted Imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
fMRI	Functional Magnetic Resonance Imaging
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
FTT	Fast Talairach-Transformation
GAF	Global Assessment of Functioning Scale
GUI	Graphical User Interface
HAM-D	Hamilton Rating Scale for Depression
ICD	International Statistical Classification of Diseases and Related Health Problems
ID3	Iterative Dichotomiser 3

Acronym	Meaning
IM	Inferior Midway
MD	Mean Diffusivity
MP-RAGE	Magnetisation-Prepared Rapid Acquisition with a Gradient Echo
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
NS	Not Significant
PANSS	Positive and negative Syndrome Scale
PC	Posterior Commissure
PEG	Pneumoencephalography
QOL	Quality of Life
ROI	Region of Interest
RPM	Raven's Progressive Matrices
SAPP	Scale for the Assessment of Passivity Phenomena
SCID	Clinical Interview for DSM Disorders
SD	Standard Deviation (stdev)
SIG	Significant
SM	Superior Midway
sMRI	Structural Magnetic Resonance Imaging
SUMD	Scale to Assess Unawareness of mental Disorders
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
VBM	Voxel Based Morphometry
WAIS	Wechsler Adult Intelligence Scale
WCST	Wisconsin Card Sorting Test
WHO	World Health Organization
WHOQOL-BREF	World Health Organization Quality of Life Bref-Scale

List of Notations

Notation	Description
Acc_m	accuracy of model *m*
$Acc_{overall}$	overall accuracy of decision support system
Accuracy	classification accuracy
$CE_{t,m}$	cost effectiveness of test *t* from model *m*
Cor	number of correctly diagnosed cases
Cor_i	number of correctly classified cases using model *i*
$Cost_t$	cost of test *t*
D	apparent diffusion tensor
D_{xx}, D_{xy}, D_{zz}	diffusion fluxes along x, y, and z directions
D_{xy}, D_{xz},	correlations between diffusion fluxes in orthogonal directions
D_{yx}, D_{yz},	
D_{zx}, D_{zy}	
FA	fractional anisotropy of the diffusion tensor
F_i	i^{th} factor (node) in the Bayesian network
FNR	false negative rate
FPR	false positive rate
M	classification model
MD	mean diffusivity of the diffusion tensor
Nr	total number of cases
Nr_i	total number of cases used by model *i*
$P_{dist}(v)$	distribution probability of patient or control
P_{prev}	prevalent patient probability
$pt\_ctrl_{classify}$	classification result of a case
RC	relative cost
$RC_{overall}$	overall relative cost of decision support system
$RC_{t,m}$	relative cost of test *t* from model *m*
t	test (neurocognitive test or neuroimaging test)

Notation	Description
TNR	true negative rate
TPR	true positive rate
u_i	possible value for a F_i (i^{th} Factor)
v	value of classification, patient or control
$\lambda_1, \lambda_2, \lambda_3$	eigenvalues of diffusion tensor along three principal directions

Chapter 1

Introduction

In this chapter, we will introduce some background knowledge of schizophrenia disease and the difficulties in its diagnosis. We will also propose our approach towards a better understanding of schizophrenia, and an alternative way to the current diagnostic procedures by using objective and quantitative criteria.

1.1 Schizophrenia

Schizophrenia is a common psychiatric disease of impaired perception or expression of reality, commonly demonstrated through disorganized speech and thinking, auditory hallucinations, delusions, or paranoid. It affects about one percent of the world population, regardless of societies and geographical areas. It usually starts in late adolescence and young adulthood, and can last for the whole life (Sadock BJ, 2003). Schizophrenia patients have severe suffering; 30% of them have attempted suicide (Radomsky, Haas, Mann, & Sweeney, 1999), and about 10% of them die by suicide (Caldwell & Gottesman, 1990).

Schizophrenia affects patients' normal mental functions and behaviors. Most likely, patients could not continue their work or study.

Schizophrenia becomes an enormous economic burden to the patients' family and the society. It is ranked the ninth in the global burden of disease (C. Murray & Lozpe, 1996). For example, the total expenses including inpatient, outpatient, primary care, pharmaceutical, and long-term care, were estimated at US$62.7 billion in year 2002 in the United State of America (Wu, et al., 2005); and the total societal cost of schizophrenia was estimated at £6.7 billion in 2004/05 in the United Kingdom of Great Britain and Northern Ireland (Mangalore & Knapp, 2007).

History

The study of schizophrenia can be traced back to 19th century. An Austrian-French physician, Benedict Augustin Morel (1809-1873) used *demence precoce* for deteriorated patient with illness beginning in adolescence. Emil Kraepelin (1856-1926), a German psychiatrist, translated it into *dementia praecox*, which distinguishes cognitive process (*dementia*) and early onset (*praecox*). Patients having dementia praecox were classified as having long-term deterioration in addition to hallucinations and delusions.

Paul Eugen Bleuler (1857-1939), a Swiss psychiatrist, started to use *schizophrenia* to express the schisms among thoughts, emotions and behaviors of the patients. Since schizophrenia comes from two roots, schizo (meaning split) phrenia (meaning mind), it is often confused with split personality (*dissociative identity disorder*) by laymen (Sadock BJ, 2003).

Symptoms

Patients with schizophrenia can show positive and/or negative symptoms. Positive symptoms include hallucinations (including auditory and visual), delusions (such as grandiose: e.g., feeling himself/herself as a great movie star, and delusion of being controlled, or passivity: feeling himself/herself being controlled by an external party), bizarre behavior (e.g., wearing odd or inappropriate makeup), and positive formal thought disorder (such as derailment: ideas slipping off the track onto another which is obliquely related or unrelated; tangentiality: replying to questions in an oblique, tangential or irrelevant manner).

Negative symptoms include affective flattening (reduction in the range and intensity of emotional expression), alogia (difficulty or inability to speak), avolition-apathy (reduction, difficulty, or inability to initiate and persist in goal-directed behavior: e.g. no longer interested in going out and meeting with friends), and inattentiveness (difficulty concentrating or focusing). Table 1.1 lists the symptoms of schizophrenia patients and divides them into positive and negative groups.

Table 1.1 Positive and negative symptoms of schizophrenia patients

Positive Symptoms	Negative Symptoms
Hallucination	Affective flattening
Auditory	Unchanging facial expression
Voice commenting	Decreased spontaneous movements
Voice conversing	Paucity of expressive gesture
Somatic-tactile	Poor eye contact
Olfactory	Affective nonresponsivity
Visual	Inappropriate affect
Delusion	Lack of vocal inflections
Persecutory	Alogia
Jealousy	Poverty of speech
Guilt, sin	Poverty of content of speech
Grandiose	Blocking
Religious	Increased response latency
Somatic	Avolition-apathy
Delusion of reference	Grooming and hygiene
Delusion of being controlled	Impersistence at work or school
Delusion of mind reading	Physical anergia
Thought broadcasting	Anhedonia-asociality
Thought insertion	Recreational interests, activity
Though withdrawal	Sexual interest, activity
Bizarre behavior	Intimacy, closeness
Clothing, appearance	Relationship with friends, peers
Social, sexual behavior	Attention
Aggressive/agitated behavior	Social inattentiveness
Repetitive/stereotyped behavior	Inattentiveness during testing
Positive formal thought disorder	
Derailment	
Tangentiality	
Incoherence	
Illogicality	
Circumstantiality	

Positive Symptoms	Negative Symptoms
Pressure of speech Distractible speech Clanging	

In this section, we briefly introduced a very common (affects 1% of population) and economically costly psychological disease, schizophrenia, its history, and its major symptoms, which can be divided into positive and negative groups. However, being such an important disease with long history (more than 100 years), its diagnosis problem is not yet solved satisfactorily, as we can see from the next section.

1.2 Diagnosis of Schizophrenia

Schizophrenia diagnosis is based on the patient's self-reported experiences, and family members', friends', and clinicians' observed behavior. There is no laboratory test for schizophrenia yet.

In 1994, American Psychiatric Association published the Diagnostic and Statistical Manual of Mental Disorder, 4th Edition (DSM-IV), which recommended the following diagnostic criteria for schizophrenia:

- Characteristic symptoms. Two or more of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated)

    - Delusions (e.g., delusion of grandeur: believing he/she is someone very famous or important, such as God)
    - Hallucinations (e.g., visual: seeing something nobody else can see; auditory: hearing things nobody else can hear)
    - Disorganized speech (e.g., frequent derailment, incoherence)
    - Grossly disorganized (e.g., shouting or cursing in public) or catatonic behavior (e.g., rapid alteration between extreme excitement and stupor)
    - Negative symptoms (e.g., affective flattening, alogia)
- Social/occupational dysfunction
- Duration. Continuous signs of the disturbance persist for at least six months.
- Schizoaffective and mood disorder exclusion
- Substance/general medical condition exclusion (disturbance not due to the direct physiologic effects of a substance or general medical condition)
- Relationship to pervasive developmental disorder

International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) (World Health Organization 2006) presents another guideline for diagnosis of schizophrenia. According to ICD-10, the most important psychopathological phenomena include: 1) thought echo, thought insertion or withdrawal, 2) thought broadcasting, 3) delusional perception and delusions of control, 3) influence or passivity, 4) hallucinatory voices commenting or discussing the patient in the third person, 5) thought disorders, and 6) negative symptoms. The duration of symptoms presenting clearly should be at least 1 month.

Schizophrenia should not be diagnosed in extensive depressive or manic symptoms unless it is clear that schizophrenic symptoms antedate the affective disturbance. It shall not be diagnosed in the presence of overt brain disease or during states of drug intoxication or withdrawal.

Schizophrenia can have different subtypes. Subtypes of schizophrenia can be identified by the most predominant and significant symptoms for each patient at the evaluation time. For example, according to DSM-IV, there are five subtypes:

- Catatonic type: when prominent catatonic symptoms are present.
- Disorganized type: when disorganized speech and behavior and flat or inappropriate affect are present.
- Paranoid type: when preoccupation with delusion or frequent hallucinations are prominent.
- Undifferentiated type: a remaining category describing prominent active-phase symptoms that are not catatonic, disorganized or paranoid types.
- Residual type: continuing evidence of disturbance but not meeting active-phase symptoms criteria.

Subtypes of schizophrenia are not mutually exclusive. Sometimes, patients may develop more than one subtypes of schizophrenia. For example, a patient may be in both catatonic and paranoid subtypes, if neither subtype trumps another significantly.

Patients' predominant symptoms may change at different stage of the disease. Hence, patients' subtype may also change over time.

1.3 Treatment and Prognosis of Schizophrenia

In this section, we will introduce the current treatment options of schizophrenia, and their prognosis.

Treatment

Treatment of schizophrenia patients needs to be comprehensive since this disorder affects many aspects of the patients, including thinking, feeling and behavior. Treatment plans should be customized to suit the individual patient's clinical status, and stages (acute stage - a period of intense psychotic symptoms; stabilization stage - a period of suffering from psychotic symptoms but less severe than in the acute stage; stable stage - severe symptoms are controlled by medication). And goals will need to evolve over time. Treatment should be continuous since schizophrenia usually affects the patient's whole life time (Herz MI, 2002).

Currently the following treatment methods are used (Sadock BJ, 2003):

- Hospitalization
- Biological therapy, including dopamine receptor antagonist, serotonin-dopamine antagonist (Resperidone, Clozapine, Olanzapine, Sertindole, Quetiapine, Ziprasidone), other drugs (Lithium, Anticonvulsants, Benzodiazepines), and other biological therapies (Electroconvulsive therapy (ECT))

- Psychosocial therapy (social skill training, family oriented therapy, case management, assertive community treatment (ACT), group therapy, cognitive behavioral therapy, individual psychotherapy)
- Vocational therapy

Prognosis

(C. M. Harding, Brooks, Ashikaga, Strauss, & Breier, 1987) reported one-half to two-third of schizophrenia patients had achieved considerable improvement or recovery in a long term retrospective follow-up study of 118 patients. However, another study on 118 (coincidently) schizophrenia or schizoaffective (a mental disorder that has symptoms of schizophrenia and affective disorder - either major depression or bipolar disorder) patients by (Robinson, Woerner, McMeniman, Mendelowitz, & Bilder, 2004) reported a much lower recovery rate of 13.7% when stricter criteria of full recovery were used, i.e. sustained improvement in both symptoms and social and vocational functioning.

(Lieberman, et al., 1996) and (Davidson & McGlashan, 1997) found that being female, being older at the first episode, having acute symptoms, having predominantly positive symptoms, and having good premorbid functioning are correlated to better prognosis.

Although currently, there are many different treatment plans (hospitalization, biological, psychological and vocational), their outcomes are not effective enough.

Merely 13.7% patients have sustained improvement in symptoms and functions. That is because the direct cause of schizophrenia is still unknown.

1.4 Motivations and Objectives

The definite etiology of schizophrenia is still not clear though the disorder has been identified over 100 years. Studies suggest that genetics, early environment, neurobiology and psychological and social processes are important contributory factors.

Many epidemiological studies have established a set of risk factors of schizophrenia. (R. Murray, Jones, Susser, Os, & Cannon, 2003) summarized 18 factors and their odds ratios (odds ratio: ratio of a factor occurring in schizophrenia patients to non-schizophrenia people). All these factors are grouped into the following 4 categories:

- Place/time of birth
    1. Winter
    2. Urban
- Infection
    3. Influenza
    4. Respiratory
    5. Rubella
    6. Poliovirus
    7. Central Nervous System (CNS)

- Prenatal
    8. Famine
    9. Bereavement
    10. Flood
    11. Unwantedness
    12. Maternal Depression
- Obstetric
    13. Rh (rhesus) factor incompatibility
    14. Hypoxia
    15. Central Nervous System (CNS) damage
    16. Low birth weight
    17. Pre-eclampsia
    18. Family history

Among them, family history has the greatest odds ratio of close to 10, followed by Central Nervous System(CNS) damage, prenatal bereavement, and rubella infection with odds ratios ranging from 5 to 7. All the rest factors have odd ratios from 1 to 4.

Modern neuroscientific studies including molecular genetics, molecular neuropathology, neurophysiology, various brain imaging, and psychopharmacology have suggested that we are now approaching the molecular basis of the disorder. Schizophrenia can be identified as a neurodevelopmental and progressive disease, which is associated with multiple biochemical abnormalities

involving dopaminergic, serotonin, glutamate, and γ-aminobutyric acidergic system (Miyamoto, et al., 2003).

(Andreasen, 2000) lists the hypotheses about the etiology of schizophrenia made by researchers as the following:

- "Hypothesis 1. The etiologies are multiple.
- Hypothesis 2. The pathophysiology is an abnormality in the regulation and expression of neurodevelopment.
- Hypothesis 3. The pathology is a disease of neuroconnectivity.
- Hypothesis 4. The phenotype is defined by a mental metaprocess rather than by clinical symptoms."

1.4.1 Problems with Existing Diagnostic Procedures

From the above introduction, we notice that the current standard procedures for diagnosing schizophrenia (DSM-IV and ICD-10) have the following problems:

Symptom-based: DSM-IV and ICD-10 diagnostic criteria are based on heterogeneous symptoms. Most symptoms are from patient's self reporting, family member's, colleague's and clinician's observations, which are subjective (A sample page for interviewing question regarding to the delusion symptoms from the Structured Clinical Interview for DSM Disorders (SCID) can be found in the web link from (SCID-I, 2007)). One common criticism of the diagnosis of schizophrenia is the lacking of scientific validity or reliability (Bentall, 1992;

Boyle, 2002). (Tsuang, Stone, & Faraone, 2000) argued that psychotic symptoms were not a good basis for schizophrenia diagnosis.

Not quantifiable: DSM-IV and ICD-10 diagnostic criteria do not have quantification components. For example the severity (degree) of delusion or hallucination is difficult to quantify.

Low/moderate diagnosis agreement: Studies show that the reliability of schizophrenia diagnosis is typically relatively low. (McGorry, et al., 1995) reported that agreement between any two psychiatrists was 66% to 76% when diagnosing schizophrenia. This converts to misdiagnosis rate of 23%-34%, assuming one psychiatrist is always correct. This misdiagnosis may have harmful clinical effect on patients.

A moderate agreement between two psychiatrists is observed by a more recent study (Cheniaux, Landeira-Fernandez, & Versiani, 2009). 100 patients are diagnosed by two psychiatrists using DSM-IV and ICD-10 procedures separately. According to DSM-IV, 39 patients received schizophrenia diagnosis; among them only 13 patients (or 33% of 39 patients) received consensus from both psychiatrists. The inter-rater agreement measured by Cohen's kappa statistic (0.59) shows a moderate agreement. Similarly, among 68 schizophrenia patients diagnosed according to ICD-10, only 24 patients (or 35% of 68 patients) received consensus from both psychiatrists. Cohen's kappa statistic (0.56) also shows a moderate agreement.

On the other hand, the congruence between DSM-IV and ICD-10, measured by Cohen's kappa statistic (0.61), is just slightly better. The number of schizophrenia diagnosis by DSM-IV criteria (39 patients, or 39% of total cases) is much lower than that by ICD-10 criteria (68 patients, or 68% of total cases) (Cheniaux, et al., 2009). Among 39 patients detected by DSM-IV, 38 are also detected by ICD-10. In contrast, there are 30 patients (or 44% of 68 patients) receiving ICD-10 diagnosis, but not DSM-IV.

The lower rate of diagnosis of schizophrenia according to DSM-IV (or DSM-III-R) than ICD-10 has also been reported in two other studies by (Hiller, Dichtl, Hecht, Hundt, & von Zerssen, 1994) and (Wciorka, et al., 1998). The reason for that may lie in the more strict criteria in DSM-IV than in ICD-10. Six months of symptom duration is required by DSM-IV, whereas only one month is required by ICD-10.

Neuroimaging is not included: The pathology of schizophrenia is believed to be a disease of neuroconnectivity. Although the modern neuroimaging techniques have been developed to quantify the brain grey and white matter abnormalities, they are still not routinely applied in diagnosis of schizophrenia.

As we can see that, since the current two standard procedures of schizophrenia diagnosis (DSM-IV and ICD-10) are generally based on objective criteria, such as symptoms from family members' observations, the diagnostic reliability becomes questionable.

In fact, personal criteria are usually applied in the diagnosis of schizophrenia in addition to the standard DSM-IV/ICD-10 procedures. As (Edlund, 1986) and (Peralta & Cuesta, 2000) have pointed out, diagnosis made by psychiatrists were actually based on the their theoretical background, clinical experience, and preference for diagnostic criteria.

For example, the initial diagnosis may be enforced or altered by medical records (including progress reports, physician orders, hospital admission and discharge summaries), following-up interviews, and/or by using some clinical scoring systems, such as the Positive And Negative Syndrome Scale (PANSS). This approach is also suggested by (Ramirez Basco, et al., 2000).

1.4.2 Hypothesis

Accuracy of a diagnosis is defined as the ratio of total number of correctly diagnosed cases (patient and non-patient) to the total number of cases.

$$Acc = \frac{Cor}{Nr} \qquad \textbf{(1.1)}$$

where *Cor* is total number of correctly diagnosed cases, *Nr* is the total number of cases.

We hypothesized that the accuracy of schizophrenia diagnosis can be improved by using objective and quantitative criteria from a wider spectrum of modalities including neuroinformatics and neuroimaging.

As we can see that schizophrenia is a complicated disease and its economic burden to the patients and society is enormous, we attempt to explore the disease from both neuroimaging and neuroinformatics directions in order to achieve a better understanding of the quantitative relationships between schizophrenia and intermediate phenotypes (as assessed by neurocognitive tests) and brain abnormalities (as assessed by neuroimaging). We will also try to develop a decision supporting system in order to provide classification results (derived from a person's neuroinformatics and neuroimaging data) as additional evidence to the current standard schizophrenia diagnostic procedures. Even though currently there is no efficient treatment, more accurate diagnosis would be useful in identification of patients and healthy persons, and might be also helpful in future potential drug development that targets at specific brain structures defects revealed by our classification models.

1.4.3 Assumptions

There are a few assumptions underlying this research project. We discuss them briefly here.

1) There are enough subjects, including schizophrenia patients and healthy controls. The study subjects' demographic data should always be collected. Neurocognitive tests and neuroimaging should be done on the study subjects and the data should be available.

We estimate that our classification models will have about 5 to 20 (the typical range of clinical prediction models (Steyerberg, 2009)) factors. At the recommended minimum subject to factor ratio of 10 to 1 (Bartlett, et al., 2001), the number of subjects required will be at least 50 to 200.

2) The ground truth (whether a subject is a schizophrenia patient or a healthy control) in the study dataset should be already diagnosed by domain experts (psychiatrists from Institute of Mental Health, Singapore) and is available to us. The ground truth diagnosis is achieved by not only DSM-IV criteria, but also on all medical records reviews (including progress reports, physician orders, hospital admission and discharge summaries), following-up interviews and some clinical scoring systems such as the Positive and negative Syndrome Scale (PANSS).

1.4.4 Major Works

Our major works consist of three parts:

1) To apply an automatic Region of Interests (ROI) selection method based on a brain atlas for neuroimaging quantification and analysis in schizophrenia study.

Patients will be scanned by using structural MRI and DTI imaging. When analyzing these medical images by Regions of Interests (ROI) methods, the placement of ROIs is usually done manually. For this study, we will apply a new method developed at Biomedical Imaging Lab, Agency of Science, Technology

and Research, Singapore using the Fast Talairach Transformation (FTT) method for electronic Talairach brain atlas registration to select ROIs and quantify the neuroimaging features automatically.

2) To discover the relationships between schizophrenia and brain abnormalities and intermediate phenotypes using neuroimaging (Diffusion Tensor Imaging (DTI)) and neuroinformatics data.

Schizophrenia is a complicated disease. Since schizophrenia was identified over 100 years ago, many efforts have been put in order to understand its etiology. It is hypothesized that schizophrenia is related to pathological neuroconnectivity through neuronal circuits (Andreasen, 2000). After the pioneer work of using DTI to study schizophrenia by (Buchsbaum, et al., 1998), researchers have found that various brain structures are associated with schizophrenia pathology. DTI has shown promise as a method to examine the brain white matter abnormalities.

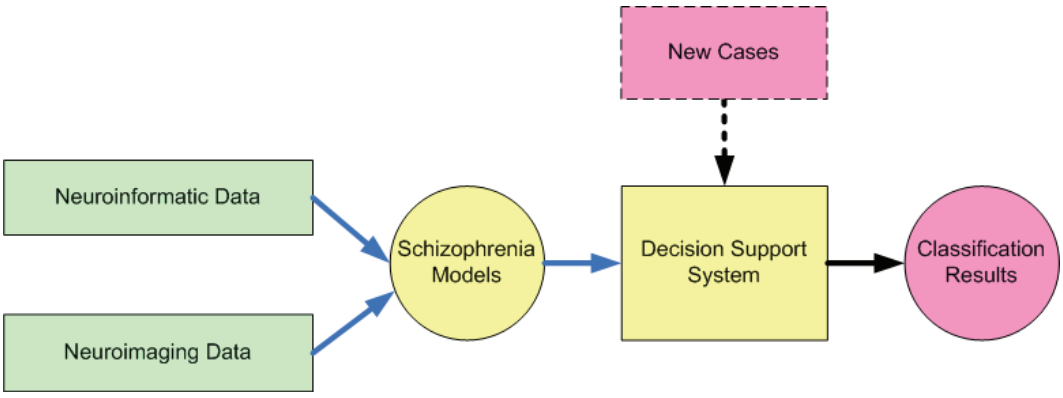

Figure 1.1 Conceptual diagram of schizophrenia modeling and decision support system

We will combine all factors from neuroinformatics data and neuroimaging data to build up schizophrenia models (Figure 1.1), which will help people to understand

this disease in a wider perspective. Multiple models of schizophrenia will be generated according to different combinations of data.

(3) To develop a decision support system based on the image analysis results and the neuroinformatics data.

Based on our schizophrenia models, we will develop a decision support system (Figure 1.1). It will choose the appropriate model automatically according to the availability of input information of new cases, and classify the cases as either patients or normal controls. It will assist clinicians by providing additional objective evidence in the schizophrenia diagnosis. Psychiatrists would be able to gain more confidence from using the objective diagnosis criteria in addition to the existing diagnosis process that rely on subjective criteria, provided that the decision support system and its underlying models have been validated in future large scale trials.

1.4.5 Major Contributions

The major contributions of this work will be:

1) The first ever schizophrenia classification models based on objective and quantitative criteria including neurocognitive tests and neuroimaging. These models quantify the relationship between schizophrenia and the relative factors from neurocognitive results and neuroimaging features, which help us to achieve a better understanding of the disease.

2) A decision support system based on our schizophrenia models that can provide the classification results as more objective evidences to clinicians in addition to the current standard diagnostic procedures. It can also help clinicians to choose the suitable further tests in order to improve the diagnosis accuracy. Our solution tries to tackle the objective criteria problems of existing diagnosis procedures. We use quantitative and objective criteria, including neurocognitive tests and neuroimaging analysis results. We think our classification results will augment the current diagnostic procedures.

3) Atlas-assisted analysis of DTI data. The structural MRI images of 156 study subjects are registered to the Talairach brain atlas. FA images are generated from the DTI images and co-registered with the structural MRI images. The automatic atlas-based ROI selection method is applied to quantify the FA image features within 48 brain anatomical structures.

1.5 Organization of the Thesis

In the rest of this thesis, we will first do a literature survey on neuroimaging analysis technologies and findings in schizophrenia, existing schizophrenia models and decision support systems in Chapter 2.

In Chapter 3, we will describe the neuroinformatics data acquisition, data items from different categories, including demographic data, clinical data, clinical

scoring systems, and neurocognitive tests, data pre-processing, data feature distribution and analysis, and schizophrenia modeling using these data.

In Chapter 4, we will introduce our novel neuroimaging analysis method for FA image quantification and results, and construct a schizophrenia model based on the image features. We will also interpret the clinical significance of the image features.

Since we have both neuroinformatics data and neuroimaging data, in Chapter 5, we will use both data to create more comprehensive models of schizophrenia, and compare their results. Effects of individual neurocognitive test and neuroimaging in improving diagnosis accuracy will be also discussed here.

In Chapter 6, we will employ various models constructed in the previous chapters to make two decision support flow charts for helping clinicians to choose the best further tests in different strategies. We will also develop a decision support system that will classify input case as either schizophrenia patient or normal control and provide suggestions to clinicians on what further tests should be done in different situations.

Finally we will conclude our work and discuss about model accuracies, validations, limitations of our work and the possible future work directions in Chapter 7.

Chapter 2

Literature Review

Since our work involves neuroimaging, schizophrenia modeling, and decision support, we review all these aspects in the following sections.

2.1 Neuroimaging Analysis in Schizophrenia Study

Various neuroimaging techniques and modalities have been used in schizophrenia study, from the very early pneumoencephalography, echoencephalography, to modern computer tomography (CT), magnetic resonance imaging (MRI), until the most recent diffusion tensor imaging (DTI). We will review all these methods and the related findings in the following sections.

2.1.1 Early Neuroimaging Techniques

In the early studies of schizophrenia to examine the anatomy of nervous system, neuropathology was the only tool. The in vivo method, pneumoencephalography (PEG) was introduced in 1919, and in 1970s non-invasive technologies, echoencephalography and computer tomography (CT), were used to study the brain (Lawrie SM, 2004).

PEG is a technology used in brain imaging. It drains cerebrospinal fluid (CSF) in the brain to a very low level, then fills the space with air to make brain X-ray images clearer. This procedure is painful and dangerous to patients. An example of PEG applied in a study by (Jacobi & Winkler, 1927) showed a high prevalence of cortical and subcortical abnormality in schizophrenia.

Echoencephalography is another early technology to scan the brain by using ultrasound. In a study at 1973, echoencephalography was used to examine 79 chronic schizophrenia patients and 79 normal controls; (Holden, Forno, Itil, & Hsu, 1973) found schizophrenia patients with ventricular widening were significantly less likely to respond to antipsychotic medication.

The first CT schizophrenia study in schizophrenia was performed by (Johnstone, Crow, Frith, Husband, & Kreel, 1976), which demonstrated schizophrenia patients had an increased lateral ventricle area compared to the normal controls.

2.1.2 Morphology Study Based on Structural MRI

In 1984, the first study in schizophrenia using magnetic resonance imaging (MRI) was carried out by (Smith, et al., 1984). The images produced by MRI are much clearer than those from CT. Since MRI doesn't require ionizing radiation, repeated scanning can be performed with less risk to the patients. Therefore MRI superseded CT in a few years time.

Since then, structural MRI (sMRI) has been applied to schizophrenia study for over 20 years. Much more knowledge about schizophrenia has been discovered, and over 100 studies have been conducted to compare schizophrenia patients with controls (Lawrie SM, 2004). The whole brain and various related brain structures have been studied, including: ventricles, cerebrospinal fluid (CSF), prefrontal cortex, temporal lobes, amygdala, hippocampus, parahippocampus, thalamus, basal gangalia, nucleus accumbens, and insula. Table 2.1 summarizes some findings using structural MRI (sMRI).

Table 2.1 Summary of structural magnetic resonance imaging findings in schizophrenia
A combined result of (Wright, et al., 2000) and (Shenton, Dickey, Frumin, & McCarley, 2001). From (McIntosh & Lawrie, 2004)

Regions	No. of Cases	Findings
Whole brain volume	50+	Reduced by 2-3%; Grey matter reduced by 4%; White matter no difference.
Frontal lobes	50+	Reduced by 3%
Temporal lobes	100+	Reduced by 5-6%
Hippocampus	10+	Reduced by 4%
Parahippocapus	10+	Reduced by 10%
Amygdala	10+	Reduced by 4%
Basal gangalia	20+	Globus pallidus increased by 20%
Thalamus	10+	Significant difference (no quantitative results)
Lateral ventricles	20+	Increased by 20%
Third ventricle	30+	Increased by 26%

From those studies, we see some hints on the neuroanatomical profiles of schizophrenia. But the numbers and types of cases studied are not large enough to be more convincing.

The most replicable brain morphometric phenomenon from MR imaging are enlarged ventricles and reduced cerebral volume, but the changes are relatively subtle. Although white matter may also be affected by schizophrenia, the white mater abnormalities are not reported since ordinary MRI images are not good at detecting white matter connectivity.

2.1.3 White Matter Study Based on Diffusion Tensor Imaging

Fortunately with the introduction of Diffusion Tensor Imaging (DTI), this situation has been changed. By using DTI, researchers can study neural fibers, spinal cord white matter and brain white matter. DTI is based on the principle that water diffusion is highly anisotropic in the nervous tissue. Since water molecules diffuse preferentially along axons rather than across them, by the diffusion tensor anisotropy, we can get detailed information on neural fiber direction and other architectural features of brain tissue.

The apparent diffusion tensor describes the molecule mobility along each direction and correlation between these directions (Le Bihan, et al., 2001).

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \qquad \textbf{(2.1)}$$

where D_{xx}, D_{yy} and D_{zz} are the diffusion fluxes along *x, y, z* directions; D_{xy}, D_{xz},D_{yx}, D_{yz} , D_{zx} *and* D_{zy} are correlations between diffusion fluxes in orthogonal directions.

The tensor is symmetric, i.e., $D_{xy}=D_{yx}$, $D_{xz}=D_{zx}$ *and* $D_{yz}=D_{zy}$ for uncharged moiety, water molecular.

The diffusion coefficients along three principal directions are represented by the eigenvalues of the tensor, λ_1, λ_2, and λ_3.

Mean diffusivity (MD) is a measurement of the overall evaluation of the diffusion, which is defined as the arithmetic average of the eigenvalues of the tensor.

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad \textbf{(2.2)}$$

where λ_1, λ_2, λ_3 are the eigenvalues of the diffusion tensor.

Another scalar index is fractional anisotropy (FA) (Pierpaoli & Basser, 1996), which provides a quantitative rotationally invariant assessment of diffusion anisotropy, and highlights the brain white matter tracts effectively (Parker, 2004).

$$FA = \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}}{\sqrt{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \quad \textbf{(2.3)}$$

(Buchsbaum, et al., 1998) reported significantly reduced diffusion anisotropy in prefrontal cortex, internal capsule and temporal lobe in a study of five chronic schizophrenia patients and six normal controls. Since then many researches have been done to study the white matter abnormalities in schizophrenia using DTI. Table 2.2 shows a summary of these studies.

Table 2.2 Summary of schizophrenia studies using DTI

Study	Subjects (Patient/Control)	Findings
(Buchsbaum, et al., 1998)	5/6	Reduced FA in frontotemporal peri-putamen
(Lim, et al., 1999)	10/10	Reduced FA in whole brain white matter
(Foong, et al., 2000)	20/20	Reduced FA in splenium of CC
(Agartz, Andersson, & Skare, 2001)	20/24	Reduced FA in splenium of CC
(Steel, et al., 2001)	10/10	No significant differences in prefrontal and occipital regions
(Foong, et al., 2002)	14/19	No significant differences
(Kubicki, et al., 2002)	15/18	Loss of asymmetry in uncinate fasciculus
(Hoptman, et al., 2002)	14/0	No control group; Right inferior frontal white matter FA was correlated with higher motor impulsiveness
(Wang, et al., 2003)	29/20	No significant difference in middle and superior cerebellar peduncles
(Z. Sun, et al., 2003)	30/19	Reduced FA in anterior cingulum
(Ardekani, Nierenberg, Hoptman, Javitt, & Lim, 2003)	14/14	Reduced FA in bilateral CC, AC, MTG, parahippocampal gyri (PHG), left STG
(Burns, et al., 2003)	30/30	Reduced FA in left arcuate fasciculus
(Wolkin, et al., 2003)	10/0	No control group; Inferior frontal FA was correlated with negative symptoms
(Minami, et al., 2003)	12/11	Reduced FA in frontal, parietal, temporal, occipital regions
(Kubicki, et al., 2003)	17/18	Reduced FA in cingulum
(Begre, et al., 2003)	7/7	No significant differences in hippocampus
(Wang, et al., 2004)	21/20	Reduced FA in anterior cingulum
(Okugawa, et al., 2004)	25/21	Reduced FA in middle cerebellar peduncles
(Kalus, et al., 2004)	15/15	Reduced FA in bilateral posterior hippocampus, left total hippocampus
(Kumra, et al., 2004)	12/9	Reduced FA in bilateral frontal WM, and right occipital WM on AC-PC plane
(Park, et al., 2004)	23/32	Significant differences in anisotropic asymmetry pattern in

Study	Subjects (Patient/Control)	Findings
		left and right hemisphere
(Hubl, et al., 2004)	26/13	Reduced FA in left and right temporoparietal section of the arcuate fasciculus; uncinate fasciculus, corpus callosum, inferior longitudinal fasciculus
(Price, Bagary, Cercignani, Altmann, & Ron, 2005)	20/29	No significant differences in splenium and genu of CC
(Szeszko, et al., 2005)	10/13	Reduced FA in left internal capsule, left middle frontal gyrus, posterior superior temporal gyrus
(Kumra, et al., 2005)	26/34	Reduced FA in the left anterior cingulate region in close proximity to caudate nucleus
(Okugawa, Nobuhara, Sugimoto, & Kinoshita, 2005)	25/21	Reduced FA in middle cerebellar peduncle
(Jones, et al., 2006)	14/14	Young schizophrenia patients have reduced FA in left superior longitudinal fasciculus than controls; old patients, the difference is less
(Buchsbaum, et al., 2006)	64/55	Reduced FA in frontal white matter, CC, and frontal longitudinal fasciculus
(Shergill, et al., 2007)	33/40	Reduced FA in superior longitudinal fasciculi, and genu of CC
(Schlosser, et al., 2007)	18/18	Reduced FA in right medial temporal lobe adjacent to the right parahippocampal gyrus
(Cheung, et al., 2008)	25/25	Reduced FA in left fronto-occipital fasciculus, left inferior longitudinal fasciculus
(Kyriakopoulos, Vyas, Barker, Chitnis, & Frangou, 2008)	19/20	Reduced FA in the white matter of the parietal association cortex bilaterally and in the left middle cerebellar penduncle
(Hoptman, et al., 2008)	23/37	Reduced FA in left superior and middle temporal gyri, left ILF, left cingulate gyri, and left inferior frontal gyrus and right perilentiform regions. Increased FA in left lingual and

Study	Subjects (Patient/Control)	Findings
		insular gyri, and right deep frontal white matter
(Rametti, et al., 2009)	25/24	Reduced FA in left sub-gyral WM of temporal lobe, involving posterior part of the fornix
(Rotarska-Jagiela, et al., 2009)	24/24	Reduced FA in the prefrontal regions, external capsule, pyramidal tract, occipitofrontal fasciculus, superior and inferior longitudinal fasciculi, and corpus callosum. Increased FA in arcuate fasciculus
(Moriya, et al., 2009)	19/19	No significant difference in FA
(Sussmann, et al., 2009)	28/38	Reduced FA in the anterior limb of the internal capsule, anterior thalamic radiation, uncinate fasciculus

Abbreviations: see Appendix B

It can be seen that studies using DTI have produced various findings at many different brain structures, but haven't drawn consistent conclusions yet. For example, (Foong, et al., 2000) and (Agartz, et al., 2001) reported the reduced FA in CC splenium, but (Price, et al., 2005) reported no significant changes in the same region. The limitations may have resulted from small sample size, inhomogeneous sample characteristics, insufficient image quality, and image processing techniques.

In this section, we reviewed various imaging technologies in schizophrenia, especially the most recent DTI imaging which helps to determine the neural fiber abnormalities. We found that DTI is a promising technology; however current findings are not consistent.

2.2 Schizophrenia Models

The schizophrenia etiology and development are very complicated. Only few research works have attempted to build up some schizophrenia models.

(Andreasen, 1999) describes a general model defining the development of schizophrenia. This model has an hour glass shape, which illustrates a many to many relationship between etiologic factors and phenomenology: multiple etiology (such as DNA, gene expression, virus, toxins, nutrition,birth injury and psychological experiences) fan in at the input level and multiple impairments in fundamental and second order cognitive processes including attention, memory language, executive functions, emotion and symptoms (such as hallucination, delusions, negative symptoms, disorganized speech and behavior) fan out at the output level. Between the input etiology and output phenomenology, there is a single lathomenologic process of anatomical and functional disruption in neuronal connectivity and communication that unifies the sickness. However this is a conceptual model only. It doesn't describe any quantitative relationships between etiology factors and symptoms. Hence it is not meant to be applied clinically.

(Hoffman & McGlashan, 2001) establishes neural network models to explore functional consequences of reduced corticocortical connectivity. The models simulate the auditory hallucinations of speech.

2.3 Decision Support System in Schizophrenia

In this section, we will briefly introduce some major types of decision support systems in schizophrenia from current literature. Typically, they include decision support in treatment planning and diagnosis.

2.3.1 Decision Support in Treatment Planning

Many works have been done on the schizophrenia decision support system, especially on the treatment planning evaluation.

For example, (Palmer, Brunner, Ruiz-Flores, Paez-Agraz, & Revicki, 2002) set up a decision tree model to evaluate different schizophrenia treatment plans by using 3 different medications (Haloperidol, Olanzapine and Risperidone) for a 5 year period. They discovered that Olanzapine therapy resulted in more symptom improvement, fewer relapses and was more cost-effective.

(Hansen, Lancon, & Toumi, 2006) pointed out that pharmacoeconomic evaluations were important in the decision making process. Five different decision tree models are developed to compare any two different strategies (A and B): relapse model (by comparing the incremental cost and the cost effectiveness of two strategies), compliance model (taking into consideration of patient's compliance with the treatment plans) , institution model (by including another factor reflecting the fact that some schizophrenia patients are not able to live with their families, and require

institutionalization), drop-out model (by adding one more factor for patients who drop out of treatment) and switch model (for patients who change treatment plans at halfway). Different models can be used at different conditions to evaluate the treatment plans based on their cost effectiveness and clinical outcome.

2.3.2 Decision Support in Diagnosis

Some other works focus on the diagnosis. (Razzouk, Mari, Shirakawa, Wainer, & Sigulem, 2006) developed a decision support system for diagnosis of schizophrenia spectrum disorders (a group psychiatric diagnoses similar to schizophrenia, such as schizoaffective disorder). The decision support system consists of four stages: knowledge acquisition, knowledge organization, computer assisted model construction (based on the parsimonious cover approach (Mitchell, 1997), which defines possible diagnosis as all diseases that can explain a patient's existing symptoms), and system performance evaluation. The decision criteria of this system are symptom based. It aims to differentiate schizophrenia from schizophreniform disease (a disease which is close related to schizophrenia, but the required length of symptoms presence is less than 6 months as in the case of schizophrenia, and some of the functional impairments of schizophrenia may not present). The system achieved an accuracy rate at 66-82% on 38 clinical cases.

(Yana, et al., 1994; Yana, et al., 1997) also proposed a classifier for the diagnosis of psychiatric disease including schizophrenia. From the 136 simple yes/no questions set by domain experts in Tokyo Medical and Dental University Hospital, 80 questions are selected to build a Pseudo Bayesian Network classifier and a

Neural Network classifier. Among the 80 selected questions, the first 32 are subjective symptoms, and the rest are items supposed to be important for prediction. Table 2.3 shows the first 20 questions. 100 cases are used for the classifier model construction. By using the 10-fold cross-validation method, the correct schizophrenia diagnosis rates are measured at 73.3% and 77.3% for the Pseudo Bayesian Network classifier and the Neural Network classifier, respectively.

Table 2.3 Question items (partial)

No.	Question	No.	Question
1	Headache	11	Became dull (decrease in brain power)
2	Nausea	12	Incorrect behavior
3	Cannot see clearly	13	Personality changes
4	Shaking	14	Irritated
5	Difficult in speech	15	Anxiety
6	Numbness	16	Difficult in thinking
7	Difficult to control arms / legs	17	Difficult in sleep
8	Convulsion	18	Diminished appetite
9	Lost consciousness	19	Lack of sexual desire
10	Amnesia	20	Lack of motivation

In this section, we introduced two decision support systems in pharmacoeconomic evaluation, and two in schizophrenia diagnosis. Among the two diagnosis systems, one tries to differentiate schizophrenia spectrum disorders, and the other diagnoses schizophrenia, mood disorder (disease where the patient's mood disturbance is the main feature) and neurosis (disorder involving distress but no delusions and hallucinations). However, the two diagnosis systems were still based on subjective

symptoms and questions, and no lab test results (such as neuroimaging) were used in the decision support.

2.4 Machine Learning Technology

Machine learning techniques need to be applied in order to extract knowledge from the neuroimaging and neuroinformatics data and set up the schizophrenia model and decision supporting systems. Decision Trees (Raiffa, 1968), Bayesian Networks (also known as Belief Networks) (Pearl, 1988) (Jensen, 2001), and Neural Networks (Bishop, 1996) are some common modeling representation forms.

Based on the Iterative Dichotomiser 3 (ID3) algorithm, (Quinlan, 1993) developed the C4.5 algorithm to generate a decision tree from a set of training data using the concept of information entropy (Shannon, 1948). The training data is a set of classified samples. Each sample contains multiple features and a class to which that sample belongs. At each node of the tree, C4.5 algorithm chooses a feature with the causes the highest information gain, and splits the training samples into subsets. It will apply the same rule recursively on each subset until each node contains only one class.

Learning a Bayesian Network involves the parameter learning and structure learning. (Neapolitan, 2004) introduced some basic concepts and methods in constructing the Bayesian Network, including learning with missing data items (incomplete datasets). Parameters (probability distribution) can be estimated by using the likelihood function. However, there is no efficient algorithm for structure

learning when the networks become complex since finding optimal structure is NP-hard when multiple parent nodes are allowed. Hence researchers developed many different heuristic techniques such as Greedy hill-climbing (Chickering, 2002) and K2 search (Cooper & Herskovitz, 1992).

There are many ways of training neural networks. Most of them are based on some form of gradient descent, which takes the derivative of the cost function with respect to the network parameters and then changes those parameters in a gradient-related direction (Haykin, 1998).

In this chapter, we briefly reviewed the current status of schizophrenia modeling and decision support systems. In the following two chapters, we will introduce our work on neuroinformatics data and neuroimaging data analysis and schizophrenia modeling by using those data separately.

Chapter 3

Neuroinformatics-Based Analysis and Modeling

This chapter will cover the neuroinformatics data acquisition (what sorts of data are collected, and their characteristics), pre-processing (missing value processing, exclusion of irrelevant data, data error correction), analysis (feature selection), and model construction.

3.1 Study Subjects

Schizophrenia patients and healthy controls are recruited from Institute of Mental Health / Woodbridge Hospital, Singapore, the national psychiatric hospital and main treatment center. This project is supported by the National Healthcare Group Research Grant (NHG-SIG/05004) and Singapore Biomedical Imaging Consortium (Agency for Science, Technology and Research) Research Grant (SBIC RP C-009/2006).

We collect 156 study subjects, including 92 schizophrenia patients and 64 normal, healthy controls for this study.

The selection criteria for the patients are: diagnosis of schizophrenia (based on clinical history, medical record review, interview with the significant others when

necessary as well as the Structured Clinical Interview for DSM IV), age between 21-65 years old, and English speaking. The patients with the following criteria will be excluded: history of significant head injury, neurological diseases such as epilepsy, cerebrovascular accident, impaired thyroid function, steroid use, DSM-IV alcohol or substance use or dependence, and contraindications to MRI.

The selection criteria for the normal controls are: having no history of any neurological or psychiatric disorders. Controls should match patients on age, gender, years of education and handedness.

All patients and controls are screened for co-morbid medical and psychiatric conditions by clinical assessment and physical and neurological examination (K Sim, 2005).

3.2 Demographic Data

Demographic data are collected for patients and controls. A brief summary of the characteristics of study subjects can be found in Table 3.1. Pearson Chi-Square is used to test the independency of two categorical factors such as sex, handedness, and ethnicity, while Independent Sample T Test (2-tailed) is used to compare the mean score of two continuous factors, such as age, years of duration (yrsedu), weight and height.

The age range of patients is from 18 to 56 years old, whereas the age range of healthy controls is 21 to 58 years old. The mean ages of patients and controls are 34 and 32 years old, respectively, their difference is not statistically significant.

Table 3.1 Characteristics of study subjects (N=156)

Characteristic	Schizophrenia Patients (N=92) (59.0%)	Healthy Controls (N=64) (41.0%)	P Value
Age, years	34.28 (SD 9.20)	32.33 (SD 10.20)	0.214 (NS)*
Sex (F/M)	18/74 (19.6%/80.4%)	26/38 (40.6%/59.4%)	0.004 (SIG)^
Handedness (Left/Right/Ambidextrous)	10/81/1 (10.9%/88.0%/1.1%)	7/57/0 (10.9%/89.1%/0%)	0.705 (NS)^
Ethnicity (Chinese/Malay/Indian/Others)	85/5/2/0 (92.4%/5.4%/2.2%/0%)	55/3/5/1 (85.9%/4.7%/7.8%/1.6%)	0.228 (NS)^
Marital Status (Single/Married/Widowed/Divorced)	78/12/0/2 (84.8%/13.0%/0%/2.2%)	42/21/1/0 (65.6%/32.8%1.6%/0%)	0.009 (SIG)^
Education level, years	11.40 (SD 2.49)	13.97 (SD 2.10)	< 0.001 (SIG)*
Father's education level, years	7.21 (SD 3.63)	8.33 (SD 3.91)	0.068 (NS)*
Mother's education level, years	6.43 (SD 3.80)	7.88 (SD 4.16)	0.027 (SIG)*
Weight, kg	64.30 (SD 14.10)	64.84 (SD 11.12)	0.80 (NS)*
Height, m	1.68 (SD 0.07)	1.67 (SD 0.08)	0.60 (NS)*
Age of first onset, years	25.37 (SD 6.92)	-	-
Duration of illness, years	8.69 (SD 8.41)	-	-

Note: ^ Pearson Chi-Square; * Independent Sample T Test; **Abbreviations**: SD, Standard Deviation; F, Female; M, Male; SIG, Significant (P<0.05); NS, Not Significant

The percentage of sex of patients is 19.6%/80.4% (female/male). The handedness distribution (Left/Right/Ambidextrous) of our patients and controls are similar: 10.9%/88.0%/1.1% for patients, and 10.9%/89.1%/0% for controls.

Note that the mean education level of patients (11.40 years) is significantly less than that of healthy controls (13.97 years) by 2.57 years. This can be considered as the consequence of schizophrenia – the patients' intelligence quotient (IQ) and capability of education are affected by the disease.

For patients, the mean age of first onset is 25.37 years (SD 6.92), and the average duration of illness is 8.69 years (SD 8.41). The mean weight and height of patients and controls have no significant difference in our study.

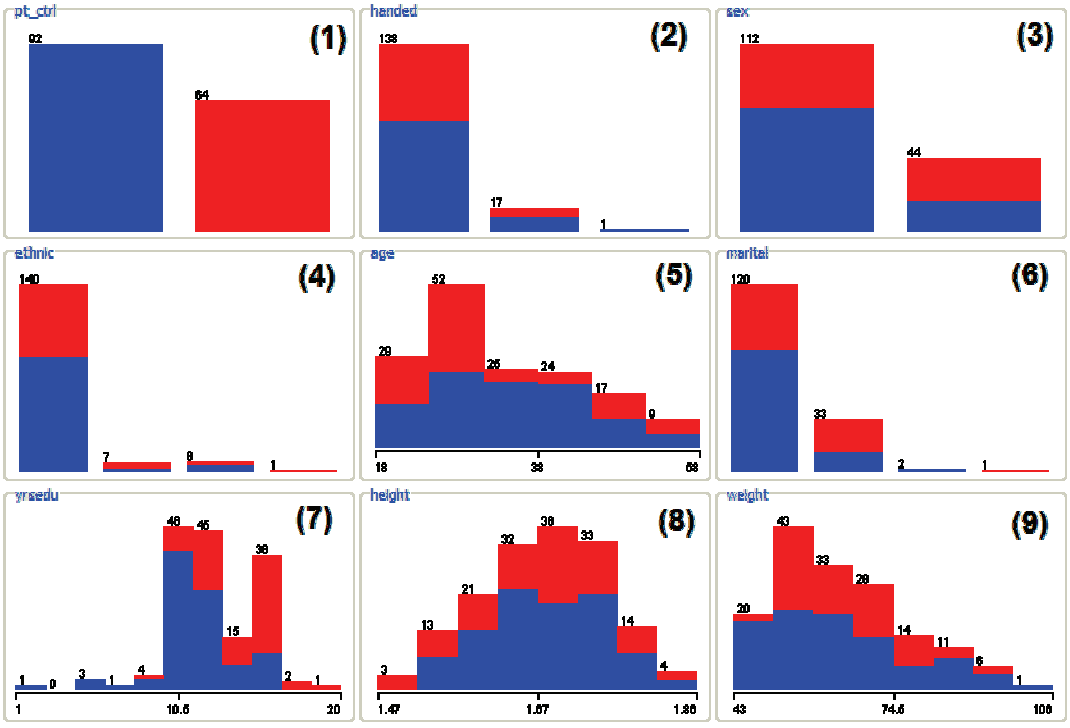

Figure 3.1 Demographic data distribution (N=156) (partial)
(Blue: Patient; Red: Control. 1: pt_ctrl (patient/control), 2: handedness (right/left/ambidextrous), 3: sex (male/female), 4: ethnic (Chinese/Indian/Malay/Others), 5: age (years), 6: marital status (single/married/divorced/widowed), 7: years of education, 8: height (in m), 9: weight (in kg). Descriptions for the data items, see Appendix A)

Figure 3.1 illustrates the distribution of major demographic data (age, sex, ethnics, handedness, years of education, weight, height, and marital status) for our samples

subjects (64 controls and 92 patients). Ethnic distribution looks severely skewed, but it is not far away from that of Singapore population in recent years (Chinese/Malay/Indian/Others: 74.2%/13.4%/9.2%/3.2%) (Singapore, 2009).

In this section, we describe the demographic data of the subjects collected. We also compare some of them (such as age, sex, handedness, weight, and height) between the patient and control group. We find that these two groups are basically matched in terms of major demographic characteristics. However since our study is conducted in Singapore, our sample contains mainly Singaporean Chinese due to the limitations in population distribution of the country.

3.3 Other Clinical Data

In addition to the demographic data, other clinical data are also collected, which include clinical information (such as date of admission to ward (patient only), medical problems, surgical problems, drug use, alcohol use, family history of psychiatric disease), medication information of patients, clinical scores (such as Positive and Negative Syndrome Scale (PANSS) (S. Kay, Opler, & Fiszbein, 1986.), and World Health Organization Quality of Life Bref-Scale (WHOQOL-BREF) (WHOQOL, 1998). A detailed list of all clinical data features can be found in Table 3.2.

Table 3.2 List of clinical data features

Feature	Feature
Diagnosis: patient or control	**Medication**
	Antipsychotics 1 (Type/ Dose)
Demographic	CPZ Equivalents 1
Sex	Typical/ Atypical/ Depot Antipsychotic
Age	Antipsychotics 2 (Type/ Dose)
Weight	CPZ Equivalents 2
Height	Typical/ Atypical/ Depot Antipsychotic
Handedness	Antipsychotics 3 (Type/ Dose)
Ethnicity	CPZ Equivalents 3
Father's Ethnicity	Typical/ Atypical/ Depot Antipsychotic
Mother's Ethnicity	Anticholinergics (Type/ Dose)
Paternal Grandfather's Ethnicity	Antidepressants (Type/ Dose)
Paternal Grandmother's Ethnicity	Mood Stabilizers (Type/ Dose)
Maternal Grandfather's Ethnicity	Benzodiazepines (Type/ Dose)
Maternal Grandmother's Ethnicity	Other Medications 1 (Type/ Dose)
Marital Status	Other Medications 2 (Type/ Dose)
Educational Level	
Years of Education	**Clinical Scores**
	Scale for the Assessment of Passivity Phenomena
Educational Level of Mother	**(SAPP)**
Educational Level of Father	1a - Time Frame
Occupation	1- Made Emotions
Father's Occupation	2 - Made Movements
Mother's Occupation	3 - Made Impulses / Decisions to Act
Living Arrangements	4 - Somatic Passivity
Living Arrangements (specify)	Total Score
Brought By	**Positive and Negative Syndrome Scale (PANSS)**
Brought By (specify)	PANSS Positive (1-7)
	PANSS Negative (1-7)
Clinical	PANSS General Psycopathology Scale (1-16)

Feature	Feature
Date of Admission to Ward	**Global Assessment of Functioning Scale (GAF)**
Diagnosis Axis 1 (DSM IV)	Total
Medical Problems (past or current)	Symptoms
Medical Problems (specify)	Disability
	Scale to Assess Unawareness of Mental Disorders
Surgical Problems (past or current)	**(SUMD)**
Surgical Problems (specify)	1 - Awareness of Mental Disorder
Alcohol Use (past or current)	2 - Awareness of Consequences of Mental Disorder
Drug Use (past or current)	3 - Awareness of Effects of Medication
Duration of Psychiatric Illness (years)	4 - Awareness of Hallucinatory Experiences
Age of First Onset of Illness	5 - Awareness of Delusions
Duration of Untreated Psychosis (in years)	6 - Awareness of Thought Disorder
Number of Hospitalizations	7 - Awareness of Flat or Blunt Affect
Number of Hospitalizations in Last 12 Months	8 - Awareness of Anhedonia
Regularity of Outpatient Attendance in Last 12	
Months	9 - Awareness of Asociality
Family History of Mental Illness	**WHO Quality of Life (WHO QOL-BREF)**
Family History of Mental Illness (specify)	WHO QOL-BREF 1-26

The Scale for the Assessment of Passivity Phenomena (SAPP) identifies passivity phenomena of patients on the basis of their total scores equal or greater than 4 on four items (Spence, et al., 1997).

The Positive and Negative Syndrome Scale (PANSS) is a medical scale to measure severity of positive and negative symptoms for schizophrenia patients. It was originally published in 1987 (S. R. Kay, Fiszbein, & Opler, 1987). It is widely used in the study of antipsychotic therapy. It consists of 30 items, which are grouped into 3 sub-categories: 7 were chosen to assess positive symptoms, 7 for

negative symptoms, and the remaining 16 for general psychopathology scale. The PANSS interview usually requires about 45 to 50 minutes to administer.

The Global Assessment of Functioning (GAF) is a 100-point scale used by mental health clinicians and physicians to subjectively evaluate the social, psychological and occupational functioning of a patient. It reports the clinician's judgment of the patient's overall level of functioning and carrying out daily activities. The scale is described in the DSM-IV-TR (*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*, 2000).

The Scale to Assess Unawareness of Mental Disorder (SUMD) (Amador, et al., 1993) is an assessment that measures the patient's discrete and global aspects of insight awareness of the his/her illness, including the awareness of 9 aspects: mental disorder, consequences of mental disorder, effects of medication, hallucinatory experiences, delusions, thought disorder, flat or blunt affect, anhedonia (lack of pleasure), and asociality.

The World Health Organization Quality of Life Bref-Scale (WHOQOL-BREF) is developed as an assessment instrument for the international cross-culturally comparable quality of life. It consists of 26 questions, which assess the participant's perceptions in the following four major aspects: physical health, psychological health, social relationships, and environment.(WHOQOL, 1998).

In this section, we introduced other clinical information collected, especially various clinical scoring systems. Next we will come to the neurocognitive tests performed for this study.

3.4 Neurocognitive Tests

Schizophrenia is a neurodevelopmental and progressive disease. We aim to build up schizophrenia models to reveal the relationship of the disease with the underlying neaurocognitive defects that can be assessed by various neurocognitive tests.

Patients and controls are administered some neurocognitive tests by psychometrists trained in standardized assessment and scoring procedures.

Table 3.3 lists the neurocognitive tests performed and the scores collected. They include Raven's Progressive Matrices (RPM) (Lezak, 1995), Wisconsin Card Sorting Test (WCST) (Heaton, Chelune, Talley, Kay, & Curtiss, 1993), Continuous Performance Task (or Continuous Performance Test) II (CPT II) (Conners, 2000), and Wechsler Adult Intelligence Scale III (WAIS-III) (D Wechsler, 1997).

The complete set of tests (4 of them) takes about 1.5–2 hours. These tests can be done in separate sessions. They assess the patients/controls' various neurocognitive functions (expressed as intermediate phenotypes) including: intelligence, attention, executive functioning, working memory, and visuo-spatial skills.

Table 3.3 List of neurocognitive tests and features

Test Feature	Test Feature
Raven's Progressive Matrices (RPM)	**Continuous Performance Task, II (CPT-II)**
RPM Raw (Raven's raw score)	Number of Omissions t score
	Number of Commissions t score
Wechsler Adult Intelligence Scale, III (WAIS-III)	Hit Reaction Time t score
Block design raw score	
Digit span forward score	**Wisconsin Card Sorting Test (WCST)**
Digit span backward score	Trials administered
Digit span total score	Total correct
Spatial span forward raw score	Total errors raw scores
Spatial span backward raw score	Perseverative Responses raw scores
Spatial span total score	Perseverative Errors raw scores
	Nonperseverative Errors raw scores
	Categories completed raw scores
	Trials to complete 1st category raw scores

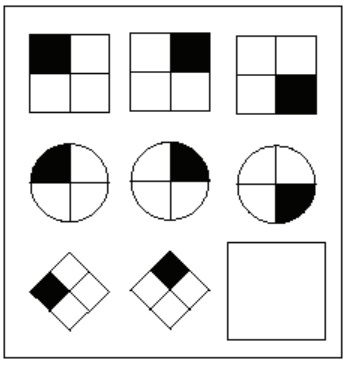

Figure 3.2 A sample RPM matrix
From (Wikipedia, 2011a)

Specifically RPM measures abstract reasoning and intelligence. It was originally developed by Dr John C. Raven in 1936 (Raven, 1936). In each test item, a candidate is required to identify the missing pattern of a series (see example in

Figure 3.2). It tests two important components of intelligence: eductive ability and reproductive ability.

The Wechsler Adult Intelligence Scale (WAIS) is used to test the intelligence of adult and adolescent (David Wechsler, 1939). The version we used in this study, WAIS-III, was published in 1997, while the latest version, WAIS-IV was published in 2008. The main difference between WAIS-III and WAIS-IV is that the later includes 5 more supplementary subtests, and the General Ability Index. As a successor and an enhanced version of WAIS-III, WAIS-IV still needs some time to be widely adopted and implemented by all hospitals. None the less, the subtests used in our study, as described in the next paragraph, remain unchanged in WAIS-III and WAIS-IV.

Specifically, in the Block Design subtest of the WAIS-III, the patient/control is required to take blocks with all white sides, all red sides, and red and white sides and arrange them according to a pattern. This test assesses the visuospatial and motor skills, which are linked to the functioning of the parietal and frontal lobes. Digit Span subset of the WAIS-III, assesses verbal working memory (or short term memory). It is the longest list of digits that a person can repeat back in correct order after they are announced. Backward digit span is a more difficult subtest which requests a person to recall digits in reverse order. Spatial Span subtest of the WAIS-III, on the other hand, assesses spatial working memory.

Continuous Performance Task (or Test) (CPT) measures a person's sustained and selective attention and impulsivity or vigilance. A person is required to click a

button when he sees the pre-set target (for instance, letter “A”) appearing on the computer screen. He must not click the button if he sees any other letters. In CPT scores, Omission Errors indicates the number of times the patient failed to respond. Commission Errors indicates the number of times the patient responded to false target. Reaction time measures the amount of time between the presentation of the target and the client's response.

Wisconsin Card Sorting Test (WCST) measures the ability to form abstract concepts, shift and maintain set, and utilize feedback. Without being told on how to match cards, the participant is required to find the rule and match the stimulus cards presented in front of him (see example in Figure 3.3). The rule may change after some trials. It generates a number of psychometric scores, including numbers, percentages, and percentiles of: categories achieved, trials, errors, and perseverative errors. It has been considered as a measure of executive function because of its reported sensitivity to frontal lobe dysfunction.

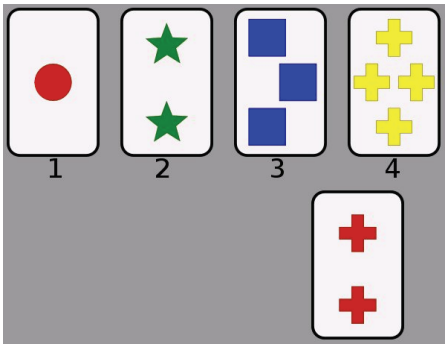

Figure 3.3 A sample WCST test
From (Wikipedia, 2011b)

The time needed to perform each neurocognitive tests varies from 15 minutes to 1 hour depending on the test setting and participant’s reaction time. In Singapore, the

costs for the neurocognitive tests are charged at two different rates, depending on whether the participants are considered as private patients or government subsidized patients. However, in this research project, as the neurocognitive tests are performed by the research assistants of Institute of Mental Health, clinical charges do not apply. In addition, some tests are not performed routinely even at the hospital clinics. We try to estimate the costs, by using information of similar tests from other hospitals, and the time required to complete them. Table 3.4 summarizes the major function tested for each neurocognitive test, as well as the time required to complete them and the costs at both government subsidized rate and private rate.

Table 3.4 Neurocognitive tests

Test	Function Tested	Time Needed	Cost	
			Subsidized Rate	Private Rate
RPM	Eductive, reproductive	30 min – 1 hour	$130*	$400*
WAIS-III	Visual spatial, motor, memory	15 min – 30 min	$65	$240
CPT-II	Attention, impulsivity	15 min – 30 min	$65^	$200^
WCST	Executive	15 min - 20 min	$65^	$200^

Note: *Estimated Cost; ^Estimated by dividing the total cost of CPT-II and WCST ($130 for subsidized, $400 for private) equally.

In this section, we introduced the four neurocognitive tests performed on the study subjects. We briefly described the functionalities, scores collected, and time and costs involved for each of them. In the next section, we will describe our method used in data pre-processing.

3.5 Data Preprocessing

We collect many data items for each study subject, and put them into a large table (156 rows by 211 columns), with each row representing a subject (patient or control), and each column representing a data item. Unfortunately, not all data items are available for each subject. Some columns are missing, for example, PANSS scores are only for patients; hence they are not collected for all healthy controls. Before we start to analyze the data, irrelevant items need to be filtered out, and errors need to be corrected.

Data Feature Extraction

We study the nature of each column of the dataset, and exclude data items that are not useful or suitable for the later analysis in schizophrenia modeling. Since the purpose of this study is to build schizophrenia models using objective criteria only. No matter whether subjective factors are relevant or not, they are excluded at the step of feature extraction. Specifically, the following items are removed: (Data item names and their descriptions can be found in Appendix A)

- 'study-id' is removed, since it serves as a key of each subject, and we already have CDNo (Compact Disk No) for this purpose.
- All columns of medication information are removed, since they are relevant to patients, but not controls, and our aim is to compare patients with controls.
- 'Date admin' (date of admission to hospital) is removed since only patients have this value but not controls.

- ‘live-spec’ (living arrangement, specify) is removed, because it is the specific detailed text information on living arrangement, and it only has values when ‘livingar’ (living arrangement) is 8 (other).
- ‘brought-sp’ (brought by, specify) is removed for the similar reasons as ‘live-spec’ (only has values when ‘brought’ is 9 (other)).
- Similarly, ‘med-spec’ (medical problems, specify) is removed. It only has value when ‘med-prob’ (medical problems) is 1 (yes).
- Similarly, ‘sur-spec’ (surgical problems, specify) is removed. It only has value when ‘sur-prob’ (surgical problems) is 1 (yes).
- Similarly, ‘fam-hxsp’ (family history of mental illness, specify) is removed. It only has value when ‘fam-hx’ (family history of mental illness) is not 1 (Nil).
- ‘dsmaxis1’ (Diagnosis Axis 1 (DSM IV)) is removed, since it specifies the schizophrenia sub-types, and we don’t deal with sub-types in this study.
- ‘first-ep’ (first episode) is removed, since it is only relevant to patients, and for all controls, this column is blank.
- ‘brought’ (brought by) is removed. Generally controls are brought by themselves. This is not rational for the patient’s disease, but rather than a consequence of the disease – patients lost the ability of taking care of themselves.
- Hospitalization features are removed, since they are relevant to patients only; and for controls, all are blank.
- ‘dur-psyc’ (duration of psychiatric illness (years)) is removed for the similar reason as above.
- Similarly ‘age-onset’ is removed.

- Similarly ‘dup-yrs’ (duration of untreated psychosis (in years)), is removed.
- Similarly ‘tcu-reg’ (visiting regularity), is removed.
- All medication information is removed, since they are relevant to patients only; and for controls, all are blank.
- All features in SAPP are removed, since they are relevant to patients only; and for controls, all are blank.
- All features in PANSS are removed, since they are relevant to patients only; and for controls, all are blank.
- All features in GAF are removed, since they are relevant to patients only; and for controls, all are blank.
- All features in SUMD are removed, since they are relevant to patients only; and for controls, all are blank.
- All features in WHO QOL (1-26) are removed, since they are for quality of life questionnaires for the last four weeks, and moreover they are subjective expressions of patients/controls’ feeling.
- ‘comments’ is removed, since it is a short text field for recording notes.

Missing Values Handling

After the removal of unnecessary data items, there are still some blank cells. We try to fill up the missing values cases by case as described below:

- ‘fam-hx’ (family history), for almost all controls (except for 2 of them) are blanks. After consulting the research assistant in charge of data entry, all blanks are filled with the default value 1 (Nil). For a patient (CDNo=1), the value is missing, and we set it to the default value 1 (Nil).

- There are two missing values in column 'yr_momedu' (years of education: mother) and 'yr_fatedu' (years of education: father) for two patients (CDNo=54, 62). We fill the blank with the most likely value – the mean value of yr_momedu (6.77, rounded to 7), yr_fatedu(7.2, rounded to 7) in the patient group.
- 'occ_dad' (father's occupation) is missing for three cases (CDNo=83, 84, 143). They are set to 8 (unspecified).

Data Correction

Some data items are input wrongly by the data input officer. The fam_hx (family history) information is initially collected in the fam_hxsp (family history specify), and later coded into the fam_hx item. We find some inconsistency between these two items. For example, in the family history, "paternal aunt" should be coded as "2nd degree" (a relative who shares about 25% of genes with an individual in a family, e.g., uncle, aunt, cousin), instead of "1st degree" (a relative who shares about 50% of genes with an individual in a family, e.g., father, mother, siblings). Hence we correct errors in the fam_hx according to its original information fam_hxsp. We summarize our corrections in the Table 3.5:

Table 3.5 Data corrections

CD No	Originally Collected Data Fam_hxsp (famiy history specify)	Encoded Data Fam_hx: change from → to
2	Paternal uncle – schizophrenia	1st degree → 2nd degree
12	Cousin, aunt	1st degree → 2nd degree
26	Paternal aunt	1st degree → 2nd degree
47	Maternal grandmother committed suicide	1st degree → 2nd degree
55	Maternal uncle	1st degree → 2nd degree
57	Maternal aunt	1st degree → 2nd degree
75	Maternal aunt and uncle	1st degree → 2nd degree
93	Nephew	1st degree → 2nd degree
124	Uncle	1st degree → 2nd degree
130	Paternal grandpa	1st degree → 2nd degree
134	Paternal nephew/niece (OCD)	Other → 2nd degree
138	Paternal uncle	Other → 2nd degree
139	Paternal uncle – schizophrenia	Other → 2nd degree

Note: 1st degree: a relative who shares about 50% of genes with an individual in a family; 2nd degree: a relative who shares about 25% of genes with an individual in a family. **Abbreviation**: CD No, Compact Disk Number (used as case number).

Neurocognitive Tests Data

A total of 76 columns of information are collected for the 4 neurocognitive tests. We keep all raw scores, and remove derived ones such as percentage or percentile of the former scores. As a result of this exercise, 19 columns remain for further analysis (as listed in Table 3.3).

Table 3.6 Number of uncompleted and completed cases of neurocognitive test

Test	Total Case	Uncompleted	Completed
RPM	156	63	93
WAIS	156	61	95
CPT	156	67	89
WCST	156	63	93

Besides that, not all subjects take all four tests for various reasons. Some are unable to complete them, while some are uncooperative. The number of completed cases (ranging from 89 to 95) for all neurocognitive tests is listed in Table 3.6. Among all completed cases, 84 cases remain in common for all 4 tests. The completed cases consist of 59 patients (70.2%) and 25 controls (29.8%). The distributions of the remaining cases are illustrated in Figure 3.4. We will use these 84 completed cases in the model constructions later.

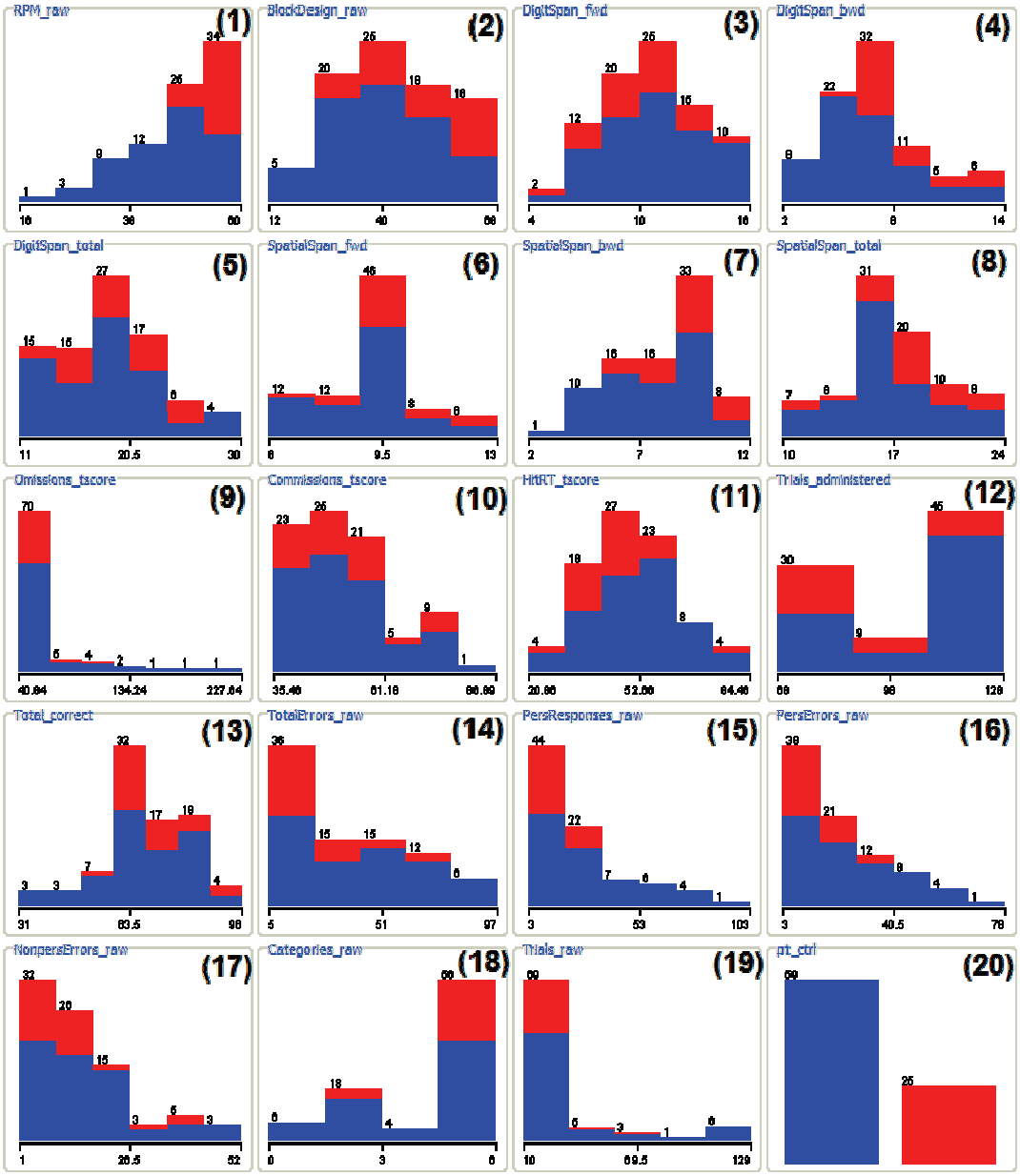

Figure 3.4 Distribution of neurocognitive test after removing missing values (N=84) (Blue: patient; Red: control. 1: RPM_raw, 2: BlockDesign_raw, 3: DigitSpan_fwd, 4: DigitSpan_bwd, 5: DigitSpan_total, 6: SpatialSpan_fwd, 7: SpatialSpan_bwd, 8: SpatialSpan_total, 9: Omissions_tscore, 10: Commissions_tscore, 11: HitRT_tscore, 12: Trials_administered, 13: Total_correct, 14: TotalErrors_raw, 15: PersResponses_raw, 16: PersErrors_raw, 17: NonpersErrors_raw, 18: Categories_raw, 19: Trials_raw. 20: pt_ctrl (patient/control). Descriptions of the items can be found in Appendix A)

In this section, we described the data preprocessing exercise. Specifically, we removed irrelevant columns, filled up missing values, corrected wrongly input data, and came up with clean datasets ready for further analysis.

3.6 Modeling Using Demographic Data and Clinical Data

The demographic data include patient's age, sex, weight, height, etc (see previous introduction in section 3.2). The clinical data include drug use, alcohol use and family history of psychiatric diseases, etc (see introduction in section 3.3). They are basic information collected at subjects recruiting time. They are available in all cases.

Table 3.7 Demographic and clinical data features

handed (handedness)	yrsedu_mum (years of education mum)
sex	yrsedu_dad (years of education dad)
ethnic	occupant (occupation)
father (father's ethnicity)	occ_dad (father's occupation)
pgfather (paternal grandfather's ethinicity)	occ_mum (mother's occupation)
pgmother (paternal grandfather's ethinicity)	med_prob (medical problems (past or current))
mother (mother's ethnicity)	sur_prob (surgical problems (past or current))
mgfather (maternal grandfather's ethinicity)	alcohol (alcohol use (past or current))
mgmother (maternal grandmother's ethnicity)	drug_use (drug use (past or current))
age	fam_hx (family history of mental illness)
marital (marital status)	height
edulevel (educational level)	weight
yrsedu (years of education)	

We start building our first schizophrenia classification model based on the demographic data and clinical data. There are 25 data features (Table 3.7), and 84 cases (see distribution in Figure 3.5). The characteristics of these cases are listed in Table 3.8. Pearson Chi-Square is used to test the independency of two categorical factorsand Independent Sample T Test (2-tailed) is used to compare the mean score

of two continuous factors. We can see that in the selected 84 cases, the patients and controls match on sex, handedness, ethnics, marital status, height and weight, except for age and education levels. The age is significantly different; and the education levels of the subject himself/herself and parents are significantly lower in patients than in controls.

Table 3.8 Characteristics of selected cases (N=84)

Characteristic	**Schizophrenia Patients (N=59) (70.2%)**	**Healthy Controls (N=25) (29.8%)**	**P Value**
Age	34.24 (SD 9.15)	28.56 (SD 6.89)	0.003 (SIG)*
Sex (F/M)	9/50 (15.3%/84.7%)	6/19 (24.0%/76.0%)	0.339 (NS)^
Handedness (Left/Right/Ambidextrous)	7/51/1 (11.9%/86.4%/1.7%)	2/23/0 (8.0%/92.0%/0%)	0.694 (NS)^
Ethnicity (Chinese/Malay/Indian/Others)	53/4/2/0 (89.8%/6.8%/3.4%/0%)	24/0/1/0 (96.0%/0%/4.0%/0%)	0.409 (NS)^
Marital Status (Single/Married/Widowed/Divorced)	49/9/0/1 (83.1%/15.3%/0%/1.7%)	19/6/0/0 (76.0%/24.0%/0%/0%)	0.526 (NS)^
Education level, years	11.22 (SD 2.67)	14.52 (SD 1.42)	< 0.001 (SIG)*
Father's education level, years	7.22 (SD 3.57)	9.04 (SD 3.31)	0.032 (SIG)*
Mother's education level, years	6.08 (SD 3.87)	8.72 (SD 3.79)	0.005 (SIG)*
Weight, kg	63.22 (SD 12.18)	66.70 (SD 11.39)	0.226 (NS)*
Height, m	1.69 (SD 0.07)	1.70 (SD 0.07)	0.494 (NS)*

Note: ^ Pearson Chi-Square; * Independent Sample T Test; **Abbreviations**: SD, Standard Deviation; F, Female; M, Male; SIG, Significant (P<0.05); NS, Not Significant

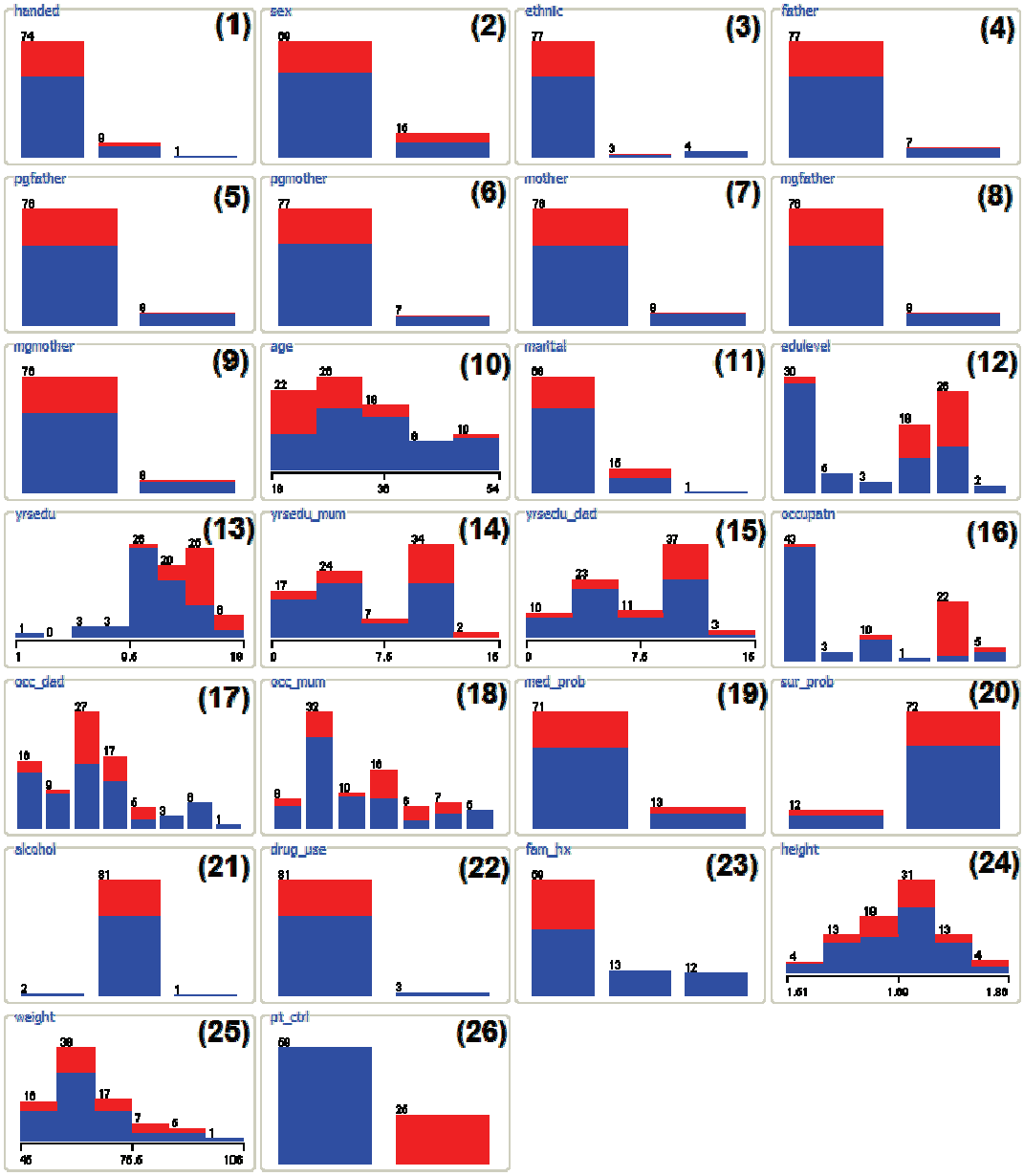

Figure 3.5 Distribution of demographic and clinical features (N=84)
(Blue: patient; Red: control. 1:handedness (right/left), 2: sex (male/female), 3: ethnic (Chinese/Indian/Malay), 4: father's ethnic (Chinese/Other), 5: paternal grandfather's ethnic (Chinese/Other), 6: paternal grantmother's ethnic (Chinese/Other), 7: mother's ethnic (Chinese/Other), 8: maternal grandfather's ethnic (Chinese/Other), 9: maternal grandmother's ethnic (Chinese/Other), 10: age (years), 11: marital status (single/married/divorced), 12: education level (Secondary/JC/Primary/University/Polytechnic/Vocational), 13: years of education, 14: years of education: mother, 15: years of education: father, 16: occupation (unemployed/manual labor/admin/homemaker/professional/other), 17: father's occupation (admin/unemployed/other/manual labor/professional/unspecified/passed away/home maker), 18: mother's occupation (admin/homemaker/unemployed/other/professional/manual labor/passed away), 19: medical problem (no/yes), 20:surgical problem (yes/no), 21: alcohol use (abuse/no/dependence), 22: drug use (no/abuse), 23: family history of mental illness (Nil/2nd degree/1st degree), 24: height (in m), 25: weight (in kg), 26: pt_ctrl (patient/control). Descriptions of data items can be found in Appendix A)

During data pre-processing, we have already removed irrelevant features such as medication and hospitalization information. If a model is constructed on all 25 data features for 84 cases, the case to feature ratio is about 3.4 to 1, which is substantially less than 10 to 1 (rule of 10), as suggested by many studies, for example (Arrindell & Ende, 1985; Bartlett, et al., 2001; Everitt, 1975; Nunnally, 1978). The model tends to be over-fitting or over-specific to the study samples and causes the lacking of generalizability due to too many parameters (Hair, Anderson, Tatham, & Black, 1995) or "the curse of dimensionality" (Hastie, Tibshirani, & Friedman, 2001). We will apply the feature selection technology to select only important features.

3.6.1 Feature Selection

Weka (Ver 3.4.13, University of Waikato, New Zealand) (Holmes, Donkin, & Witten, 1994; Witten & Frank, 2005) is an open source software package written in Java programming language. It contains implementations of many machine learning algorithms for data mining tasks. It also contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

Correlation-based Feature Subset algorithm selects significant features by evaluating and comparing each feature's predictive ability and degree of redundancy (M. A. Hall, 1998). We apply this algorithm implemented in the Weka software package for feature selection.

The features selected by this algorithm are: age, yrsedu (years of education), occupatn (occupation of patient), and fam_hx (family history of mental illness).

Some other feature selection algorithms have also been used, such as, Info Gain (Mitchell, 1997; Y. Yang & Pedersen, 1997), and Support Vector Machine (Bishop, 1996; Guyon, Weston, Barnhill, & Vapnik, 2002). The results are the same for our study. We will use the Correlation-based Feature Subset algorithm for the feature selection process throughout this work.

Among the selected features, yrsedu and occupatn are socioeconomic status of the patient; they are considered as the consequence of the sickness instead of reasons. Hence they shall be excluded from the selected feature list.

After removing another similar socioeconomic feature (edulevel: education level), the remaining selected features become: age and fam_hx. However, age difference is caused by the selection of subjects who have completed the neurocognitive tests; hence it shall not be included in model construction. Finally the only relevant feature to be used in model construction is the fam_hx (family history).

3.6.2 Definitions and Terminologies

Before we start to build up the classification models, we recall some definitions and terminologies here.

We want to identify “Patient”. In other words, our target is “Patient”. When we describe the status of a case, we mean its ground truth, or what it is actually. So when we say the status is true, we mean the subject is actually a patient; when we say the status is false, we mean the subject is actually a control.

When describing the test result, we use the terms: positive and negative. When we say the test result is positive, we mean the test result classifies it as a patient. When we say the test result is negative, we mean the test result classifies it as a control.

In this situation, we define true positive (TP), false positive (FP), true negative (TN), false negative (FN) as follows:

- True Positive (TP) test result: the test result classifies this case as a patient, and ground truth is actually a patient;
- False Positive (FP) test result: the test result classifies this case as a patient, and ground truth is actually a control;
- True Negative (TN) test result: the test result classifies this case as a control, and ground truth is actually a control;
- False Negative (FN) test result: the test result classifies this case as a control, and ground truth is actually a patient.

The following Table 3.9 shows the format of the confusion matrix of the supervised learning that will be used throughout the rest of thesis. For example, the total number of ground truth patients correctly classified as patients will be filled in the True Position (TP) blank.

Table 3.9 Confusion matrix of supervised learning

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
True Positive (TP)	False Negative (FN)	**Patient**	**Ground Truth**
False Positive (FP)	True Negative (TN)	**Control**	

The True Positive Rate, False Positive Rate, True Negative Rate, and False Negative Rate are defined as follows:

True Positive Rate: TP Rate = TP / (TP + FN) **(3.1)**

False Positive Rate: FP Rate = FP / (FP + TN) **(3.2)**

True Negative Rate: TN Rate = TN / (FP + TN) **(3.3)**

False Negative Rate: FN Rate = FN / (TP + FN) **(3.4)**

True Positive Rate is also known as Sensitivity. True Negative Rate is also known as Specificity.

Type I Error (also known as α error, False Positive Rate) and Type II Error (also known as β error, or False Negative Rate) are used to describe possible errors made in a statistical decision process.

Accuracy is defined as a measurement of how well a binary classification test correctly identifies or excludes a condition.

Accuracy = (TP + TN) / (TP + FP + TN + FN) **(3.5)**

Error Rate is the ratio of incorrectly classified cases to the total number of cases.

Error Rate = (FP + FN) / (TP + FP + TN + FN) **(3.6)**

3.6.3 Bayesian Network Classifier Evaluation

We choose the format of models from point of usefulness. Bayesian network model, alternating tree model and logistic regress model are considered. Different formats of models are generated and compared, but alternating tree model and logistic regression model produce lower classification accuracy than the Bayesian Model (see Section 7.2.5). Besides that, suitable presentation format is also a concern. Bayesian Network model is good for its simple graphical format (Acyclic Directed Graph (DAG)) to represent and understand the relationship between schizophrenia and the significant factors.

Bayesian Network is a graphic model that shows a set of interrelated factors (random variables) and their joint probability distributions. It is expressed as a Directed Acyclic Graph (DAG). Each node in the graph represents a random variable. Each arc represents a direct qualitative dependence relationship. And the local distribution of a node represents the quantitative strength of the dependence relation. (Cooper & Herskovitz, 1992; Howard, 1990; Jensen, 2001)

To classify a new case using the Bayesian Network Classifier, we first calculate the probability distribution of the case using equation (3-7).

$$P_{dist}(v) = P(pt\_ctrl = v)\frac{P(F_1 = u_1,\cdots,F_n = u_{n|} \mid pt\_ctrl = v)}{P(F_1 = u_1,\cdots,F_n = u_n)} \quad \textbf{(3.7)}$$

where P_{dist} is the distribution probability of patient or control; v takes value of "Patient" or "Control"; F is a chance node (a node in the Bayesian network that represents a factor, for example a test result) other than target node (pt_ctrl); u is the possible value for a chance node (factor); n is the number of chance nodes except for the target node pt_ctrl.

In case of Naïve Bayesian network (which consists of one target node and several child nodes, and no links between child nodes), since all factors are conditionally independent on pt_ctrl, equation (3-7) becomes:

$$P_{dist}(v) = P(pt\_ctrl = v)\prod_{i=1}^{n} P(F_i = u_i \mid pt\_ctrl = v) \quad \textbf{(3.8)}$$

Classification result is determined by equation (3-9):

$$pt\_ctrl_{classify} = \arg\max_{v} P_{dist}(v) \quad \textbf{(3.9)}$$

where $pt\_ctrl_{classify}$ is the classification result of a case: either "Patient" or "Control".

3.6.4 Baseline Model Construction

We use cases that complete all tests to construct our models. The total number of such cases is 84, including 59 patients and 25 controls. Hence the prior probability of patient in our sample space is 70.2%.

Weka software is used for model construction. We have tried different forms of model including Logistic Regression, Decision Tree and Bayesian Network (see a discussion in Chapter 7). It appears that Bayesian network models perform quite well in terms of classification accuracy. In addition, the Bayesian Network model has a simpler and clearer presentation format and is easier to interpret compared to the other two models. Hence, we decide to use Bayesian Network for all our model construction.

Based on the selected feature (fam_hx), we construct a Bayesian Network classification model with the heuristic local K2 searching algorithm (Cooper & Herskovitz, 1992) by using the simple Bayes estimator (Bouckaert, 2004) to calculate the conditional probability table and comparing the posterior probabilistic ratio of any pair of possible Bayesian Network structure at the given dataset, which learns the structure effectively. Other searching algorithms, such as Greedy Search (Chickering, 2002), which searches for local maximum from initial structure, are also attempted. However, In our study, the same results are generated. K2 algorithm is often used by researchers since it can be implemented easily and evaluated fast. We will use the K2 searching algorithm throughout all Bayesian Network Model construction process in this study.

Figure 3.6 Bayesian network model on clinical data

A Bayesian Network model is generated, as illustrated in Figure 3.6. This is a very basic model, which has only two nodes. This over simplified model demonstrates the fact that there is a stronger association between schizophrenia and the family history than other features (such as sex, ethnic), as people have already pointed out in (R. Murray, et al., 2003).

The probability distribution table generated for the model is displayed in Table 3.10.

Table 3.10 Probability distribution of fam_hx

Pt_ctrl	fam_hx		
	Nil	2nd degree	1st degree
Patient	0.5702	0.2231	0.2066
Control	0.9623	0.0189	0.0189

Note: 1st degree: a relative who shares about 50% of genes with an individual in a family; 2nd degree: a relative who shares about 25% of genes with an individual in a family.

Since the sample data collection is very costly ($1,497 per case at private rate, see Section 7.1), the validation of the model is done by using the 10-fold cross-validation method, where all cases are randomly split into 10 subsets, and every subset is used as validation set to validate the model generated from the other 9 subsets (as training set) (Kohavi, 1995). From the validation results (Table 3.11, Table 3.12), we can see that all cases are classified as patients, so that the Type I error is 100%. The accuracy is 70.2%, which is actually the same as the prior probability of patient. We will use this model as the baseline model, and gradually enhance it by adding other features.

Table 3.11 Confusion matrix (clinical data: fam_hx)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
59 (TP)	0 (FN)	**Patient (59)**	**Ground Truth**
25 (FP)	0 (TN)	**Control (25)**	

Table 3.12 Summary of model (clinical data: fam_hx)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	59
Incorrectly Classified Instances	25
Accuracy	**70.2%**
Error Rate	29.8%
True Positive Rate (Patient classified as patient)	100%
Type I Error Rate (Control classified as patient)	100%
True Negative Rate (Control classified as control)	0%
Type II Error Rate (Patient classified as control)	0%

As a comparison, we tried to use the feature yrsedu (years of education) to build another model. The results are listed in Table 3.13 and Table 3.14. This model does not classify all cases as patients (as the baseline model does); actually it can correctly classify 7 control cases (TN). It also incorrectly classifies 11 patients as controls (FN). However, the total correctly classified number of cases is 55 (TP+TN). Though this model has a lower Type I error (72.0%) than the baseline model (100%), its accuracy is only 65.5%, which is worse than that of the baseline model (70.2%).

From the comparison, we can see that, even though the baseline model has a Type I error of 100%, it can still be considered as the best choice for the aim of high accuracy, at the current restriction (with only one feature: fam_hx). In other words, the extremely high Type I error (100%) of the baseline model is not a coincidence, it is the result of the best fitting of model for the purpose of achieving the highest accuracy.

In fact, because of the availability of the feature it used (family history is considered as also available), this model serves as a starting point (the meaning of baseline). We can add additional features into the baseline models: (1) to increase the accuracy and (2) to decrease the Type I error. Since the baseline model's Type I error is 100%, we have to point out that this model alone shall not be applied in practice.

As we have discussed earlier, being a consequence of the schizophrenia disease, yrsedu shall not be included into our classification models as a predictor, and we will not consider it in our further model constructions.

Table 3.13 Confusion matrix (yrsedu)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
48 (TP)	11 (FN)	**Patient (59)**	**Ground Truth**
18 (FP)	7 (TN)	**Control (25)**	

Table 3.14 Summary of model (yrsedu)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	55
Incorrectly Classified Instances	29
Accuracy	**65.5%**
Error Rate	34.5%
True Positive Rate (Patient classified as patient)	81.4%
Type I Error Rate (Control classified as patient)	72.0%
True Negative Rate (Control classified as control)	28.0%
Type II Error Rate (Patient classified as control)	18.6%

In this section, we first use the feature selection method to select the important features from demographic data and clinical data. Then we build a baseline Bayesian Network model with accuracy rate 70.2%. Next we will include neurocognitive test results in the model.

3.7 Modeling Using Neurocognitive Tests Results

Four neurocognitive tests on study subjects, namely RPM, WAIS, CPT and WCST have been done, which generates 19 data features (See Table 3.3). Among all 156 cases, 84 participants have completed all 4 tests, including 59 patients (70.2%), and 25 controls (29.8%). We will use this subset for model construction.

In the previous section, we select only one significant feature from demographic data and clinical data, the fam_hx (family history). Since any neurocognitive test can be done separately and independently, various models should be constructed to

reflect this situation. In this section, we will use fam_hx and various combinations of neurocognitive tests to build more comprehensive models. However, we will examine these tests individually before we combine them with clinical data.

3.7.1 Neurocognitive Tests Only

By using the method described in section 3.6, we select significant features from each of the 4 neurocognitive tests, and build 4 Bayesian Network models based on each of them separately by using the same approach described in Section 3.6.4.

Specifically, feature RPM_raw is selected from the RPM test, since it is the only result in the test. Feature DigitSpan_bwd is selected from the WAIS test results (BlockDesign_raw, DigitSpan_fwd, DigitSpan_bwd, DigitSpan_total, SpatialSpan_fwd, SpatialSpan_bwd, and SpatialSpan_total); feature Omissions_tscore is selected from the CPT test results (Omissions_tscore, Commissions_tscore, HitRT_tscore); and features PersResponses_raw and PersError_raw are selected from the WCST test results (Trials_administered, Total_correct, TotalErrors_raw, PersResponses_raw, PersErrors_raw, NonpersErrors_raw, Categories_raw, Trials_raw).

Four Bayesian Network classification models are constructed by using significant features from the neurocognitive tests separately, and their results are summarized in Table 3.15, Table 3.16, Table 3.17 and Table 3.18 for RPM, WAIS, CPT and WCST tests respectively.

The model on RPM test has an accuracy of 66.7%, which is even lower than the baseline model (70.2%). This shows that when this test is applied alone, the effect in reaching correct diagnosis is worse than the family history. However, the Type I error (56.0%) decreases compared to the baseline model (100%), which means it can correctly identify some control cases, whereas the baseline model never does.

The model on WAIS test results (DigitSpan_bwd) alone has the same accuracy (70.2%) as the baseline model. In fact, it also classifies all controls as patients incorrectly as the baseline model does; hence it's Type I error is also 100%. This test alone has the same effect in terms of classification accuracy as the family history. Again, this model shall not be applied in practice because of its high Type I error.

Similarly, the model on CPT test results alone generates a classification accuracy of 70.2%, and a Type I error of 100%.

In the case of WCST test, the model accuracy is 56.0%, which is the lowest of all 4 models. However, it is able to correctly identify some controls (21 cases), which makes it's Type I error 16.0%, also the lowest of all. This shows its ability of being a potential predictor in further model construction.

Table 3.15 Summary of model on RPM test results (RPM_raw)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	56
Incorrectly Classified Instances	28
Accuracy	**66.7%**
Error Rate	33.3%
True Positive Rate (Patient classified as patient)	76.3%
Type I Error Rate (Control classified as patient)	56.0%
True Negative Rate (Control classified as control)	44.0%
Type II Error Rate (Patient classified as control)	23.7%

Table 3.16 Summary of model on WAIS test results (DigitSpan_bwd)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	59
Incorrectly Classified Instances	25
Accuracy	**70.2%**
Error Rate	29.8%
True Positive Rate (Patient classified as patient)	100.0%
Type I Error Rate (Control classified as patient)	100.0%
True Negative Rate (Control classified as control)	0.0%
Type II Error Rate (Patient classified as control)	0.0%

Table 3.17 Summary of model on CPT test results (Omission_tscore)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	59
Incorrectly Classified Instances	25
Accuracy	**70.2%**
Error Rate	29.8%
True Positive Rate (Patient classified as patient)	100.0%
Type I Error Rate (Control classified as patient)	100.0%
True Negative Rate (Control classified as control)	0.0%
Type II Error Rate (Patient classified as control)	0.0%

Table 3.18 Summary of model on WCST test results (PersResponse_Raw + PersError_raw)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	47
Incorrectly Classified Instances	37
Accuracy	**56.0%**
Error Rate	44.0%
True Positive Rate (Patient classified as patient)	44.1%
Type I Error Rate (Control classified as patient)	16.0%
True Negative Rate (Control classified as control)	84.0%
Type II Error Rate (Patient classified as control)	55.9%

In summary, four neurocognitive tests are investigated separately. Two tests show similar contribution to classification accuracy as the baseline model. The other two tests show lower classification accuracy, however they can correctly identify some normal controls. In the next few sections, we will add these tests into the baseline model, and construct new models based on the combined features.

3.7.2 Clinical Data + RPM

We first combine all the clinical data and the RPM test results (RPM_raw). The following features remain after the feature selection:

- fam_hx
- RPM_raw

Weka software is used for model construction. Based on the selected features, we construct a Bayesian Network classification model with the local K2 searching algorithm (Cooper & Herskovitz, 1992).

The model is illustrated in Figure 3.7. A classification accuracy of 82.1% is achieved, which is substantially higher than the baseline model accuracy (70.2%). This is a promising result, as we start to see the power of combining factors from different categories. Other rates are listed in Table 3.20. Compared to the baseline model's Type I error (100%), this model is able to classify 20 cases of controls correctly (Table 3.19), and it's Type I error drops to 20%. Note that this Bayesian Network model degenerates to Naïve Bayesian Network. That is because the two features (fam_hx and RPM_raw) are conditionally independent given pt_ctrl, since only highly independent features remain after the feature selection step (which selects features that generates lowest redundancy.

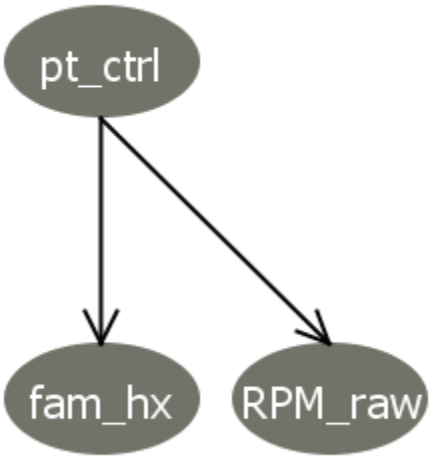

Figure 3.7 Model on clinical data + RPM

Table 3.19 Confusion matrix (clinical data + RPM)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
49 (TP)	10 (FN)	**Patient (59)**	**Ground Truth**
5 (FP)	20 (TN)	**Control (25)**	

Table 3.20 Summary of model on clinical data + RPM

Item	Value
Total Number of Instances	84
Correctly Classified Instances	69
Incorrectly Classified Instances	15
Accuracy	**82.1%**
Error Rate	17.9%
True Positive Rate (Patient classified as patient)	83.1%
Type I Error Rate (Control classified as patient)	20.0%
True Negative Rate (Control classified as control)	80.0%
Type II Error Rate (Patient classified as control)	16.9%

3.7.3 Clinical Data + WAIS

Now we add the WAIS test results (BlockDesign_raw, DigitSpan_fwd, DigitSpan_bwd, DigitSpan_total, SpatialSpan_fwd, SpatialSpan_bwd, and SpatialSpan_total) to all clinical data. The following features remain after the feature selection:

- fam_hx
- DigitSpan_bwd

A Bayesian Network model is built on these features (as illustrated in Figure 3.8). It contains 3 nodes. Since fam_hx and DigitSpan_bwd are conditionally independent given pt_ctrl, the model degenerates to a Naïve Bayesian network model. The accuracy of this model is 79.8%, which is substantially higher than the baseline model accuracy (70.2%). Other rates are listed in Table 3.22. This model further increases the number of correctly classified controls to 24 (Table 3.21), which leads to an even lower Type I error of 4%. As a trade-off, 16 patients are incorrectly classified as controls, which causes a 27.1% Type II error. This model seems to be good at identifying controls.

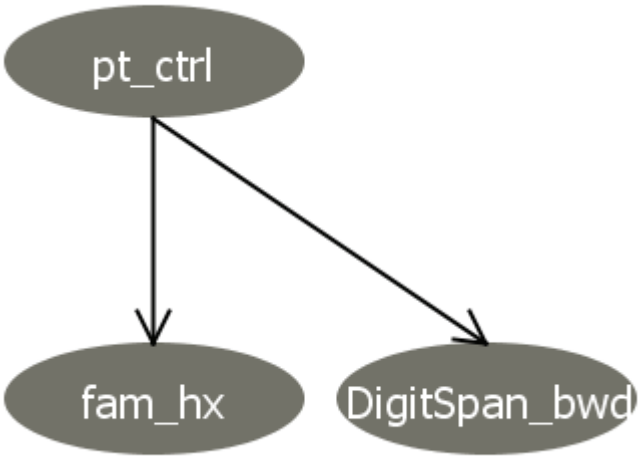

Figure 3.8 Model on clinical data + WAIS

Table 3.21 Confusion matrix (clinical data + WAIS)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
43 (TP)	16 (FN)	**Patient (59)**	**Ground Truth**
1 (FP)	24 (TN)	**Control (25)**	

Table 3.22 Summary of model on clinical data + WAIS

Item	Value
Total Number of Instances	84
Correctly Classified Instances	67
Incorrectly Classified Instances	17
Accuracy	**79.8%**
Error Rate	20.2%
True Positive Rate (Patient classified as patient)	72.9%
Type I Error Rate (Control classified as patient)	4.0%
True Negative Rate (Control classified as control)	96.0%
Type II Error Rate (Patient classified as control)	27.1%

3.7.4 Clinical Data + CPT

We next combine all clinical data and the CPT test results (Omissions_tscore, Commissions_tscore, HitRT_tscore). The following features remain after the feature selection:

- fam_hx

No CPT test result appears in the selected feature list. This means CPT test's contribution to the classification is insufficient. However we can still examine the effect of CPT, by including fam_hx and CPT test results as factors to build a Bayesian Network classification model. The following results are generated as in Table 3.23 and Table 3.24. We can see that it does not change the classification results of model on fam_hx alone. The accuracy is still 70.2%.

In fact, by examining the conditional probabilities of all factors from CPT test (Table 3.25), we notice that the probability distribution of classification result (pt_ctrl) is always 1 when it takes value of either "patient" or "control", regardless of the value of Omissions_tscore, Commissions_tscore and HitRT_tscore. That means the CPT test results have no relation with the classification target node (pt_ctrl), and they don't affect the probability distribution at all. Hence, the CPT test shall be excluded from the model construction.

From clinical point of view, the result of this model suggests that the attention and impulsivity functions assessed by the CPT test do not show significant difference

between patients and controls. These two functions do not increase the classification accuracy when combined with other factor (fam_hx).

Table 3.23 Confusion matrix (clinical data + CPT)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
59 (TP)	0 (FN)	**Patient (59)**	**Ground Truth**
25 (FP)	0 (TN)	**Control (25)**	

Table 3.24 Summary of model on clinical data + CPT

Item	Value
Total Number of Instances	84
Correctly Classified Instances	59
Incorrectly Classified Instances	25
Accuracy	**70.2%**
Error Rate	29.8%
True Positive Rate (Patient classified as patient)	100.0%
Type I Error Rate (Control classified as patient)	100.0%
True Negative Rate (Control classified as control)	0.0%
Type II Error Rate (Patient classified as control)	0.0%

Table 3.25 Probability distribution tables of factors from CPT

Pt_ctrl	Omissions_tscore
Patient	1
Control	1

Pt_ctrl	Commissions_tscore
Patient	1
Control	1

Pt_ctrl	HitRT_tscore
Patient	1
Control	1

3.7.5 Clinical Data + WCST

We combine clinical data with the WCST test results (Trials_administered, Total_correct, TotalErrors_raw, PersResponses_raw, PersErrors_raw, NonpersErrors_raw, Categories_raw, Trials_raw). The following features remain after the feature selection:

- fam_hx
- PersResponses_raw

A model is built on these features (as illustrated in Figure 3.9). Its accuracy is 75.0%. Other results are listed in Table 3.26 and Table 3.27. The low Type I error (4.0%) shows its good ability in identifying controls. In contrast, a big portion of patients (20) are wrongly classified as controls, which causes a high Type II error (33.9%). This model seems to be more biased to controls.

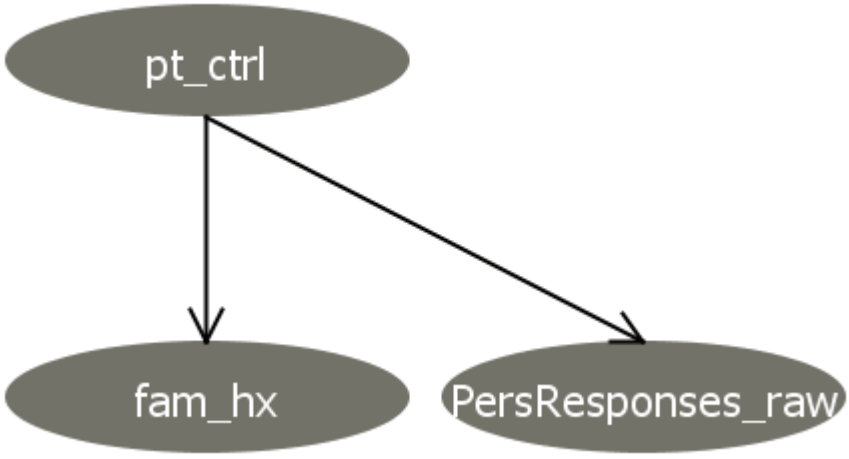

Figure 3.9 Model on clinical data + WCST

Table 3.26 Confusion matrix (clinical data + WCST)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
39 (TP)	20 (FN)	**Patient (59)**	**Ground Truth**
1 (FP)	24 (TN)	**Control (25)**	

Table 3.27 Summary of model on clinical data + WCST

Item	Value
Total Number of Instances	84
Correctly Classified Instances	63
Incorrectly Classified Instances	21
Accuracy	**75.0%**
Error Rate	25.0%
True Positive Rate (Patient classified as patient)	66.1%
Type I Error Rate (Control classified as patient)	4.0%
True Negative Rate (Control classified as control)	96.0%
Type II Error Rate (Patient classified as control)	33.9%

3.7.6 Clinical Data + RPM + WAIS

We combine all clinical data with two neurocognitive tests, namely, RPM test result (RPM_raw) and the WAIS test results (BlockDesign_raw, DigitSpan_fwd, DigitSpan_bwd, DigitSpan_total, SpatialSpan_fwd, SpatialSpan_bwd, and SpatialSpan_total). The following features remain after the feature selection:

- fam_hx
- RPM_raw
- DigitSpan_bwd

A model is built on these features (as illustrated in Figure 3.10). Its accuracy is 84.5%. This is substantially higher than the baseline model (70.2%). Other results are listed in Table 3.28 and Table 3.29.

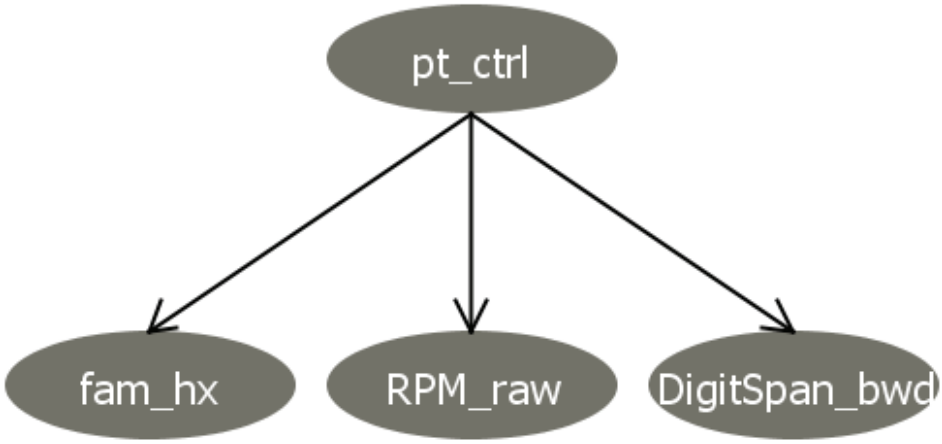

Figure 3.10 Model on clinical data + RPM + WAIS

Table 3.28 Confusion matrix (clinical data + RPM + WAIS)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
52 (TP)	7 (FN)	**Patient (59)**	**Ground Truth**
6 (FP)	19 (TN)	**Control (25)**	

Table 3.29 Summary of model on clinical data + RPM + WAIS

Item	Value
Total Number of Instances	84
Correctly Classified Instances	71
Incorrectly Classified Instances	13
Accuracy	**84.5%**
Error Rate	15.5%
True Positive Rate (Patient classified as patient)	88.1%
Type I Error Rate (Control classified as patient)	24.0%
True Negative Rate (Control classified as control)	76.0%
Type II Error Rate (Patient classified as control)	11.9%

3.7.7 Clinical Data + RPM + WCST

We combine all clinical data with another two neurocognitive tests, namely, WAIS test and the WCST test. The following features remain after the feature selection:

- fam_hx
- RPM_raw
- PersResponses_raw

A model is built on these features (as illustrated in Figure 3.11). Its accuracy is 83.3%. Again a substantial increment in accuracy is achieved compared to the baseline model. Other results are listed in Table 3.30 and Table 3.31.

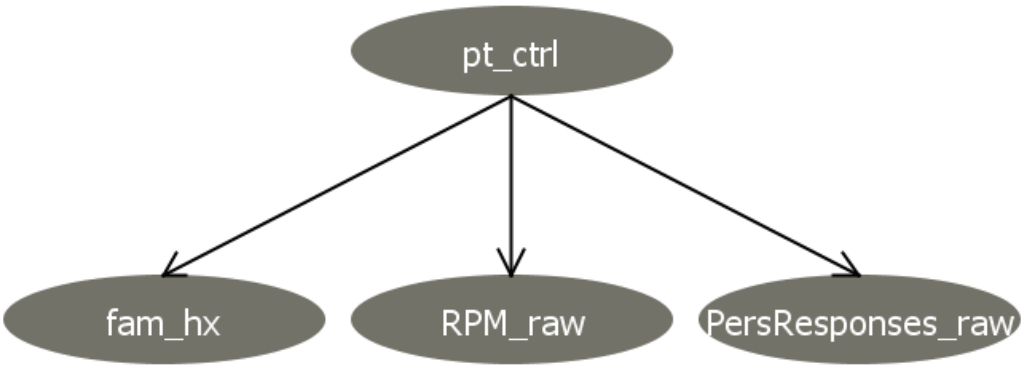

Figure 3.11 Model on clinical data + RPM + WCST

Table 3.30 Confusion matrix (clinical data + RPM + WCST)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
51 (TP)	8 (FN)	**Patient (59)**	**Ground Truth**
6 (FP)	19 (TN)	**Control (25)**	

Table 3.31 Summary of model on clinical data + RPM + WCST

Item	**Value**
Total Number of Instances	84
Correctly Classified Instances	70
Incorrectly Classified Instances	14
Accuracy	**83.3%**
Error Rate	16.7%
True Positive Rate (Patient classified as patient)	86.4%
Type I Error Rate (Control classified as patient)	24.0%
True Negative Rate (Control classified as control)	76.0%
Type II Error Rate (Patient classified as control)	13.6%

3.7.8 Clinical Data + WAIS + WCST

We combine all clinical data with another two neurocognitive tests, namely, WAIS test and WCST test. The following features remain after the feature selection:

- fam_hx
- DigitSpan_bwd
- PersResponses_raw

A model is built on these features (as illustrated in Figure 3.12). Its accuracy is 84.5%. Compared to the baseline model, the accuracy improvement is also substantial. Other results are listed in Table 3.32 and Table 3.33.

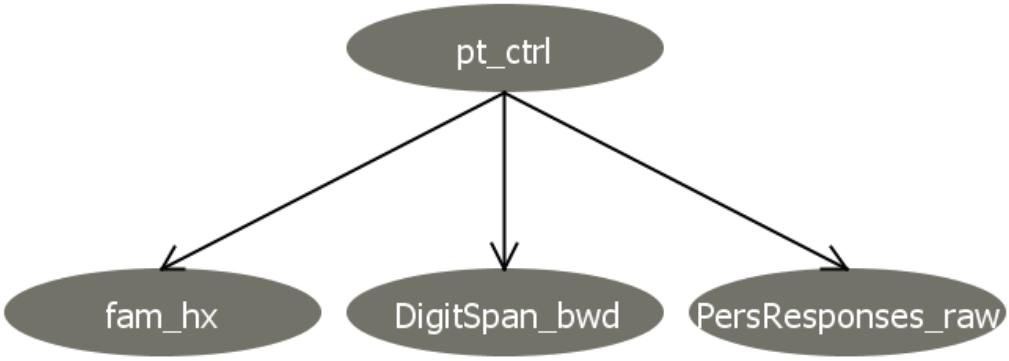

Figure 3.12 Model on clinical data + WAIS + WCST

Table 3.32 Confusion matrix (clinical data + WAIS + WCST)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
48 (TP)	11 (FN)	**Patient (59)**	**Ground Truth**
2 (FP)	23 (TN)	**Control (25)**	

Table 3.33 Summary of model on clinical data + WAIS + WCST

Item	Value
Total Number of Instances	84
Correctly Classified Instances	71
Incorrectly Classified Instances	13
Accuracy	**84.5%**
Error Rate	15.5%
True Positive Rate (Patient classified as patient)	81.4%
Type I Error Rate (Control classified as patient)	8.0%
True Negative Rate (Control classified as control)	92.0%
Type II Error Rate (Patient classified as control)	18.6%

3.7.9 Clinical Data + RPM + WAIS + WCST (All Tests)

Finally we combine all clinical data with all three tests: RPM, WAIS and WCST. The following features remain after the feature selection:

- fam_hx
- RPM_raw
- DigitSpan_bwd
- PersResponses_raw

A model is built on these features (as illustrated in Figure 3.13). Its accuracy is 85.7%. Other results are listed in Table 3.34 and Table 3.35. This model has the highest accuracy among all models based on clinical information and various combinations of neurocognitive tests. It also achieves the highest sensitivity (91.5%), which seems to be very sensitive to patients. However, as a trade-off, 7 controls are wrongly classified as patients, which causes a Type I error of 28.0%.

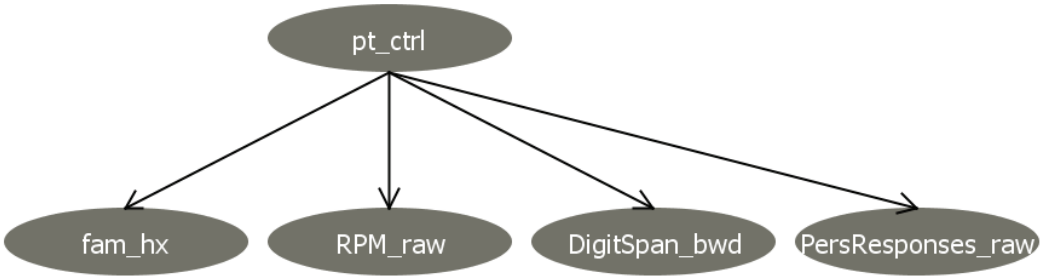

Figure 3.13 Model on clinical data + RPM + WAIS + WCST

Table 3.34 Confusion matrix (clinical data + RPM + WAIS + WCST)

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
54 (TP)	5 (FN)	**Patient (59)**	**Ground Truth**
7 (FP)	18 (TN)	**Control (25)**	

Table 3.35 Summary of model on clinical data + RPM + WAIS + WCST

Item	**Value**
Total Number of Instances	84
Correctly Classified Instances	72
Incorrectly Classified Instances	12
Accuracy	**85.7%**
Error Rate	14.3%
True Positive Rate (Patient classified as patient)	91.5%
Type I Error Rate (Control classified as patient)	28.0%
True Negative Rate (Control classified as control)	72.0%
Type II Error Rate (Patient classified as control)	8.5%

3.7.10 Summary of All Models

Since neurocognitive tests may be done on patients separately, not all testes results are always available. So we tried to build models on clinical data plus different combinations of neurocognitive tests.

We found that CPT test results do not contribute to the model performance. The patients and controls' test results have no differences. Hence this test is not necessary in differentiating patients with controls.

We summarize the results for various models in Table 3.36. The baseline model on clinical data only (Model C) has an accuracy of 70.2%, which is the same as the prior probability of patient in our sample space. In fact, it classifies every case as patient, so the sensitivity and type I error are both 100%. Hence clinical information alone is not practically sufficient in schizophrenia diagnosis by using this model.

All other models have accuracy higher than the baseline probability. Their accuracy gains range from 4.8% (Model C+WC) to 15.5% (Model C+R+WA+WC).

We observe that when a model contains 2 or more neurocognitive tests, its accuracy is usually more than 10% higher than the baseline model (Figure 3.14). In general, models containing two neurocognitive tests achieve higher accuracy than

models containing one. Also, the model with 3 tests has higher accuracy than models with two, though the increase is not as big as the former.

Table 3.36 Summary of models on clinical data + neurocognitive tests

Model→	C	C+R	C+WA	C+WC	C+R+WA	C+R+WC	C+WA+WC	C+R+WA+WC
Nr	84	84	84	84	84	84	84	84
Cor	59	69	67	63	71	70	71	72
Incor	0	15	17	21	13	14	13	12
Acc	**70.2%**	**82.1%**	**79.8%**	**75.0%**	**84.5%**	**83.3%**	**84.5%**	**85.7%**
Err	29.8%	17.9%	20.2%	25.0%	15.5%	16.7%	15.5%	14.3%
Sen	100.0%	83.1%	72.9%	66.1%	88.1%	86.4%	81.4%	91.5%
Type I	100.0%	20.0%	4.0%	4.0%	24.0%	24.0%	8.0%	28.0%
Spe	0.0%	80.0%	96.0%	96.0%	76.0%	76.0%	92.0%	72.0%
Type II	0.0%	16.9%	27.1%	33.9%	11.9%	13.6%	18.6%	8.5%

Abbreviations: Nr, Total Number of Instances; Cor, Correctly Classified Instances; Incor, Incorrectly Classified Instances; Acc, Accuracy; Err, Error Rate; Sen, Sensitivity; TPR, True Positive Rate (Patient classified as patient); Type I, Type I Error Rate (Control classified as patient); Spe, Specificity, TNR, True Negative Rate (Control classified as control); Type II, Type II Error Rate (Patient classified as control); C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Most models have Type II error below 30% except for Model C+WC (Figure 3.15), which has the highest Type II error of 33.9%. This model tends to classify patients as healthy controls.

One model has extremely high Type I error. Model C's Type I error is 100%. It always classifies controls as patients in our cross-validation test. This simple model takes into consideration of only patient's family history, and its prediction is not practically reliable.

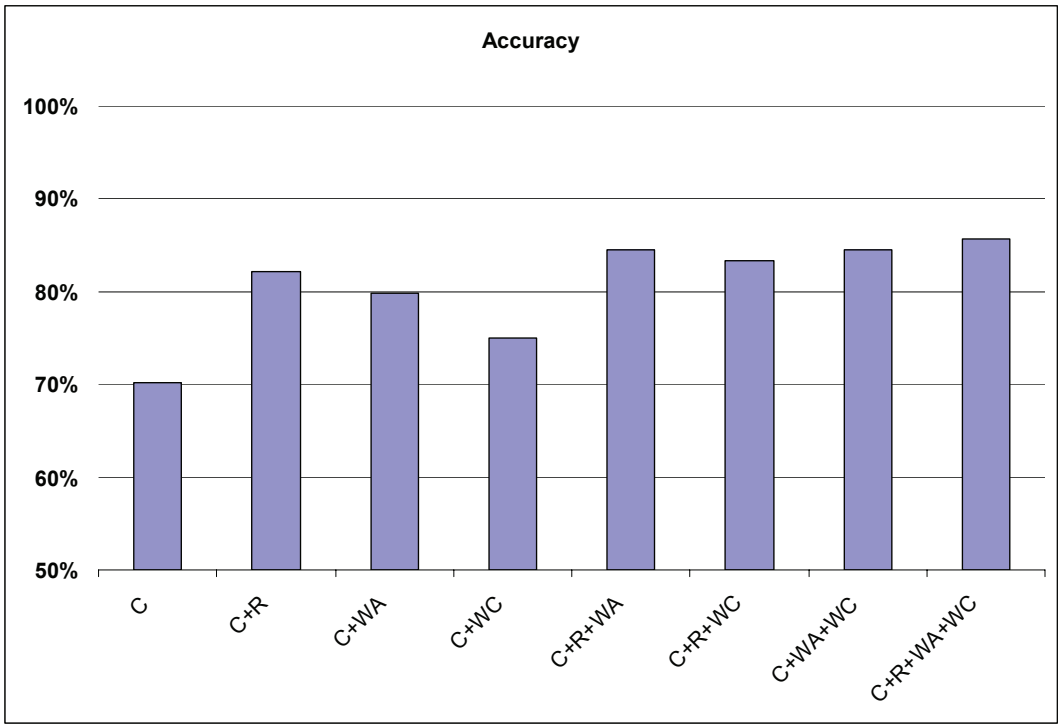

Figure 3.14 Accuracy chart for models on clinical data + neurocognitive tests

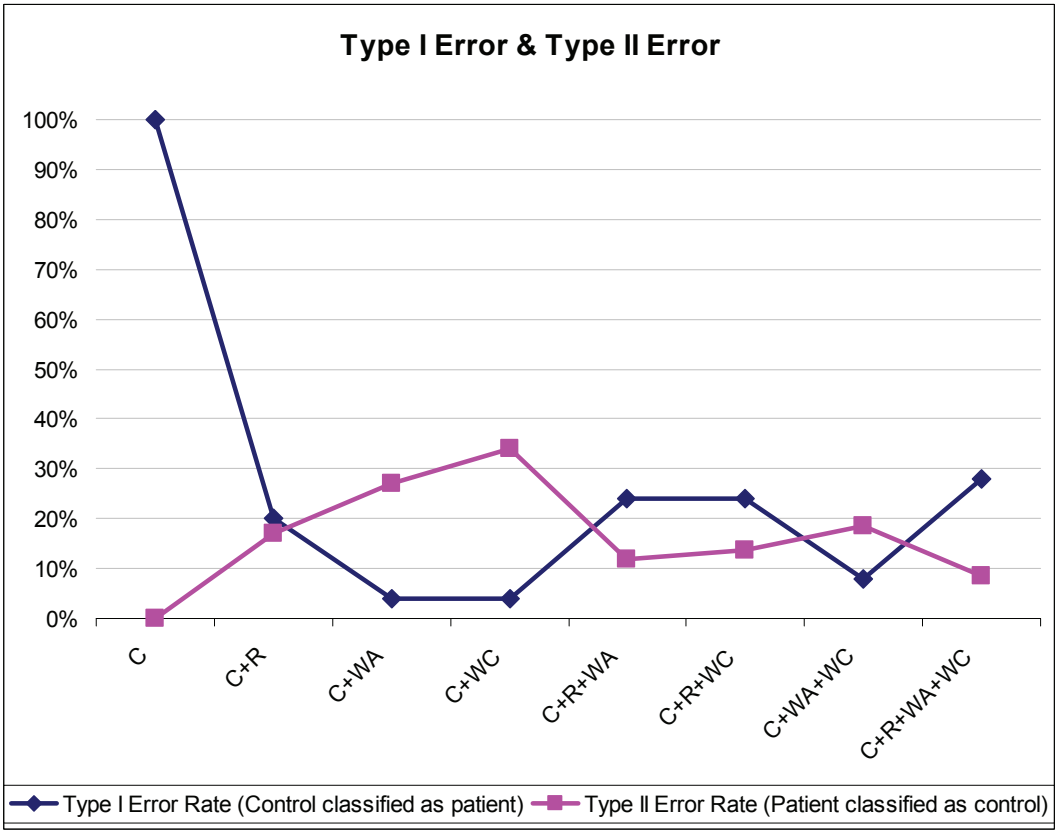

Figure 3.15 Type I and II error chart for models on clinical data + neurocognitive tests

3.8 Conclusions

We have recruited 156 subjects (92 patients, and 64 healthy controls). Their neuroinformatics data are collected: including demographic information, clinical information such as family history, medication information, clinical scores and neurocognitive tests results. In total there are 211 data features.

Data preprocessing is done to eliminate features that are not appropriate in constructing schizophrenia classification models. Forty-five features remain as candidate predictors for the classification model. Then the missing values in the datasets are filled up by reasonable estimations from domain knowledge (such as most likely values). Errors in data input (fam_hx) are corrected according to the original data (fam_hxsp).

Feature selection is done on demographic and clinical information. Only fam_hx (family history of psychiatric disease) is selected to be the significant feature. After feature selection, a basic classification model in the form of Bayesian Network is generated on the significant feature. Then various combinations of neurocognitive tests results together with clinical data features are used to build a set of classification models. PCT test is found not contributing to model accuracy gain at all.

All classification models built on clinical data plus neurocognitive tests have better accuracy than the baseline model. Their accuracy gain ranges from 4.8% to 15.5%. The best performing model has an accuracy of 85.7%. It consists of features from

clinical data, and all three neurocognitive tests (RPM test, WAIS test, and WCST test). Specifically, the features are:

- fam_hx
- RPM_raw
- DigitSpan_bwd
- PersResponses_raw

We notice that, family history always remains in all the models. This means it is an important factor in distinguishing patients and controls. This finding aligns well with literature (R. Murray, et al., 2003).

The summary of the three significant neurocognitive tests comparison results between patients and controls is displayed in Table 3.37 and Figure 3.16. Independent Sample T Test (2-tailed) is used to compare the mean score between patients and controls.

Table 3.37 Neurocognitive tests results comparison

Neurocognitive Test	Mean		Mean Difference (Patient – Control)	P Value (2-tailed)
	Patient (N=59)	Control (N=25)		
RPM_raw	45.85 (SD 9.28)	54.48 (SD 3.50)	-8.633	<0.001 (SIG)^
DigitSpan_bwd	7.05 (SD 2.66)	8.64 (SD 2.25)	-1.589	0.007 (SIG)^
PersResponse_raw	29.20 (22.97)	13.16 (SD 9.23)	16.04	0.001 (SIG)^

Note: ^Independent Samples T-Test. **Abbreviations**: SD, Standard Deviation; SIG, significant (P<0.05); NS, Not significant.

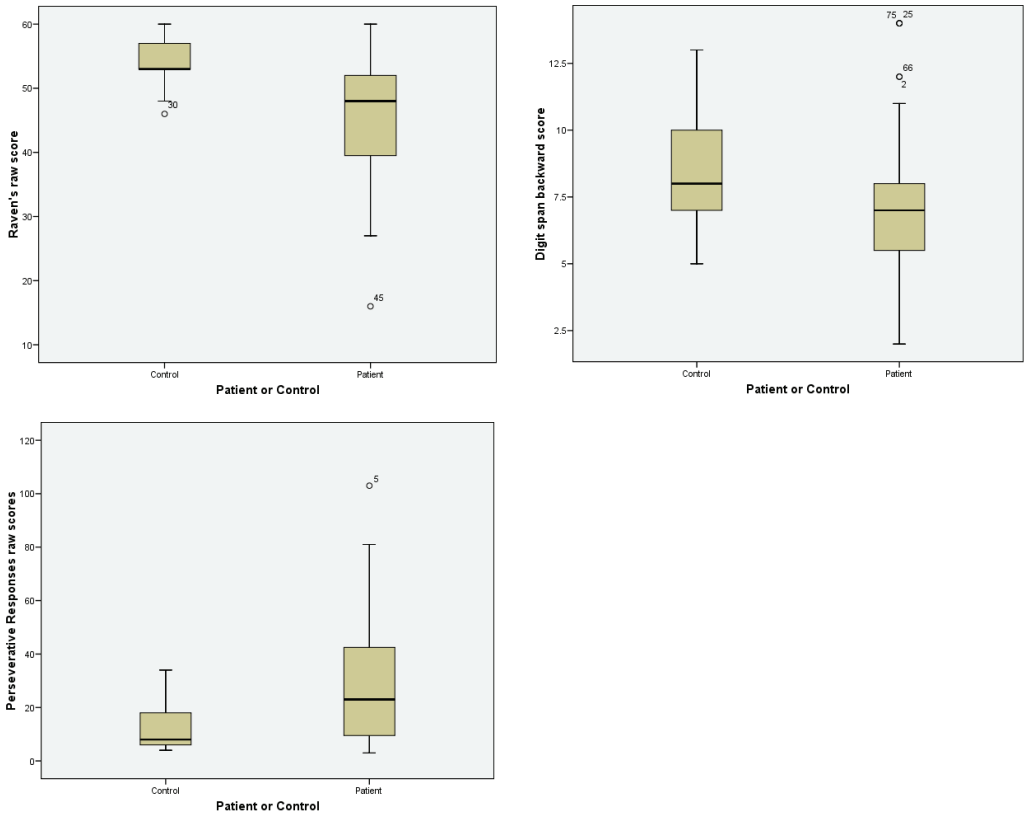

Figure 3.16 Box plot of selected neurocognitive tests results grouped by patient / control Upper left: RPM_raw; Upper right: DigitSpan_bwd; Lower-left: PersRepsonses_raw. Note that scales for each box plot are different.

We find that RPM (RPM_raw) and WAIS (DigitSpan_bwd) are significant factors in our schizophrenia model; they are related to a person's intellectual abilities.

RPM_raw is the number of missing patterns correctly identified by a person in a test. DigitSpan_bwd is the longest of number of digits that a person can repeat correctly in the reverse order after they are announced. In our study subjects, patients' mean RPM_raw score (45.85, SD: 9.28) is significantly lower by 8.63 than that of controls (54.48, SD 3.50). Likewise, patients' mean DigitSpan_bwd (7.05, SD 2.66) is also significantly lower by 1.59 than controls (8.64, SD: 2.25). That means patients have significantly lower intellectual abilities in terms of non-verbal logic (as assessed by RPM) and memory capacity (as measured by Digit

Span sub-test of WAIS) than normal controls. Some other studies (David, Malmberg, Brandt, Allebeck, & Lewis, 1997), (Zammit, et al., 2004) also demonstrated low intellectual ability as a risk factor for schizophrenia.

WCST (PersResponses_raw) is a significant factor in schizophrenia classification too. Reduced executive capability (tested by WCST), is believed to relate to frontal lobe dysfunction of schizophrenia patients. Our study results show that schizophrenia patients (29.20, SD: 22.97) have significantly higher perseverative response (by 16.04) than normal controls (13.16, SD: 9.23), which suggests patients show lower adaptability in learning new rules. Many studies have also shown the similar trends (Abbruzzese, Bellodi, Ferri, & Scarone, 1995; Mahurin, Velligan, & Miller, 1998; Pae, et al., 2004).

CPT test doesn't affect the model, which suggests that the differences in capability of attention and impulsivity between patients and controls are not significantly enough when compared with other factors (Table 3.38). Independent Sample T Test (2-tailed) is used to compare the mean scores between patients and controls. Actually even if we include CPT test into our models, the accuracy does not increase, as discussed in section 3.7.4 before.

Table 3.38 CPT test results comparison

CPT Test Results	Mean		Mean Difference	P Value
	Patient (N=59)	Control (N=25)	(Patient – Control)	(2-tailed)
Omissions_tscore	63.07 (SD 37.58)	52.62 (SD 17.75)	10.45	0.088 (NS)^
Commissions_tscore	52.44 (SD 11.23)	51.29 (SD 10.54)	1.16	0.654 (NS)^
HitRT_tscore	52.15 (SD 13.06)	45.38 (SD 10.54)	6.77	0.024 (SIG)^

Note: ^Independent Samples T-Test. **Abbreviations**: SD, Standard Deviation; SIG, significant (P<0.05); NS, Not significant.

We notice all our Bayesian Network models degenerate to Naïve Bayesian Network. This is because after feature selection (which selects features with lowest redundancy on highest prediction ability), all factors remaining for model construction contribute significantly and independently to the target pt_ctrl. They are conditionally independent to each other given the target node (pt_ctrl).

In summary, we have developed a data analysis and model construction approach and successfully applied it in building a set of schizophrenia classification models based on various neuroinformatics data. We will use this approach in modeling neuroimaging data in the next chapter.

Chapter 4

Neuroimaging-Based Analysis and Modeling

In this chapter, we will first describe the neuroimaging acquisition procedure and parameters. Then we will introduce our brain atlas based automatic ROI placement method for the MRI and DTI image analysis. Statistics of Fractional Anisotropy (FA) values within all selected brain structures are calculated for further analysis.

We will follow the procedure as we have used in Chapter 3, to select significant image features. These features will be used in the schizophrenia classification model construction.

4.1 MRI and DTI imaging

As we can see from the literature review in section 2.2 that schizophrenia is associated with not only the brain morphometric changes but also the white matter abnormalities, and MRI and DTI imaging are useful tools to quantify the neuroconnectivities.

All 156 study subjects (92 patients and 64 controls) have taken the MRI and DTI scans.

MRI Scan: Single session MRI scans are performed on a clinical 3-Tesla MRI scanning system (Intera 3T, Philips Medical Systems, Netherlands) with whole brain, high resolution, using 3D MP-RAGE (Magnetisation-Prepared Rapid Acquisition with a Gradient Echo) protocol. The volumetric scans parameters are TR/TE/TI (repetition time, echo time and, and inversion time) 8.4/3.8/3000; flip angle 8; matrix 256x204; Field of View (FOV) 240 mm^2, with axial orientation, covering the whole brain for structural-anatomic detail.

DTI Scan: In the same session, the diffusion tensor imaging (DTI) in 15 directions are also performed using single-shot, spin-echo EPI (echo planar imaging) sequence, b value (a diffusion weighted sequences factor, which summarizes the influence of the gradient) of 0 and 800 s/mm^2, at TR/TE (repetition time and echo time) 10,000/80, matrix 128x128, slice thickness 3 mm with no gap and field of view 240 mm^2 (K Sim, 2005).

A sample structure MRI image set is shown in Figure 4.1. Only the images in axial orientation are available. The coronal and sagittal images are reconstructed by the image processing software.

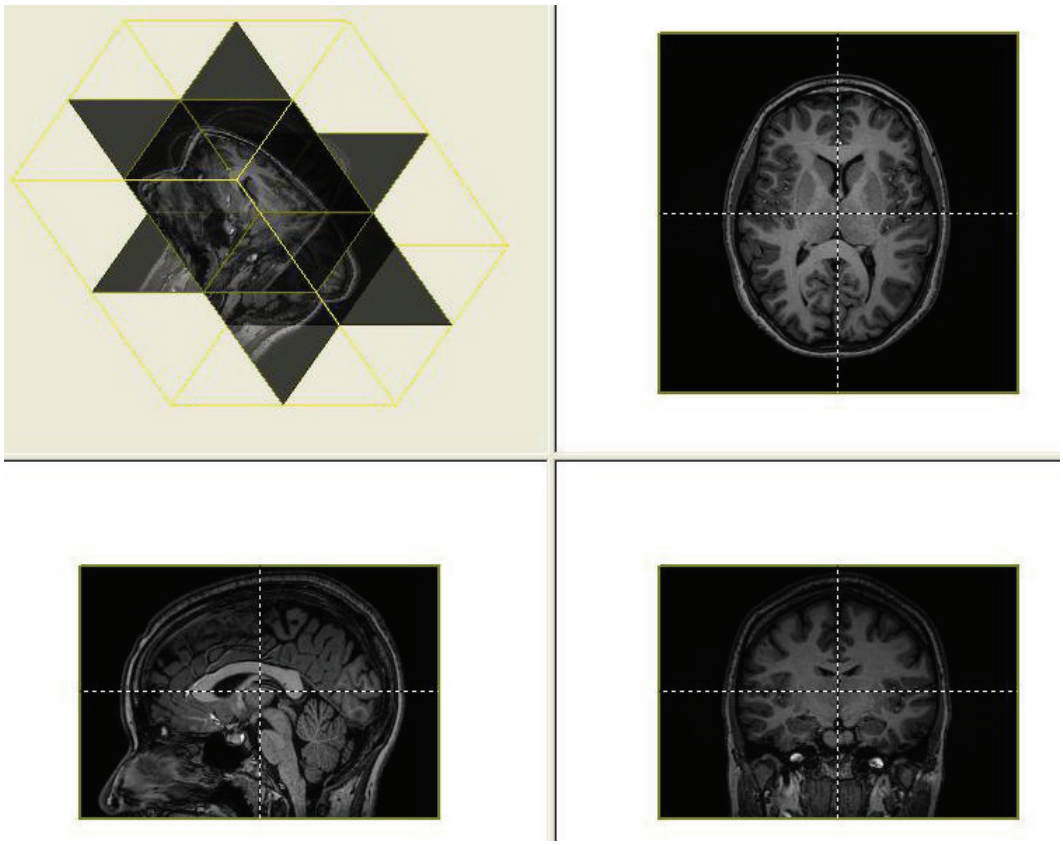

Figure 4.1 Structural MRI images
Upper right corner shows the original image scanned in axial orientation; lower right, coronal image; lower left, sagittal image; upper left, triplanar display

Figure 4.2 shows a set of Diffusion Weighted Images (DWI) in one of the 15 directions. DWI images of 15 directions will be used to construct the diffusion tensors.

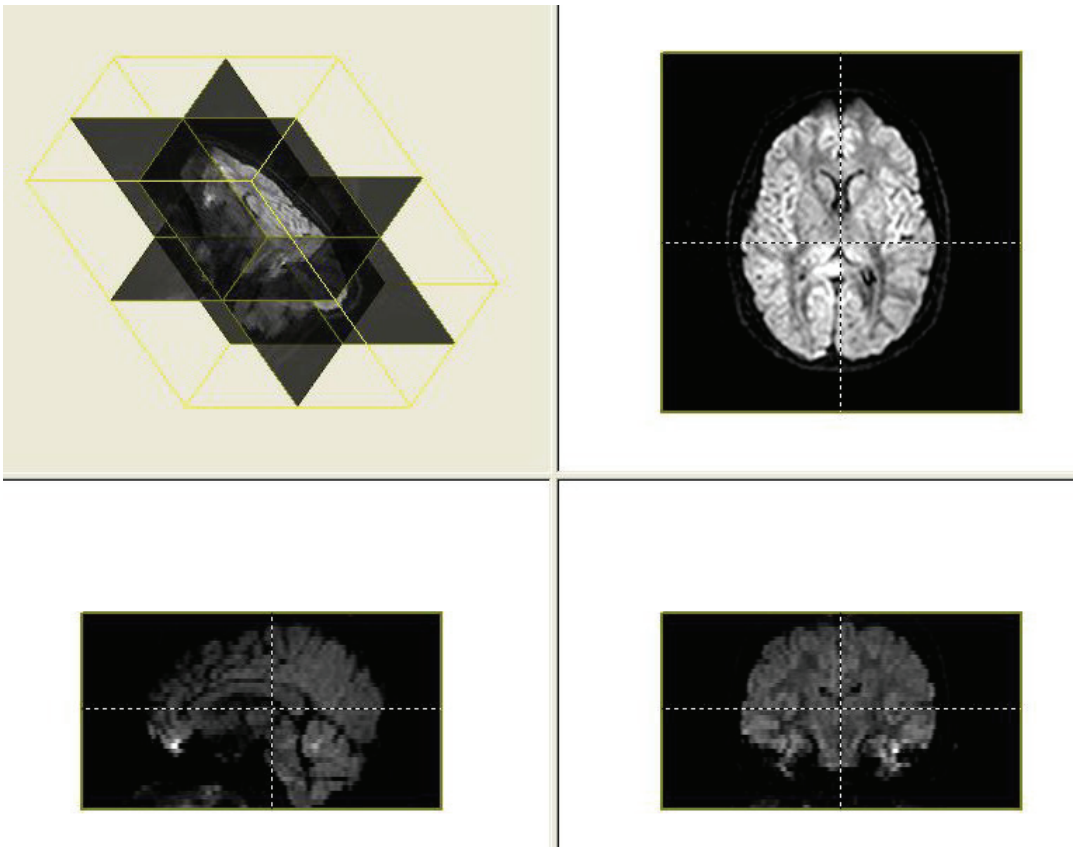

Figure 4.2 DWI images
Upper right corner shows the original image scanned in axial orientation; lower right, coronal image; lower left, sagittal image; upper left, triplanar display

4.2 Image Analysis Methods

Generally there are two different approaches in analysis of DTI images, the voxel based morphometry (VBM) (Ashburner & Friston, 2000) (Honea, Crow, Passingham, & Mackay, 2005) and region of interested (ROI) method (Giuliani, Calhoun, Pearlson, Francis, & Buchanan, 2005). VBM compares local concentration of brain images of two groups of patients voxel by voxel, while ROI method studies images within a specific ROI, usually selected manually.

In our previous study of computer aided diagnosis for acute stroke, image morphormetry based segmentation method was developed (G. L. Yang, et al., 2005)

to quantify brain CT images. However, it lacks the ability of automatically identifying multiple ROIs for FA quantification. Hence we propose a novel image analysis algorithm – brain atlas-based automatic ROI selection in DTI study, as illustrated in Figure 4.3.

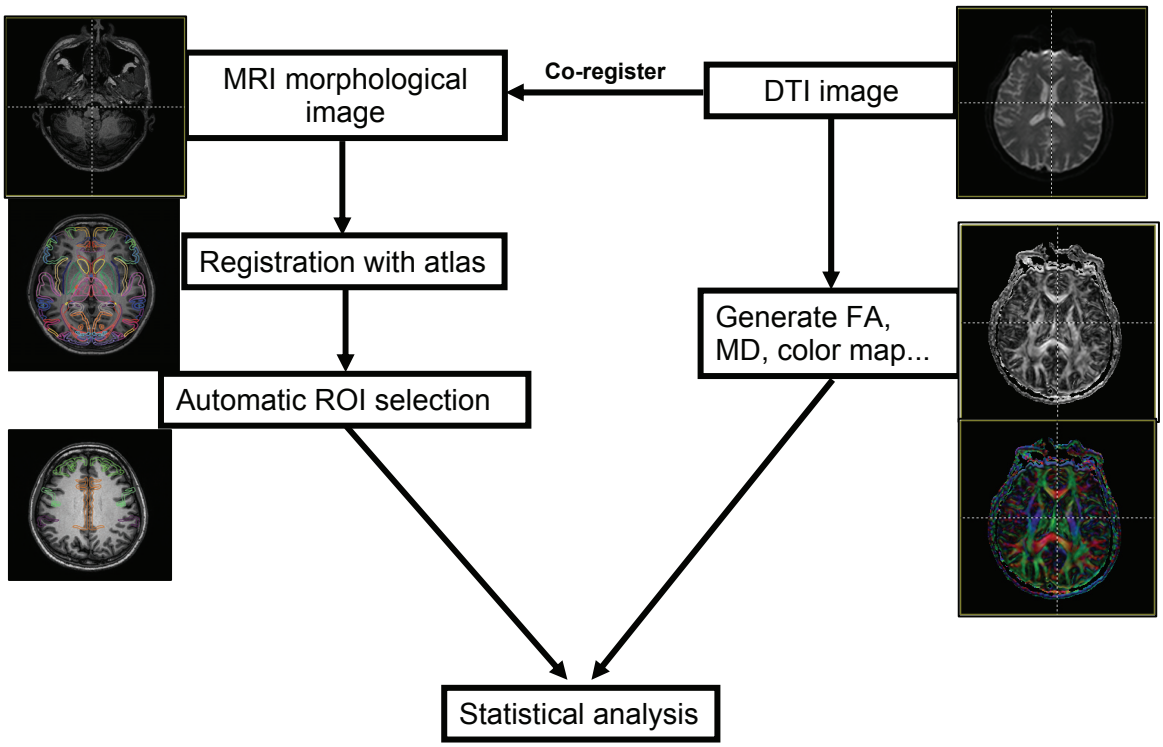

Figure 4.3 Image analysis algorithm

We describe our algorithm step by step as following:

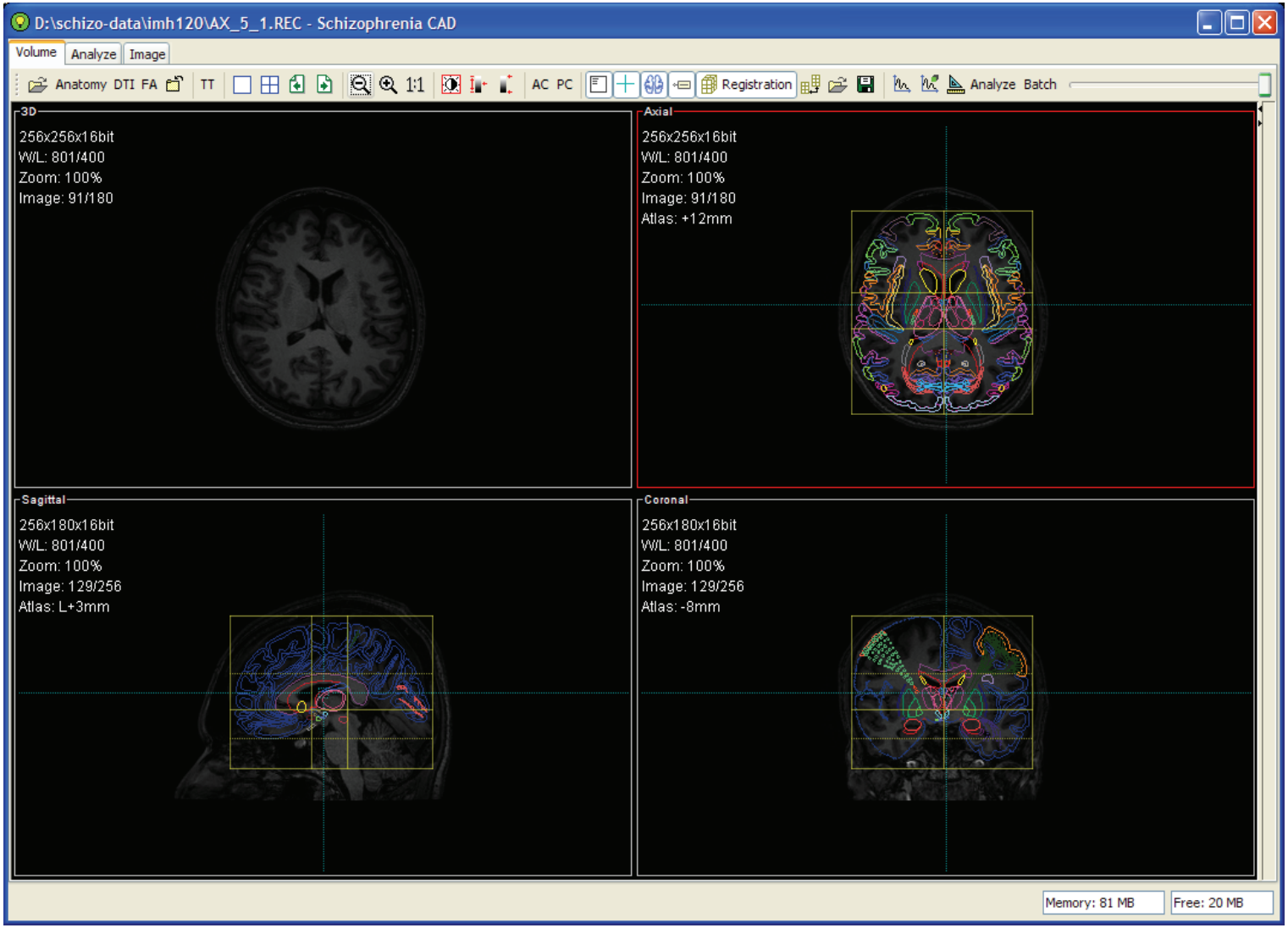

Figure 4.4 Step 1: Structural MRI images and the brain atlas are registered

Structural MRI images are displayed in axial (top right), coronal (bottom right), sagittal (bottom left), and triplanar (top left) views. Brain atlas is overlaid on the MRI images. Grids drawn in yellow color are the Talairach registration grid.

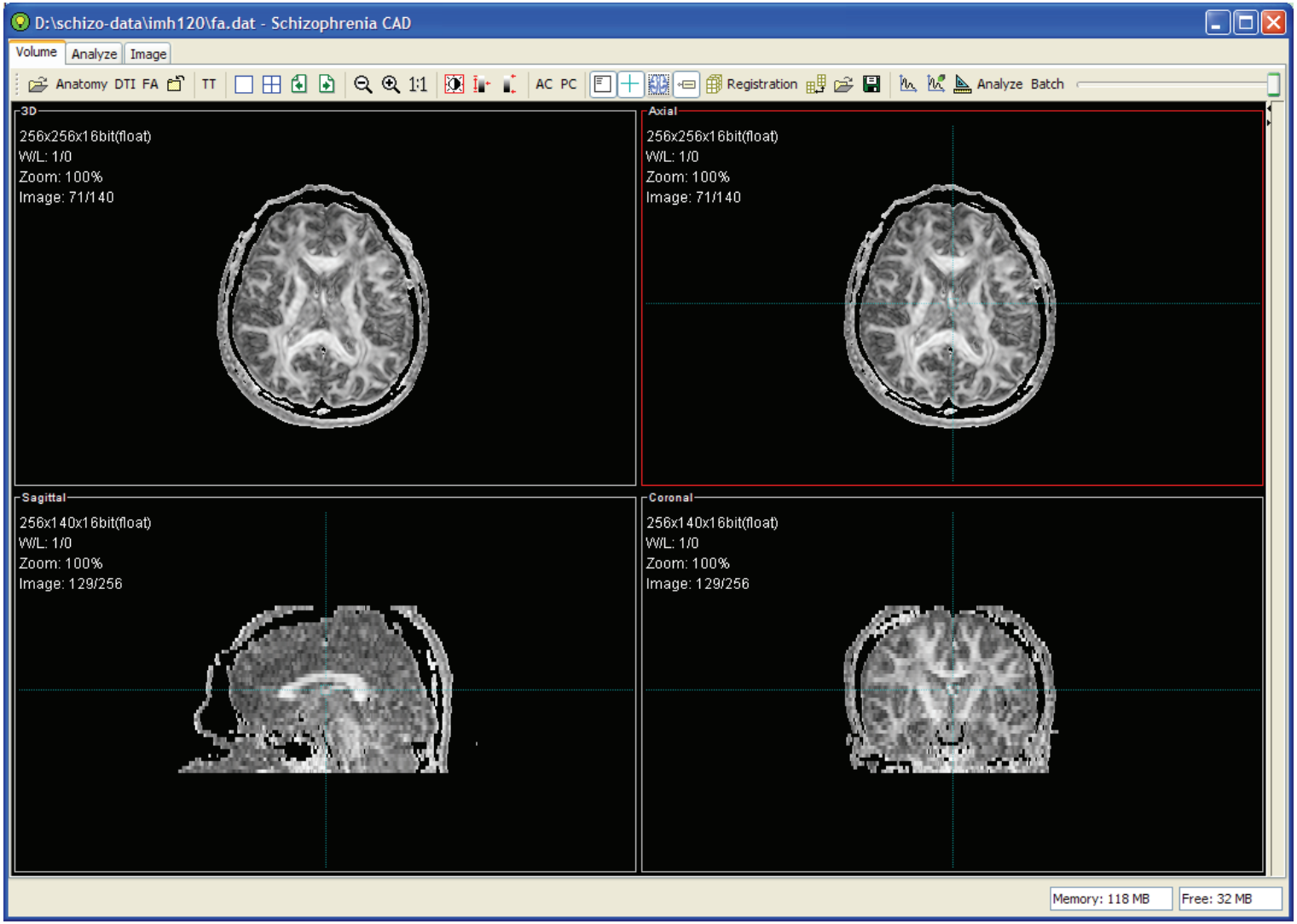

Figure 4.5 Step 2: Generating FA images

Images are displayed in triplanar view (top left), axial (top right), coronal (bottom riht) and sagittal (bottom left). Axial images are calculated from DWI images directly. Coronal and sagittal images are reformatted from axial images.

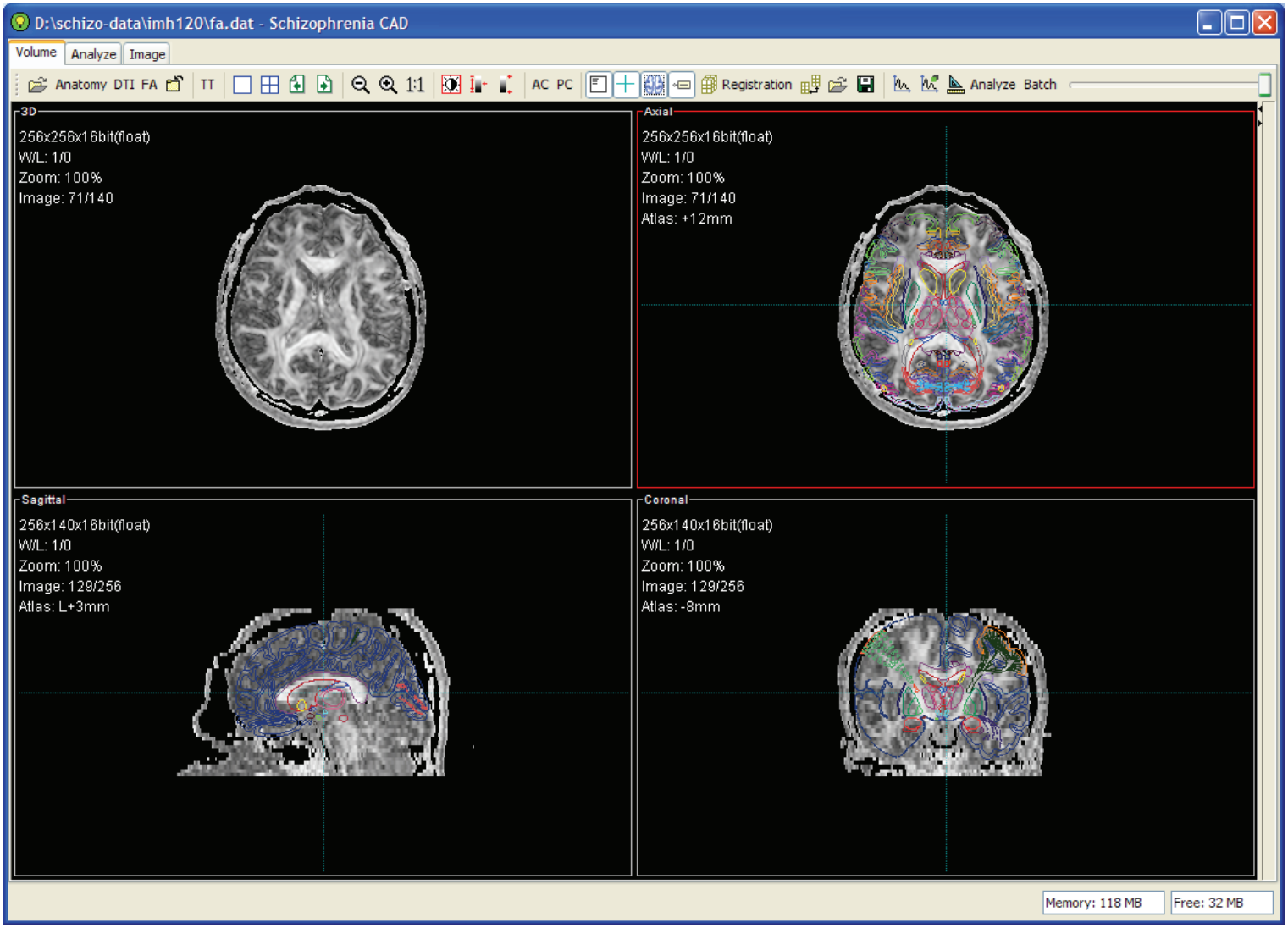

Figure 4.6 Step 3: FA images are registered with brain atlas

According to the co-registration information between structural MRI images and DTI images, brain atlas is automatically registered to FA images. FA images are delineated into brain structures. Top left: triplanar view; Top right: axial atlas overlaid on FA images in axial orientation; Bottom right: coronal atlas overlaid on FA images in coronal orientation; Bottom left: sagittal atlas overlaid on FA images in sagittal view.

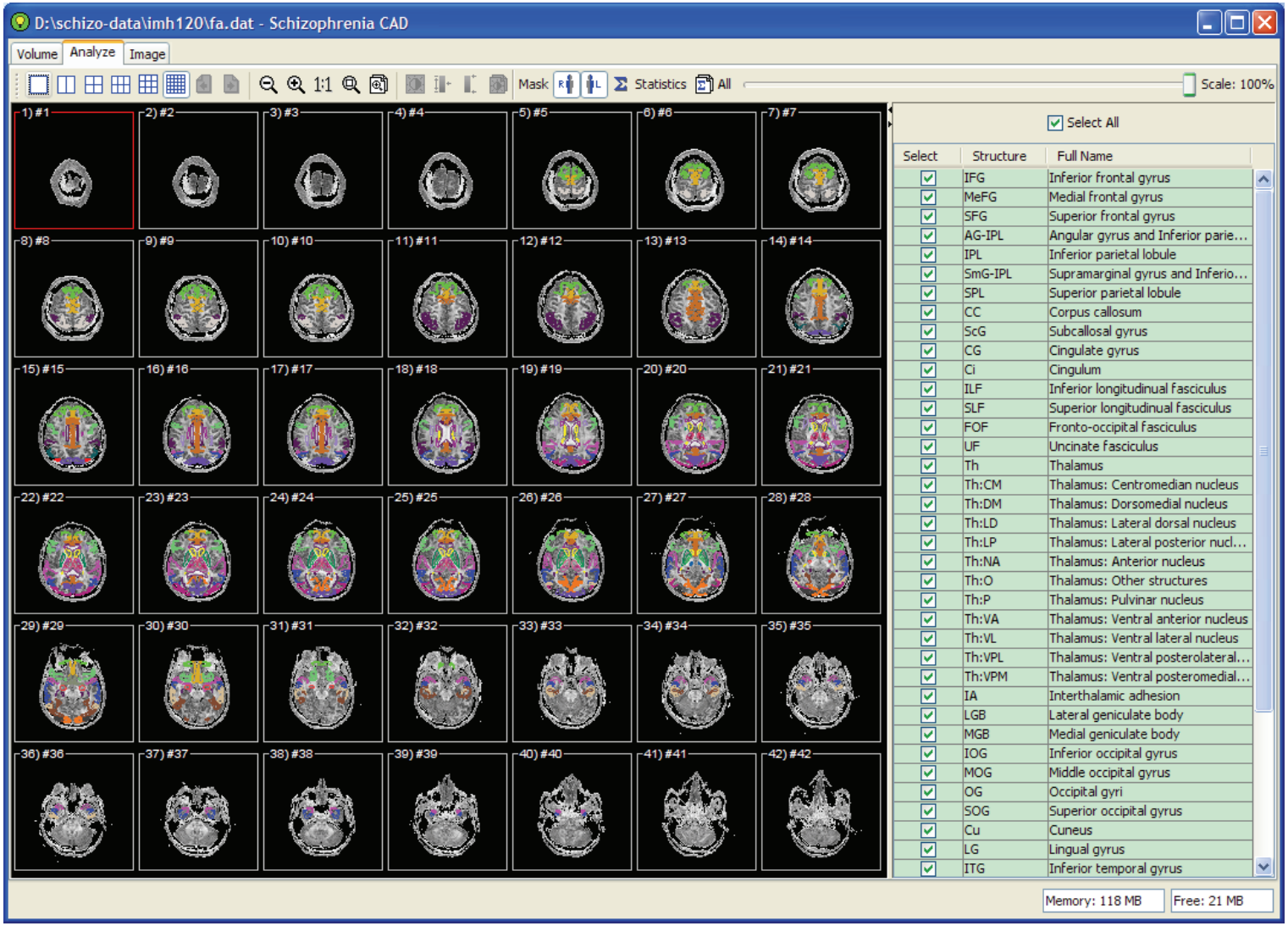

Figure 4.7 Step 4: FA images with selected brain structures

Brain structures are identified on all FA image slices. Statistics (voxels, mean, stdev) can be calculated within every ROI across all FA image slices delineated by brain atlas. Left: all FA images; Right: list of brain structures used as ROIs.

Step 1. Registering Brain Atlas to the structural MRI images

Structural MRI images contain the brain morphological information. The images are used to identify the anatomical structures, including anterior commissural (AC), posterior commissural (PC), and brain extends, which are useful for registration with the brain atlas. The structural images are in high resolution. In our current study, the pixel size is 0.9x0.9 mm, and the slice distance is also 0.9 mm, which makes cubic voxels.

Talairach-Tournoux brain atlas (W. L. Nowinski, 2005; W. L. Nowinski, et al., 1997; Talairach & Tournoux, 1988) is overlaid on top of the MRI structural images, as demonstrated in Figure 4.4. Brain atlas is registered by setting the original Talairach landmarks: anterior commissural (AC), posterior commissural (PC), and brain extends in all three directions, i.e. Right extend (R), Left extend (L), Anterior extend (A), Posterior extend (P), Superior extend (S) and Inferior extend (I), as well as the extended Talairach landmarks: Superior Midway (SM) and Inferior Midway (IM) (W. L. Nowinski & Prakash, 2005). The registration can be done automatically by using the Fast Talairach transformation algorithm (W. L. Nowinski, Qian, Bhanu Prakash, Hu, & Aziz, 2006), or by setting the landmarks manually.

Currently, the setting of Talairach landmarks is done by manual operation. The Fast Talairach Transformation (W. L. Nowinski, Qian, et al., 2006) can be used to accelerate the registration process and achieve more consistent registration results.

However, in our study, the time for registering brain atlas is not critical; we choose to place the landmarks manually in order to get a higher precision.

Step 2. Generating FA images

Diffusion Tensor Images (DTI) contains the diffusion tensor information, which can be used to calculate the Mean Diffusivity (MD), Fractional Anisotropy (FA), etc according to the method described in (Pierpaoli & Basser, 1996). Currently DTI Studio (Version 2.10, Johns Hopkins University, USA) (Jiang, van Zijl, Kim, Pearlson, & Mori, 2006) is used to generate the FA images. Figure 4.5 shows FA images displayed in axial, coronal, sagittal directions and the 3D view.

Step 3. Co-registering DTI/FA images and structural images

Structural images and the DTI images are scanned in the same session without changing the patient's position. Their geometric relation can be retrieved from the scanning parameters. This information is used to co-register the DTI images and the structural MRI images, because only rigid transformation is needed for the same subject's MRI and DTI images.

As structural images are already registered with the brain atlas in step 1, DTI images and FA images are also registered to the brain atlas. Figure 4.6 shows the registered brain atlas overlaid on the FA images.

Step 4. ROI selection and Statistical Analysis

Since the brain atlas delineated anatomical and functional regions of the brain, single or multiple region of interests (ROI) can be selected by just simply specifying their anatomical structure names. Figure 4.7 demonstrates FA images with some selected ROIs overlaid.

Statistics can then be calculated in each ROI. Currently we are interested in the volume (number of voxels), mean value, standard deviation of the FA images in the selected ROIs.

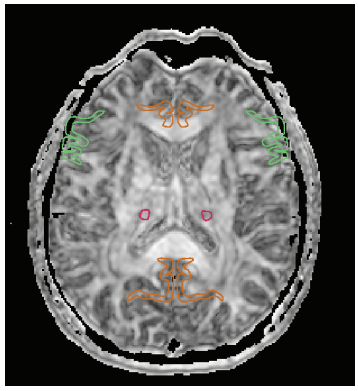

Figure 4.8 FA image with significant brain structures overlaid
Green: IFG; Brown: CG; Pink: ThLP. (Abbreviations: see Appendix B)

This image analysis algorithm has also been applied on a subset of the study subjects that consists of 36 patients (11 with passivity and 25 without passivity) and 32 age, gender and handedness matched controls. This sub-study identifies brain structure difference between schizophrenia with and without passivity. The results (K. Sim, et al., 2009) show that passivity is associated with the increased FA in right inferior frontal gyrus (IFG), cingulate gyrus (CG), left globus pallidus

(GP) and the decreased FA in left latero-posterior thalamic nuclei (ThLP) (Figure 4.8).

In this section, we introduced our image analysis algorithm. We will use this algorithm to quantify the FA images in the next section.

4.3 Quantification of FA Images

From literature review in Section 2.1.3, we notice that the Fractional Anisotropy changes of schizophrenia patients have been found almost all over the brain, from frontal, parietal, temporal, occipital regions, till deep brain structures such as corpus callosum, thalamic regions, and brain connections such as superior longitudinal fasciculus, inferior longitudinal fasciculus. We also notice that those finding are not consistent which might be due to the insufficient number of cases (mostly less than 50 to 60), and image processing methods – by using manual ROI (region of interest) placement, which may cause inconsistency in the regions identification studied. After discussion with a group of domain experts including neuroradiologist, neurologist, neuroscientist and psychiatrists, we select a wide spectrum of brain structure as potential relevant factors to schizophrenia in order not to miss potential findings. Our selected brain structures cover almost all brain regions found in the previous studies. Specifically, they include:

- Frontal (inferior frontal gyrus, medial frontal gyrus, middle frontal gyrus, and superior frontal gyrus)

- Parietal (angular gyrus and inferior parietal lobule, inferior parietal lobule, supramarginal gyrus and inferior parietal lobule, and superior parietal lobule)
- Corpus callosum
- Subcallosal gyrus
- Cingulate gyrus
- Cingulum
- Tracts (inferior longitudinal fasciculus, superior longitudinal fasciculus, fronto-occipital fasciculus, and uncinate fasciculus)
- Thalamus (all thalamic nucleus)
- Subthalamic nucleus
- Inter-thalamic adhesion
- Lateral geniculate body
- Medial geniculate body

According to above list, 48 brain structures from the brain atlas are chosen for our study. Each brain structure can be further subdivided into the portions in the left and right brain hemispheres. So each structure will generate 3 results: as a whole, as well as the left and right parts. Hence the total number of ROIs becomes 144 (48 x 3). The complete list of brain structures can be found in Table 4.1. Their full names can be found in Appendix B.

Table 4.1 Complete list of ROIs for the study

No	Structure	No	Structure	No	Structure
1	IFG	17	Th	33	MOG
2	MeFG	18	ThCM	34	OG
3	MiFG	19	ThDM	35	SOG
4	SFG	20	ThLD	36	Cu
5	AGIPL	21	ThLP	37	LG
6	IPL	22	ThNA	38	ITG
7	SmGIPL	23	ThO	39	MTG
8	SPL	24	ThP	40	STG
9	CC	25	ThVA	41	U
10	ScG	26	ThVL	42	CN
11	CG	27	ThVPL	43	GPL
12	Ci	28	ThVPM	44	GPM
13	ILF	29	IA	45	Pu
14	SLF	30	LGB	46	STN
15	FOF	31	MGB	47	FG
16	UF	32	IOG	48	AB

Abbreviations: see Appendix B.

FA images are registered to the brain atlas by using the method described in the previous section. With brain atlas, 48 ROIs are placed, and their statistics are calculated. Table 4.2 shows a part of the statistical results calculated in the selected brain structures. In each ROI, the number of voxels is proportional to its volume, since the size of each voxel is 0.9mm x 0.9mm x 3mm. The mean and stdev values are the mean and stdev FA values of all voxels within the specific ROI in all slices. These results will be used in the learning of schizophrenia model.

Table 4.2 Statistical results for the selected ROIs (partial)

CDNo	IFG			MeFG			MiFG		
	Voxel	mean	stdev	voxel	mean	stdev	voxel	mean	stdev
1	5804	0.2642	0.1773	5816	0.2507	0.1393	8918	0.2675	0.1597
2	7923	0.2192	0.1448	7837	0.2304	0.1320	12131	0.2153	0.1183
3	8330	0.2297	0.1566	8245	0.2743	0.1475	11960	0.2437	0.1693
4	9077	0.2429	0.1758	8420	0.2557	0.1337	13213	0.2458	0.1790
5	6854	0.2110	0.1511	6387	0.2224	0.1268	9781	0.2262	0.1417
6	9007	0.2220	0.1529	7813	0.2415	0.1529	12351	0.2146	0.1630
7	7574	0.2571	0.1604	7084	0.2351	0.1438	9956	0.2290	0.1469
8	8835	0.2220	0.1628	8586	0.2314	0.1370	12372	0.2125	0.1460
9	7981	0.2224	0.1713	7374	0.2232	0.1149	10791	0.2222	0.1614
10	7750	0.2407	0.1754	7621	0.2433	0.1428	9989	0.2312	0.1576

Abbreviations: CDNo: Compact Disk No (used as case number); IFG: Inferior frontal gyrus; MeFG: Medial frontal gyrus; MiFG: Middle frontal gyrus. **Note**: for each brain structure, voxel presents total number of voxels in all slices; mean is the mean FA values of all voxels in all slices; stdev is the standard deviation.

4.4 Model Construction

Our study involves 48 brain structures and 144 features (the mean of FA value in all the 144 ROIs). In order to build up a robust model, a subset of highly relevant candidate features must be chosen from the 144 features.

Feature Selection

We apply the method described in chapter 3 for feature selection. The following ROIs remain after feature selection:

- CG (cingulate gyrus)
- ScG_left (left subcallosal gyrus)

- ThLD_left (left thalamus: lateral dorsal nucleus)
- ThNA_right (right thalamus: anterior nucleus)

Table 4.3 Mean FA values of selected ROIs

Brain Structure	Mean		Mean Difference (Patient – Control)	P Value (2-tailed)
	Patient (N=59)	Control (N=25)		
CG	0.2547 (SD 0.0214)	0.2662 (SD 0.0133)	-0.0115	0.004 (SIG)^
ScG_left	0.2319 (SD 0.1032)	0.2840 (SD 0.0757)	-0.0521	0.013 (SIG)^
ThLD_left	0.2829 (SD 0.0750)	0.3497 (SD 0.0857)	-0.0668	0.002 (SIG)^
ThNA_right	0.3310 (SD 0.0522)	0.3595 (SD 0.0408)	-0.0285	0.009 (SIG)^

Note: ^Independent Samples T-Test. **Abbreviations**: SD, Standard Deviation; SIG, significant (P<0.05); NS, Not significant.

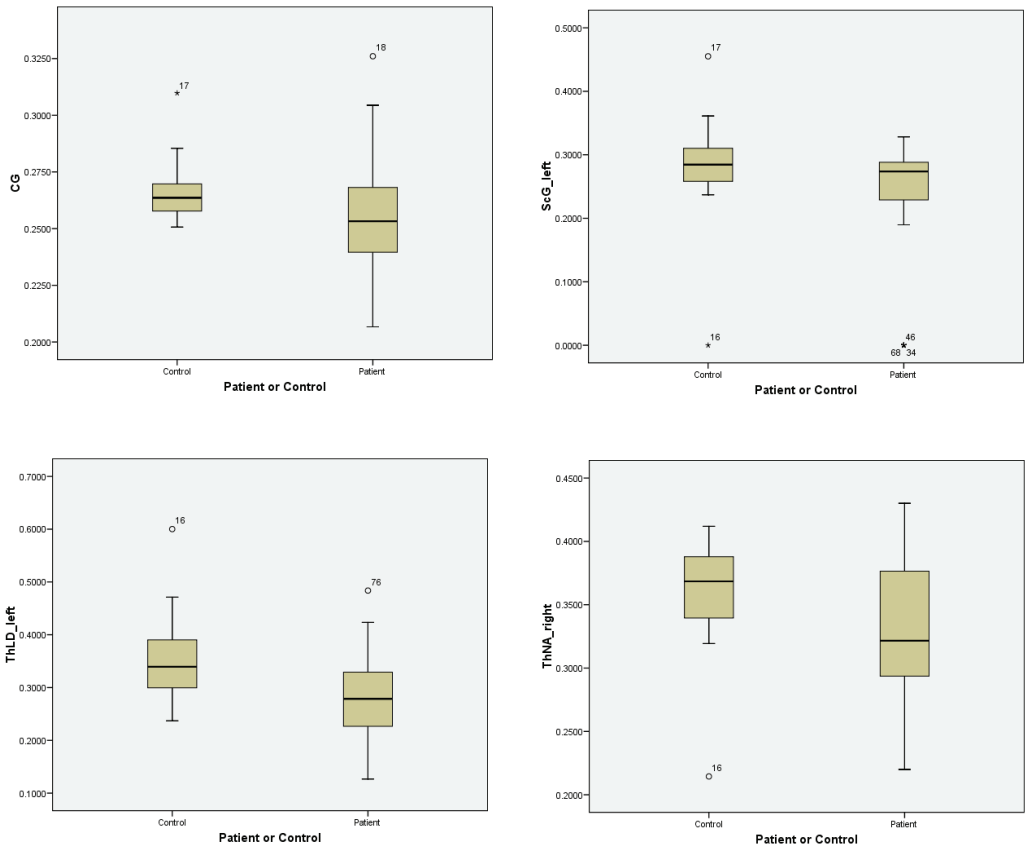

Figure 4.9 Box plot of FA values in selected image ROIs
FA values are groupped by patients and controls. Upper-left: CG; Upper-right: ScG_left; Lower-left: ThLD_left; Lower-right: ThNA_right. Note that scales for each box plot are different.

We compare the mean FA values using Independent Sample T Test (2-tailed) between patients and controls in the 4 selected ROIs, CG, ScG_left, ThLD_left and ThNA_right and display the results in Table 4.3, and show their box plots in Figure 4.9. We find that there is consistent decrease in FA across the four regions in patients compared to controls. That means in these brain structures, the brain connectivity is weaker in patients than in controls. Figure 4.10 illustrates the locations of the selected brain structures. Though ThLD_left and ThNA_right are tiny brain structures, they are still clearly visible and identifiable in the DTI images with the help of brain atlas. In our sample data, their mean volumes are 99.9 mm^3 (SD 35.5 mm^3) and 241.8 mm^3 (SD 47.1 mm^3), respectively.

From the anatomy point of view, the cingulate gyrus (CG) is important for focussed attentional tasks (Carter, et al., 2000; Sharp, Scott, Mehta, & Wise, 2006; Whittle, Allen, Lubman, & Yucel, 2006). Patient's mean FA value in cingulate gyrus, 0.2547 (SD 0.0214), is 0.0115 lower than that of controls, 0.2662 (SD 0.0133) (P=0.004). The reduced FA value in this region suggests that schizophrenic patients may have an anatomical basis for poorer attention capability than healthy controls. Besides some similar findings from (Kumra, et al., 2005) and (Hoptman, et al., 2008), recent functional magnetic resonance imaging (fMRI) study also shows schizophrenia patients having a reduction in blood oxygenated level dependent (BOLD) in cingulate gyrus compared to the healthy participants during attention processes (Filbey, Russell, Morris, Murray, & McDonald, 2008). A study also reported that schizophrenia patients and subjects of higher genetic risk for schizophrenia (with family history of schizophrenia) are found to have

reduced FA in cingulate gyrus, which suggests the cingulate gyrus as a good predictor even before the onset of psychotic sickness (Hoptman, et al., 2008).

The thalamus in the human brain is an important "way station" for many pathways and connections (Cipolotti, et al., 2008). The lateral dorsal nucleus (LD) in the thalamus is reported to have contributions to the links between hippocampus and thalamus (Aggleton & Brown, 1999). It connects to the parahippocampal and posterior cingulate cortex (Yeterian & Pandya, 1988) as well as medial temporal regions (including the hippocampus, presubiculum and entorhinal cortex) (Aggleton, Desimone, & Mishkin, 1986) (Saunders, Mishkin, & Aggleton, 2005), LD is involved in higher order somatosensory and visuo-spatial functions (Broman, 1994). Patient's mean FA value in LD thalamus, 0.2829 (SD 0.0750), is significantly lower than that of controls 0.3497 (SD 0.0857) (P=0.002). The reduced FA value suggests that deficits in visuo-perceptual tasks (Green, et al., 2009) or reported psychopathology such as passivity phenomenon may be related to disruptions in white matter integrity involving the LD thalamus.

On the other hand, the anterior nuclei (NA) thalamus has reciprocal connections with limbic regions subserving functions such as memory and emotional memory (A. Harding, Halliday, Caine, & Kril, 2000). It is reported that NA receives a key input from the hippocampus via the mamillothalamic tract (Saunders, et al., 2005). We find that patients' mean FA value in this region, 0.3310 (SD 0.0522), is 0.0285 lower than that of controls, 0.3595 (SD 0.0408) (P=0.009). The reduced FA value in this region points towards possible neural basis underlying memory deficits found not uncommonly in schizophrenia (Barch, Csernansky, Conturo, & Snyder,

2002; J. Hall, Harris, McKirdy, Johnstone, & Lawrie, 2007; Herbener, 2008; Lysaker, Bell, Greig, & Bryson, 2000).

Similarly, the subcallosal gyrus is part of corpus callosum and located immediately anterior to the anterior commissure of the brain. It is responsible for left and right prefrontal interhemispheric communication (Belin, Faure, & Mayer, 2008; Milner, 1982; Milner & Lines, 1982).Thus, the observation that patients' mean FA value in this region, 0.2319 (SD 0.1032), is 0. 0521 lower than that of controls 0.2840 (SD 0.0757) (P=0.013), is probably related to poorer connectivity between cerebral hemispheres. Disruption of white matter integrity involving this region may affect the information processing between the cerebral cortices and underlie information processing deficits implicated in the origin of symptoms such as delusions and hallucinations (Doty, 1989; Wright, et al., 1995).

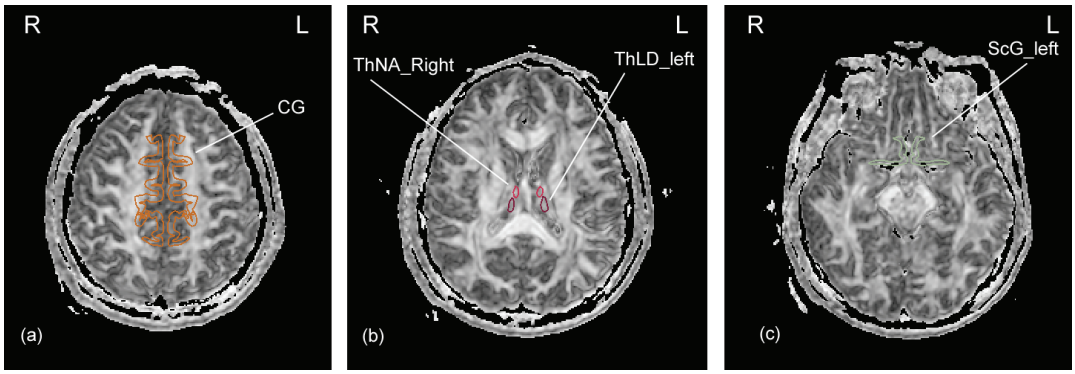

Figure 4.10 Selected brain structures
(a) CG (Cingulate Gyrus) (b) ThLD_left (Thalamus: Lateral dorsal nucleus, in left hemisphere) and ThNA_right (Thalamus: Anterior nucleus, in right) (c) ScG_left (Subcallosal gyrus, in left hemisphere). Note the images are displayed in the neuroradiology convention, i.e., the left side on the image shows the right side of the brain.

Model Construction

In order to combine with other neurocognitive tests, we use cases that complete all tests to construct our models. The total number of cases is 84, including 59 patients and 25 controls. Hence the prevalence probability of patient is 70.2% in our study sample space.

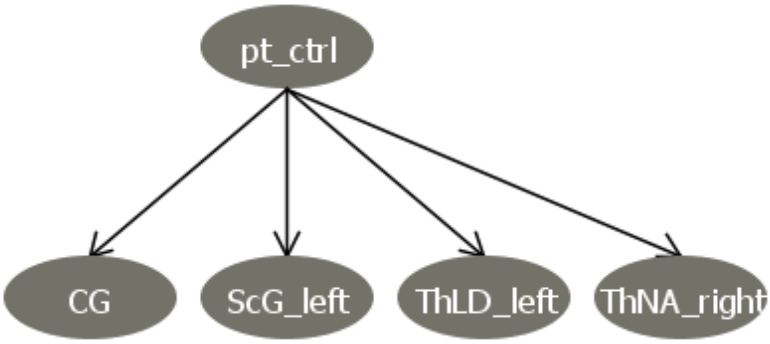

Figure 4.11 Bayesian network model on image features

Based on the 4 selected features, we construct a Bayesian Network classification model. This model (as illustrated in Figure 4.11) degenerates to a Naïve Bayesian network, since selected features are conditionally independent to each other on pt_ctrl target node.

Validation of the model is done by using the 10-fold cross-validation method. From the validation results (Table 4.4, Table 4.5, Table 4.6), we can see that 46 cases are correctly identified as patient, and 19 cases are correctly identified as control. The total number of correctly identified instances is 65. The accuracy is

77.4%, which is higher than the prior probability of patient. That shows image features are good predictors even without clinical data.

Table 4.4 Confusion matrix of model on image features

Test Outcome			
Patient (Positive)	**Control (Negative)**	**← Classified As**	
46 (TP)	13 (FN)	**Patient (59)**	**Ground Truth**
6 (FP)	19 (TN)	**Control (25)**	

Table 4.5 Detailed accuracy by class (image features)

TP Rate (Sensitivity)	FP Rate (Type I Error)	TN Rate (Specificity)	FN Rate (Type II Error)	Class
0.78	0.24	0.76	0.22	**Patient**
0.76	0.22	0.78	0.24	**Control**

Table 4.6 Summary of model (image features)

Item	Value
Total Number of Instances	84
Correctly Classified Instances	65
Incorrectly Classified Instances	19
Accuracy	**77.4%**
Error Rate	22.6%
True Positive Rate (Patient classified as patient)	78%
Type I Error Rate (Control classified as patient)	24%
True Negative Rate (Control classified as control)	76%
Type II Error Rate (Patient classified as control)	22%

For the identification of patient, the True Positive rate is 78%, and the True Negative rate is 76%. This shows imaging features have balanced abilities in identifying patients and controls. We will enhance this model by adding other features in the next chapter.

4.5 Conclusion

In order to determine the brain structure abnormalities, patients and controls are scanned to obtain their brain images in structural MRI and DTI formats. We developed an image analysis algorithm to automatically place 144 ROIs on the brain images. The ROIs are used to quantify the FA values.

Our image analysis algorithm has the following advantages over the conventional manual ROI placement:

- It introduces a systematic way for ROI selection. All ROIs are placed automatically after the brain atlas is registered to the patient's brain images. Human errors in placing ROIs are avoided. It also increases the consistency of image data quantification among multiple researchers and multiple centers.
- The results are more consistent among different studies, since the ROI placement is done by a computer program that implements our method.
- With the help of brain atlas, tiny structures such as subthalamic nucleus, which are usually difficult to identify from the image, can also be quantified.

- Diffusion Tensor Imaging (DTI) images can be warped to the Talairach space, image values can be compared among studies, and averaging of values in the Talairach space can be also performed.
- Large amount of ROIs and studies can be performed automatically, which is usually difficult for manual methods. This makes studies involving large amount of patients/controls more feasible.

We first apply this technique to examine the brain structure differences between schizophrenia patients with and without passivity (K. Sim, et al., 2009), and then use it in our study to extract 144 image features for all patients and controls.

Four significant features are chosen from the 144 features using a feature selection algorithm, namely, CG, ScG_left, ThLD_left and ThNA_right. Reduced FA values are found in the above 4 brain structures in schizophrenia patients compared to healthy controls. From the anatomy point of view, cingulate gyrus is important for attentional tasks, LD thalamus for higher order somatosensory and visuo-spatial functions and anterior nuclei for connections with limbic regions (memory, emotional memory etc) and subcallosal gyrus is part of corpus callosum (which plays a role in the left and right prefrontal interhemispheric communication).

We build a Bayesian Network classification model on these image features alone. A higher accuracy (77.4%) is achieved compared to the baseline model accuracy (70.2%) as described in Chapter 3. Based on this result, the image features appear to be promising factors in schizophrenia classification.

In summary, an image analysis algorithm is developed to extract features (FA values in brain structures) from the brain images. The image features are used as objective and quantifiable criteria in building the schizophrenia classification model.

In the next chapter, we will use the image features together with the neuroinformatics data to construct more comprehensive models.

Chapter 5

Neuroinformatics and Neuroimaging Data Based Modeling

In chapters 3 and 4, we created some schizophrenia models based on neuroinformatics and neuroimaging data separately. In this chapter, we will combine them, and build more comprehensive models.

5.1 Model Construction

We have already identified the 8 significant features (Table 5.1). They break down into: 1 feature from the clinical data, 3 features from different neurocognitive tests, and 4 from neuroimaging.

Various models have already been created based on different combinations of neuroinformatics features. Now, we add the neuroimaging features, and build more comprehensive models by using the same approach as discussed in chapters 3 and chapter 4. The complete list of models and their characteristics (including total number of correctly classified cases, total number of wrongly classified cases, model accuracy, sensitivity, specificity and Type I and Type II errors) are summarized in Table 5.2.

Table 5.1 Significant neuroinformatics and neuroimaging features

Category			Feature
Neuroinfomatics	Clinical Data		1) fam_hx
	Neurocognitive Tests	RPM	2) RPM_raw
		WAIS	3) DigitSpan_bwd
		WCST	4) PersResponses_raw
Neuroimaging			5) CG 6) ScG_left 7) ThLD_left 8) ThNA_right

The top half of the Table 5.2 contains 8 models, all of them are built on clinical information (fam_hx: family history of psychiatric disease) and neurocognitive tests results. Their accuracies range from 70.2% to 85.7%. The first model is the baseline model which contains only one factor (fam_hx), and it has the lowest accuracy among all models (70.2%). It also has the highest Type I error (100%), which means it is completely biased to patient - all cases are classified as patient. It is used as a starting point, and we gradually enhance it by adding more and more neurocognitive tests.

Table 5.2 Summary of models on neuroinformatics and neuroimaging

Model →		**C**	**C+R**	**C+WA**	**C+WC**	**C+R+ WA**	**C+R+ WC**	**C+WA +WC**	**C+R+ WA+ WC**
Nr		84	84	84	84	84	84	84	84
Cor		59	69	67	63	71	70	71	72
Incor		0	15	17	21	13	14	13	12
Acc		**70.2%**	**82.1%**	**79.8%**	**75.0%**	**84.5%**	**83.3%**	**84.5%**	**85.7%**
Err		29.8%	17.9%	20.2%	25.0%	15.5%	16.7%	15.5%	14.3%
Sen		100.0%	83.1%	72.9%	66.1%	88.1%	86.4%	81.4%	91.5%
Type I		100.0%	20.0%	4.0%	4.0%	24.0%	24.0%	8.0%	28.0%
Spe		0.0%	80.0%	96.0%	96.0%	76.0%	76.0%	92.0%	72.0%
Type II		0.0%	16.9%	27.1%	33.9%	11.9%	13.6%	18.6%	8.5%
Model →	**I**	**I+C**	**I+C+R**	**I+C+ WA**	**I+C+ WC**	**I+C+R +WA**	**I+C+R +WC**	**I+C+ WA+ WC**	**I+C+R +WA+ WC**
Nr	84	84	84	84	84	84	84	84	84
Cor	65	71	74	74	72	73	72	74	75
Incor	19	13	10	10	12	11	12	10	9
Acc	**77.4%**	**84.5%**	**88.1%**	**88.1%**	**85.7%**	**86.9%**	**85.7%**	**88.1%**	**89.3%**
Err	22.6%	15.5%	11.9%	11.9%	14.3%	13.1%	14.3%	11.9%	10.7%
Sen	78%	84.7%	98.3%	89.8%	88.1%	91.5%	93.2%	91.5%	93.2%
Type I	24%	16.0%	36.0%	16.0%	20.0%	24.0%	32.0%	20.0%	20.0%
Spe	76%	84.0%	64.0%	84.0%	80.0%	76.0%	68.0%	80.0%	80.0%
Type II	22%	15.3%	1.7%	10.2%	11.9%	8.5%	6.8%	8.5%	6.8%

Abbreviations: Nr, Total Number of Instances; Cor, Correctly Classified Instances; Incor, Incorrectly Classified Instances; Acc, Accuracy; Err, Error Rate; Sen, Sensitivity; TPR, True Positive Rate (Patient classified as patient); Type I, Type I Error Rate (Control classified as patient); Spe, Specificity, TNR, True Negative Rate (Control classified as control); Type II, Type II Error Rate (Patient classified as control); C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test; I, Imaging

The bottom half of the table contains 9 models. The first one (model I) is built on imaging features only (as discussed in Chapter 4). The rest of the 8 models are built on clinical features and various combinations of neurocognitive tests results

plus the neuroimaging features. There is a one to one relationship between the top half of the table and the bottom half, except for model I, namely, a column in the bottom half is the result of a column in the top half plus the imaging features.

We can see that, starting from model I+C, till model I+C+R+WA+WC, all models with imaging features plus some other features can achieve accuracy of more than about 85%. This is a big improvement compared to the baseline model (70.2%). All sensitivities are improved too. The lowest sensitivity (84.7%) is observed at model I+C, which is still reasonably good; the rest are close to or higher than 90%. This makes these models good in predicting schizophrenia. On the other hand, specificities range from 64% to 84%, which shows a generally lower ability in detecting controls than patients.

Among all models, the most comprehensive one is the model with all the 8 significant features, fam_hx, RPM_raw, DigitSpan_bwd, PersResponses_raw, CG, ScG_left, ThLD_left, and ThNA_right, which has the highest accuracy of 89.3%. This model is illustrated in Figure 5.1. As we can see, it has the form of Naive Bayesian Network with one target node (pt_ctrl) and 8 child nodes, each representing a selected feature. All features are independent to each other.

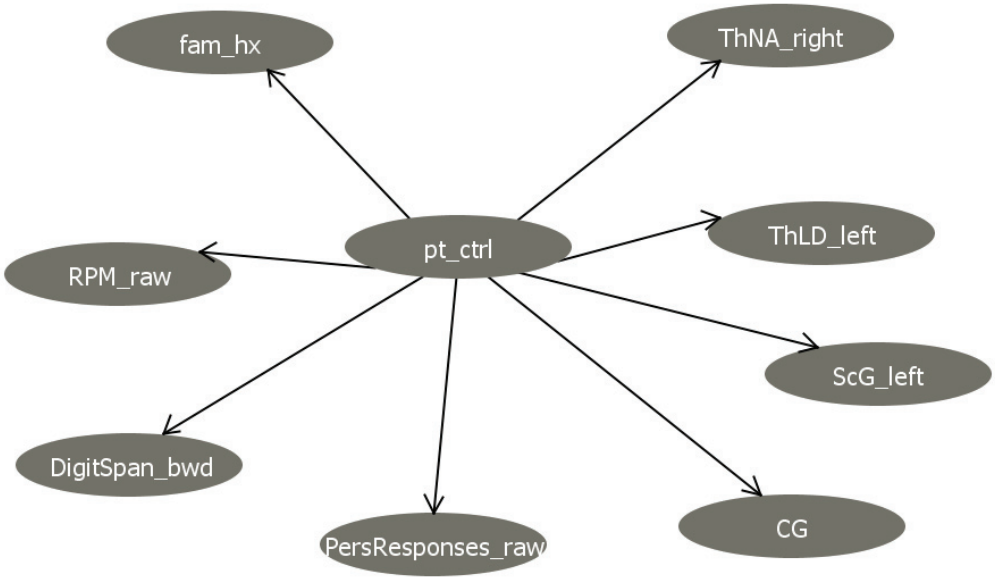

Figure 5.1 The Most comprehensive model on all information

5.2 Results and Conclusions

We have constructed 17 models on different combinations of input features. Since clinical information is considered to be always available (our assumption), we use it as the baseline model for comparing model accuracies.

Figure 5.2 shows the accuracies of all models. Most models can achieve accuracy from about 80% to 90%, which is about 10%-20% gain compared to the baseline model (model with family history only). We also notice that in general, the accuracy has an increasing trend by adding more and more tests.

The model on image features alone is a special case. Its accuracy is 77.4%. It is used to demonstrate the usefulness of neuroimaging in schizophrenia classification.

When combined with clinical information, the accuracy increased to 84.5%, a gain of 14.3% compared to the baseline model.

The other models have accuracy of less than 80%. They are models on clinical data plus a single neurocognitive test (WAIS or WCST). This shows that single neurocognitive test plus clinical information (without imaging) are not sufficient to get good diagnosis accuracy.

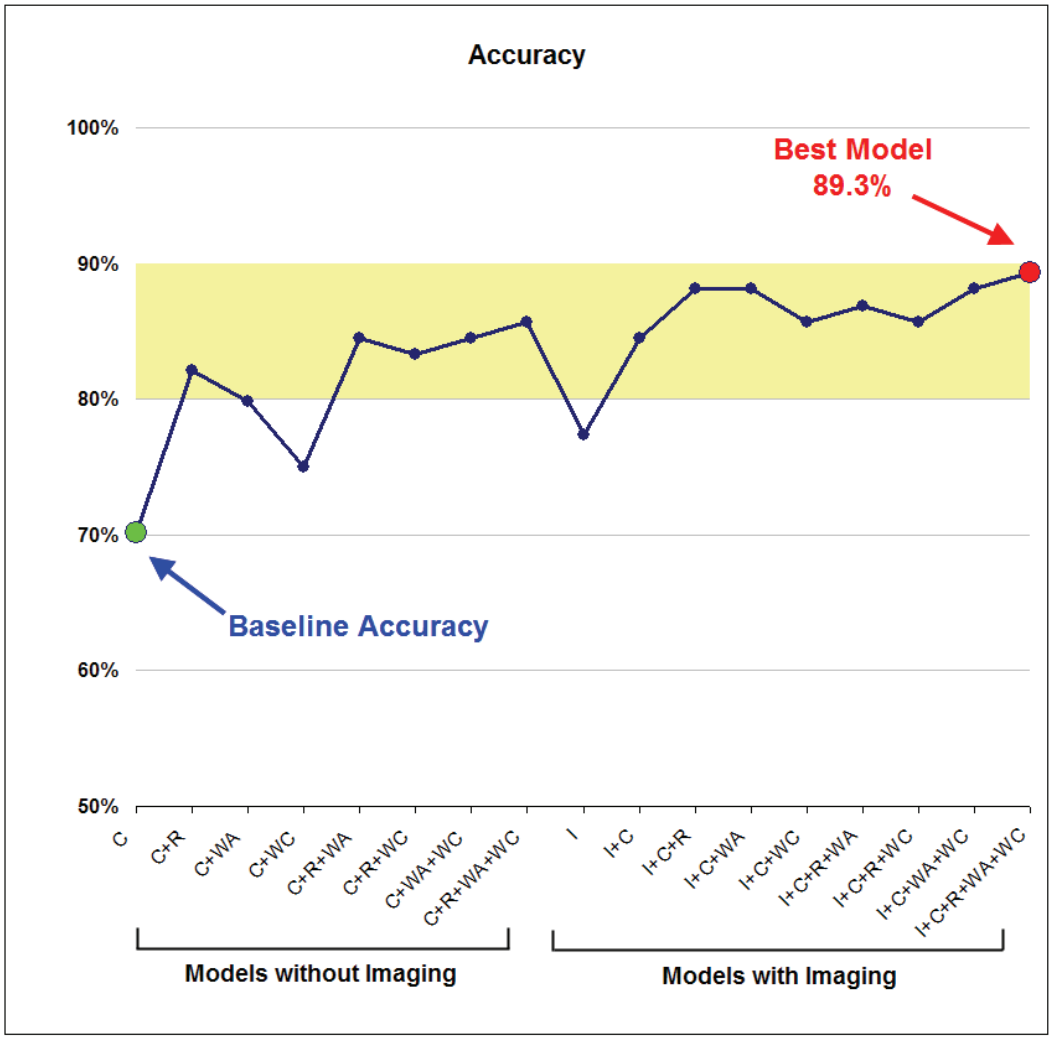

Figure 5.2 Accuracy chart of all models

Accuracy of the baseline model is 70.2%. The most comprehensive model with all features has the highest accuracy of 89.3%. Most other models have accuracy of about 80-90%. Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Type I Error (Figure 5.3) of baseline model (clinical data only) is 100%, meaning all cases are classified as patient. That restricts its usefulness in practice. Other models have Type I error below 30%, except for model I+C+R, and model I+C+R+WC, whose type I error reaches 36% and 32% respectively. Almost all Type II errors are below 30%, with an exception of model C+WC, whose Type II error reaches 33.9%.

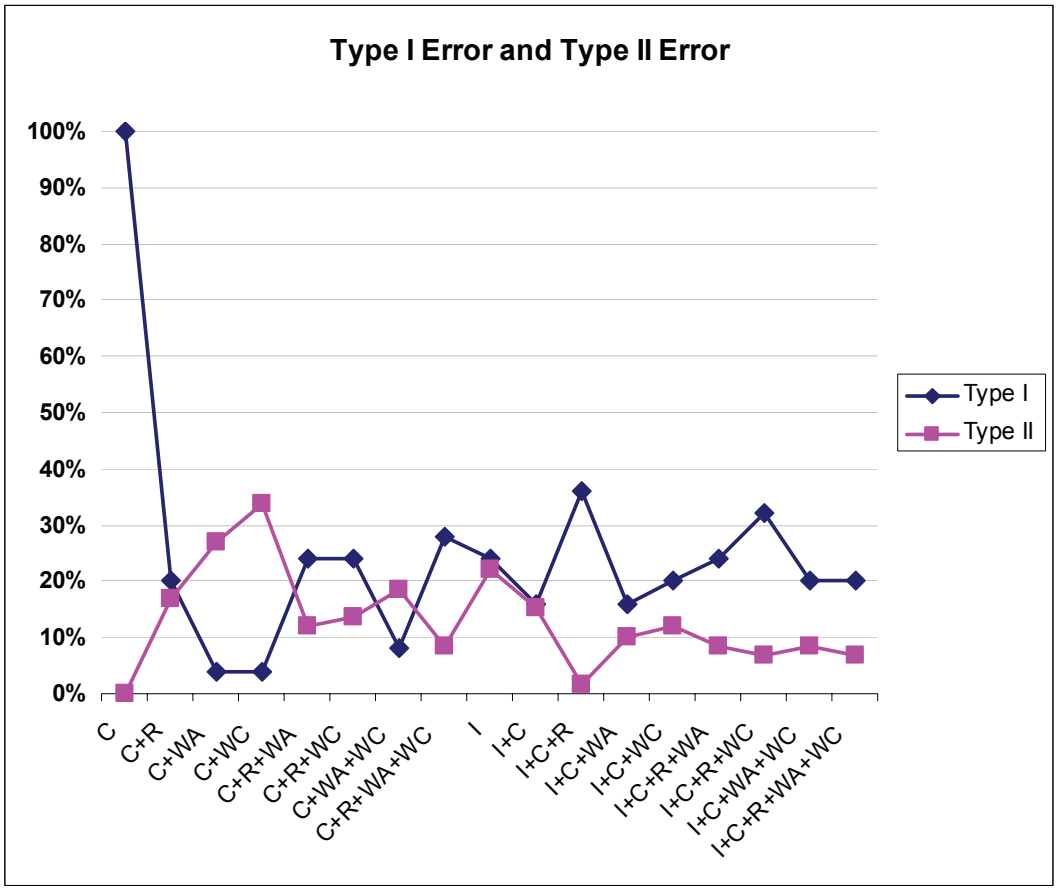

Figure 5.3 Type I and II error chart of all models
Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Effect of Neuroimaging

We examine the effect of neuroimaging, by comparing model accuracy with and without neuroimaging features (Figure 5.4). We notice a substantial increment in

accuracy (from 6.0% to 14.3%) is obtained when adding neuroimaging feature into any models with none or single neurocognitive test only. However, if a model already contains multiple neurocognitive tests, the accuracy gain is only marginal (2.4% to 3.6%).

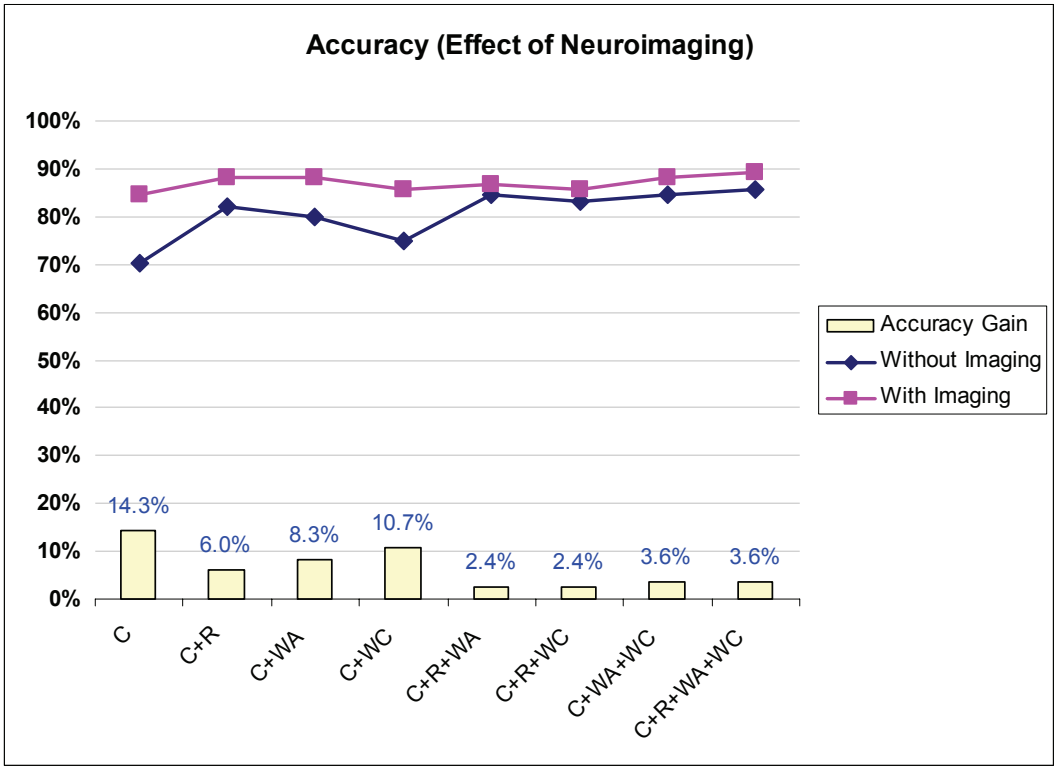

Figure 5.4 Accuracy (effect of neuroimaging)
Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Effect of RPM Test

We also examine the effect of RPM test by comparing the accuracy of models with and without RPM (Figure 5.5). Adding RPM test to existing models that contain none or only one neurocognitive test (mode C, C+WA, C+WC), the accuracy gain is substantial (4.7% to 11.9%). This is similar to what we have observed in neuroimaging effect.

On the other hand, if existing model already contains two other neurocognitive tests (WA+WC), the accuracy gain is merely 1.2%.

For models that already contain neuroimaging features, adding RPM doesn't increase their accuracy substantially. For instance, adding RPM to model I+C+WA doesn't increase the accuracy at all. For model I+C+WA, adding RPM test even decreases its accuracy marginally by 1.2%. Since the validation is done on 84 cases, 1.2% decreasing means just 1 case difference (1/84 = 1.2%). This fluctuation may be caused by the small number of validation cases. In this situation, RPM test's additional contribution to the classification accuracy is already small; when the number of cases is small, the irregularity of the sample data may affect a small number (e.g., 1 or 2 cases) of classification results, and cause the decreased accuracy. In other words, the model generated from the training cases represents a classification rule for the objective of optimal probability of Bayesian Network structure given the current training dataset. However the test subset of the sample data may not always follow the distribution pattern that the optimal model required, hence the model accuracy can be reduced by a small amount sometimes.

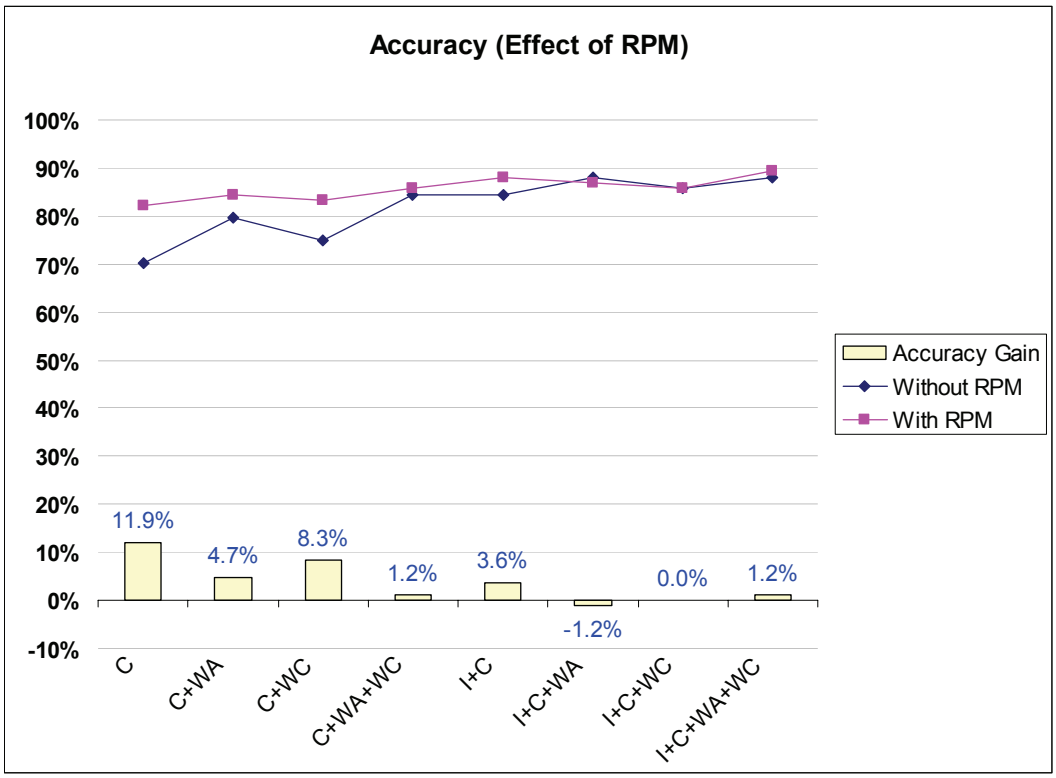

Figure 5.5 Accuracy (effect of RPM test)
Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Effect of WAIS Test

Similar accuracy change effect can be observed in the case of WAIS test (Figure 5.6). Adding WAIS test into models with none or one neurocognitive test (WCST) does boost the accuracy substantially by 9.6% and 9.5% respectively. For the rest of the models, it doesn't contribute much to models accuracy: accuracy gains are 2.4%-3.6%.

In the worst situation (for model I+C+R), the accuracy even drops 1.2%. Since the validation is done on 84 cases, 1.2% decreasing means just 1 case difference. This fluctuation may be caused by the small number of validation cases as discussed earlier.

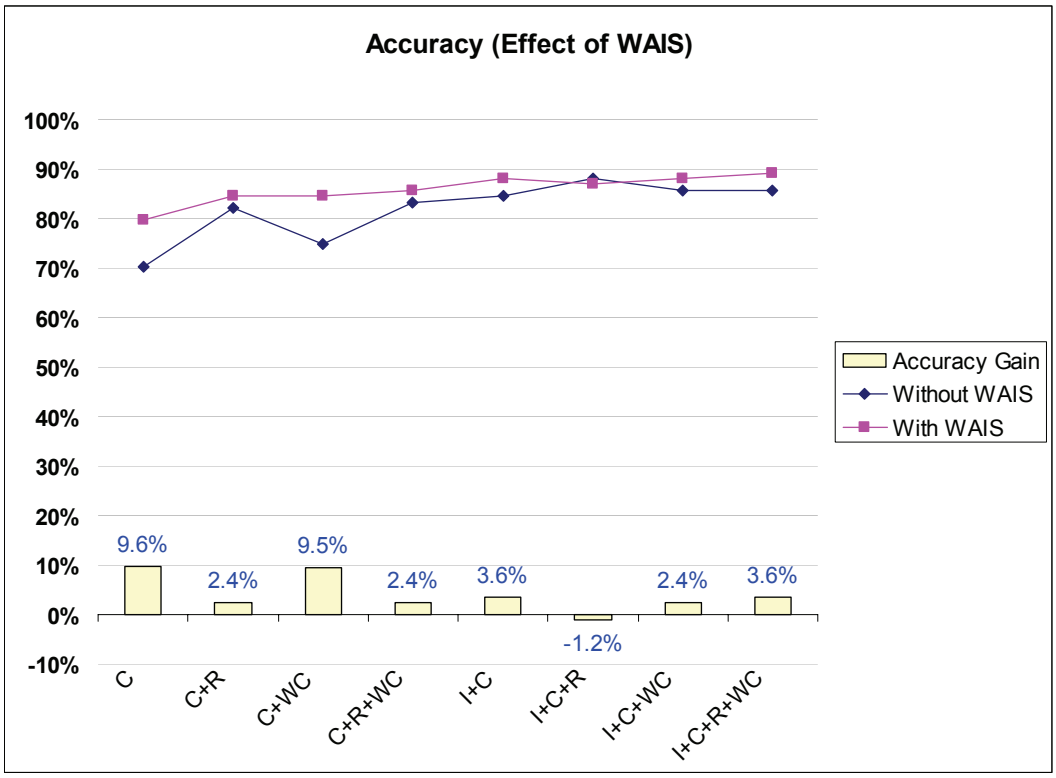

Figure 5.6 Accuracy (effect of WAIS test)
Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

Effect of WCST

The contribution of WCST is the smallest among all 3 other tests (RPM, WAIS, neuroimaging) (Figure 5.7). It only increases the accuracy by 4.8% and 4.7% for model C and model C+WA, respectively.

For the rest of the models, WCST doesn't contribute much, or even decreases the model accuracy. Again, this fluctuation may be caused by the small number of validation cases as discussed earlier.

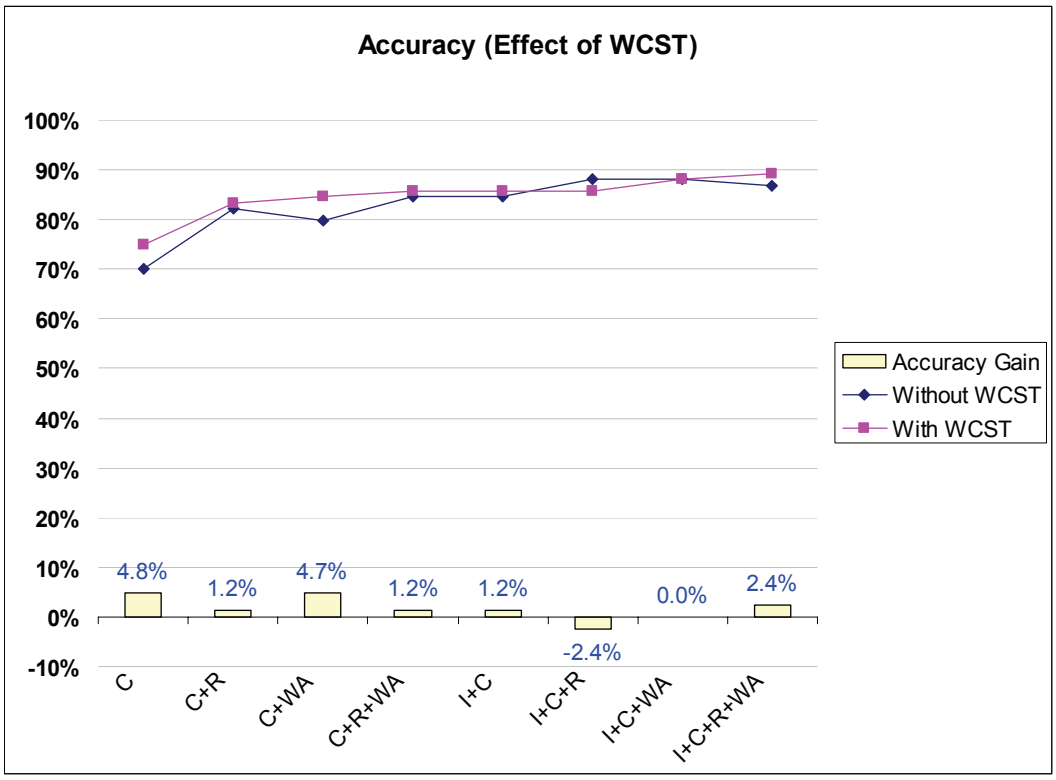

Figure 5.7 Accuracy (effect of WCST test)
Abbreviations: I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

In summary, combined schizophrenia models are constructed based on neuroinformatics and neuroimaging data. Models with two or more neurocognitive tests or the neuroimaging data can achieve classification accuracy at about 80%-90%, an increase of 10%-20% compared to the baseline model (with clinical information only).

In the next chapter, we will use the models that we have already built to develop a decision support system for schizophrenia diagnosis.

Chapter 6

Decision Support System for Schizophrenia

The models we have built in previous chapters are based on clinical information (fam_hx), neurocognitive tests (RPM, WAIS, WCST), and neuroimaging.

All the features selected for model construction are objective. They are also quantifiable (their values are integer numbers or real numbers), except for fam_hx (which takes three possible nominal values: Nil, 1st degree and 2nd degree). Hence they are more reliable than the subjective criteria used in DSM-IV and ICD-10.

In this chapter, we will develop a schizophrenia diagnosis decision support system based on the features we selected and various models we constructed.

6.1 Decision Support System

In order to augment the exiting standard diagnosis and provide a diagnosis result based on objective criteria that they are lacking of, we decide to construct a decision support system by combining and using all schizophrenia models in one integrated system.

In previous chapters, we have built some 17 models (see Table 5.2). Different model can be applied at different situation depending on availability of the test results. For example, when patient clinical information is only available, model C can be applied. After the person does the WAIS test, model C+WA can be applied to generate the classification results.

All models data are stored in Bayesian Interchange Format (BIF) (Cozman, 1998) files. A BIF file is a text file using XML schema, which stores all the nodes information, including their names, possible nominal values, and ranges of numeric values. It also stores the relationship (arcs) between nodes, and the conditional probability tables for all nodes.

We develop computer software to provide the decision support in schizophrenia diagnosis based on all models. The decision support system consists of 4 components: a model repository, a data input GUI, a decision support engine, and a report display GUI. A block diagram of the system is shown in Figure 6.1.

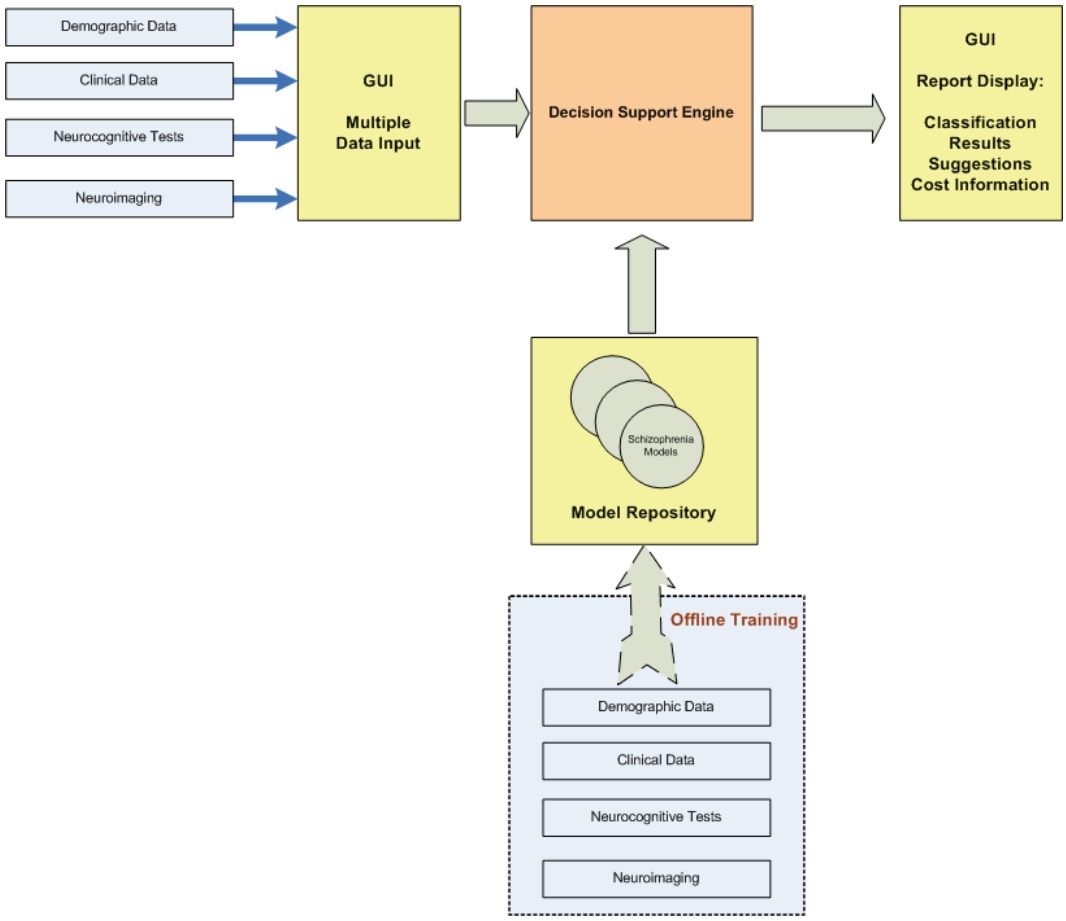

Figure 6.1 Decision support system block diagram

The Model Repository: This component stores all Bayesian Network models, and their properties, as well as some other information like the cost for the tests. We have collected the cost information for various neurocognitive tests in previous chapter; we compile them together with the cost information for the neuroimaging (Table 6.1). The cost information will be used by the decision support engine to calculate the cost effectiveness of models. The model repository provides accessing functions to models and their properties.

Table 6.1 Cost of tests

Test	Time Needed	Cost	
		Subsidized Rate	Private Rate
RPM	30 min – 1 hour	$130*	$400*
WAIS-III	15 min – 30 min	$65	$240
CPT-II	15 min – 30 min	$65^	$200^
WCST	15 min - 20 min	$65^	$200^
Neuroimaging (MRI+DTI)	30 min – 1 hour	$220	$457

Note: * Estimated; ^Estimated by dividing total cost of CPT and WCST by 2.

The Data Input Component: It is a simple Graphical User Interface (GUI) that allows the user to input patient's clinical information, various neurocognitive tests results, and neuroimaging results.

The Decision Support Engine: This component receives the user input data, and automatically chooses an appropriate Bayesian network model from the Model Repository, depending on the availability of different types of data. (For example, if the RPM test and WAIS test results are available, the model C+R+WA will be chosen.) Then the model is queried to generate the probability distribution of the input case by using Equation (6-1) (because all our models are in the form of Naïve Bayesian Network). The classification result is calculated using Equation (6-2). In addition, the decision support engine also searches in the model repository for other models that may generate higher accuracy.

$$P_{dist}(v) = P(pt\_ctrl = v)\prod_{i=1}^{n} P(F_i = u_i \mid pt\_ctrl = v) \qquad \textbf{(6.1)}$$

where P_{dist} is the distribution probability of patient or control; v takes value of "Patient" or "Control"; F is a chance node (factor) in the Naïve Bayesian network (other than the target); u is the possible value of a chance node; n is the number of other chance nodes.

$$pt\_ctrl_{classify} = \arg\max_{v} P_{dist}(v) \qquad \textbf{(6.2)}$$

where $pt\_ctrl_{classify}$ is the classification result of a case: either "Patient" or "Control"; P_{dist} is the distribution probability of patient or control; v takes value of "Patient" or "Control".

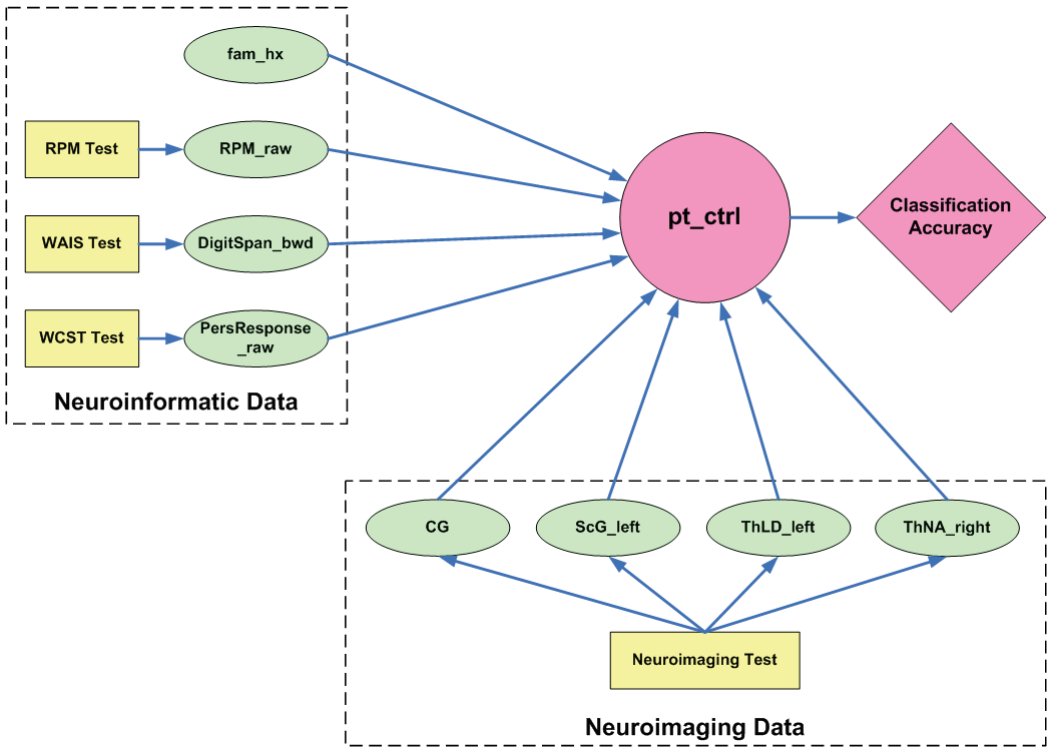

Figure 6.2 Component diagram of decision support system

Figure 6.2 illustrates the component diagram for the decision support engine. The target node is the classification result, pt_ctrl, which is related to two types of data, neuroinformatics data and neuroimaging data. More specifically, chance node

pt_ctrl is related to chance node fam_hx, RPM_raw, DigitSpan_bwd and PersResponse_raw (in the neuroinformatics domain), and CG, ScG_left, ThLD_left, ThNA_right (in the neuroimaging domain). There are four decision nodes (RPM Test, WAIS Test, WCST Test and Neuroimaging Test). Each generates it respective test results (chance nodes). The value node represents the classification accuracy of the specific Bayesian Network model used in the classification for a given combination of input data. For example, when only RPM test is done, model C+R will be used for classification, and hence the value node, classification accuracy takes the value of accuracy of model C+R.

We describe the algorithm for searching suggested models (with further tests) that have higher accuracy than any given model as follows:

Model searching algorithm (pseudo code)

```
// ---------------------------------------------
// Find suggested models to improve the accuracy
// Input parameter: M0: the given model
// Output: a list of suggested models
// ---------------------------------------------
create an empty list for the suggested models
FOR each model M in the model repository
       check the tests contained in existing model M0
       check the tests contained in existing model M
       compare the model M with existing model M0
       IF M contains one more test than M0 THEN
              compare their classification accuracies
              IF accuracy of M is greater than that of M0 THEN
                     add M into the suggested model list
              END IF
       END IF
END FOR
RETURN the suggested model list
```

The Report Display Component: This is a simple Graphical User Interface (GUI) to display the report generated by the decision support engine. It also provides a save function allowing the user to save the report in a text file, and a print function allowing the user to print a hard copy report.

6.2 Results

6.2.1 Decision Support Flow Charts

Based on all schizophrenia classification models, their accuracies, their tests included, and relationships between them, we make a decision support flow chart (Figure 6.3) that can help clinicians to decide what test to choose in order to increase the diagnosis accuracy in various situations.

In the flow chart, a circle represents a model; the size of the circle represents its accuracy: the larger the size, the higher the accuracy. An arrow represents a test with different colors representing different tests: brown for RPM, green for WAIS, blue for WCST and pink for Imaging.

Models are arranged in 5 different layers. The “Baseline” layer contains model C, which does not include any test. The “1 Test” layer contains models with any single test, namely models C+R, C+WA, C+WC and C+I. The “2 Tests” layer contains models with any combinations of two tests, for example C+R+WA. The “3 Tests” layer contains models with any combinations of 3 tests such as model C+R+WA+WC. Finally the “4 Tests” layer contains the most comprehensive model I+C+R+WA+WC (which has all 4 tests: 3 neurocognitive tests and the neuroimaging test).

In each layer, models are sorted by accuracy from left to right. For example, at "1 Tests" layer, models C+WC (75%), C+WA (79.8%), C+R (82.1%) and I+C (84.5%) are arranged from the lowest accuracy to the highest accuracy.

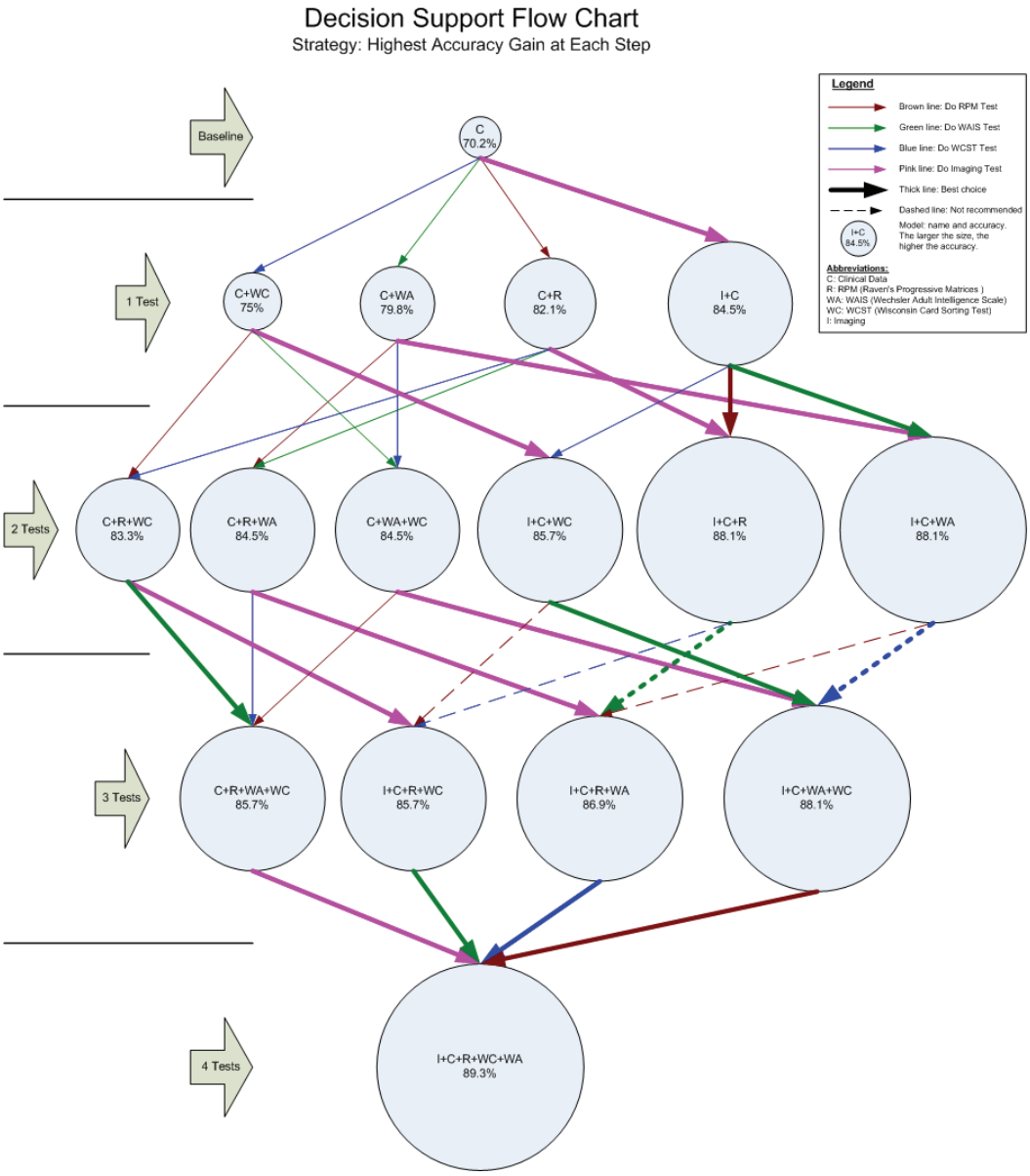

Figure 6.3 Decision support flow chart (strategy: highest accuracy gain)
Circles represent models, with size representing accuracy. Models are arranged in layers according to the number of tests (baseline: 0 test, other layers: 1 to 4 tests) they contain, and they are sorted by accuracy in each layer. Arrows of different colors represent different tests. This flow chart helps clinicians to choose the best test in different situations by following the thick arrows.

Starting from any model, if a test is done, it will flow to a new model along the arrow representing that test. For example, at model C+R, if test WCST is done, it will flow to model C+R+WC, by following the blue arrow (representing WCST test).

An arrow of thick line represents the locally best choice test (meaning it leads to a model with highest accuracy gain). For example, if a person has already done the RPM test, then the clinician can choose the next test from WAIS, WCST and Imaging. In the flow chart, we can see that starting from model C+R (accuracy 82.1%), there are three outgoing arrows pointing out to other models. The best choice is the arrow with thick line, in this case, the pink colored arrow, which represents Imaging Test. That means, the Imaging Test should be chosen as the next test. After the Imaging Test, the classification accuracy will become 88.1% (model I+C+R).

If a model has two (or more) thick arrows pointing out, it means doing these two tests will have the same accuracy gain. For example, from model I+C, a brown arrow (representing RPM test) and green arrow (WAIS test) are both in thick line. That means, doing these two tests will lead to two models (model I+C+R, and model I+C+WA) with same accuracy (88.1%).

An arrow with dashed line represents a test which is not recommended (meaning it does not increase the accuracy at all). For example, from model I+C+R (accuracy 88.1%), doing test WCST (blue arrow) or WAIS (green arrow) leads to new models with lower accuracy, hence they are not suggested.

In addition, this flow chart can also be used by clinicians to choose combinations of multiple tests directly since it give the clinicians a global view of all possible tests. For example, if clinicians want to do two tests, they can immediately find the best combinations (model I+C+R, and model I+C+WA) from the "2 Tests" layer. Another example is that if the clinicians want to find out what combinations of tests can give more than 88% of accuracy, they can directly locate model I+C+R, I+C+WA, I+C+WA+WC, and I+C+R+WA+WC from the chart by looking at areas with large circles.

Although from the baseline model C (accuracy 70.2%) to the most comprehensive model I+C+R+WA+WC (accuracy 89.3%), the accuracy increment is at substantial level of 19.1%, this increment is not always noticeable for each test added. For example, from model C+R+WC (accuracy 83.3%), to model I+C+R+WC (accuracy 85.7%), the accuracy increment is only 2.4% by adding the imaging test. In such a case with small steps of accuracy increment, clinicians should consider the effectiveness of the test recommended in the flow chart.

We also make another Decision Support Flow Chart (Figure 6.4) based on a different strategy. We choose the best further test by selecting the test that has the highest cost effectiveness. Cost effectiveness (*CE*) is a measurement of a test. It is defined as accuracy gain from a model to new model (in terms of percentage) divided by the cost of the additional test that the new mode has. Throughout the study, the private rate (non-subsidized) costs are used for the calculation.

$$CE_{t,m} = \frac{Acc_{m+t} - Acc_m}{Cost_t} \quad \textbf{(6.3)}$$

where $CE_{t,m}$ is the Cost Effectiveness of test t from model m; Acc_m is the accuracy of model m, Acc_{m+t} is the accuracy of another model $m+t$ (a model with additional test t), and $Cost_t$ is the cost of test t.

The unit of Cost Effectiveness is percent/$. For example a Cost Effectiveness of 0.02%/$ means, for every dollar ($) spent on the test, an accuracy gain of 0.02% can be achieved.

A more meaningful measurement is Relative Cost, which is defined as the reciprocal of Cost Effectiveness:

$$RC_{t,m} = \frac{1}{CE_{t,m}} = \frac{Cost_t}{Acc_{m+t} - Acc_m} \quad \textbf{(6.4)}$$

where $RC_{t,m}$ is the Relative Cost of test t from model m; $CE_{t,m}$ is the Cost Effectiveness of test t from model m; Acc_m is the accuracy of model m, Acc_{m+t} is the accuracy of another model $m+t$ (a model with additional test t), and $Cost_t$ is the cost of test t.

The unit of Relative Cost is $/percent. It can be interpreted as: for each percent of accuracy gain, how much money is spent. In the decision support flow chart, the Relative Cost is also displayed along the thick arrows (the most cost effective choice).

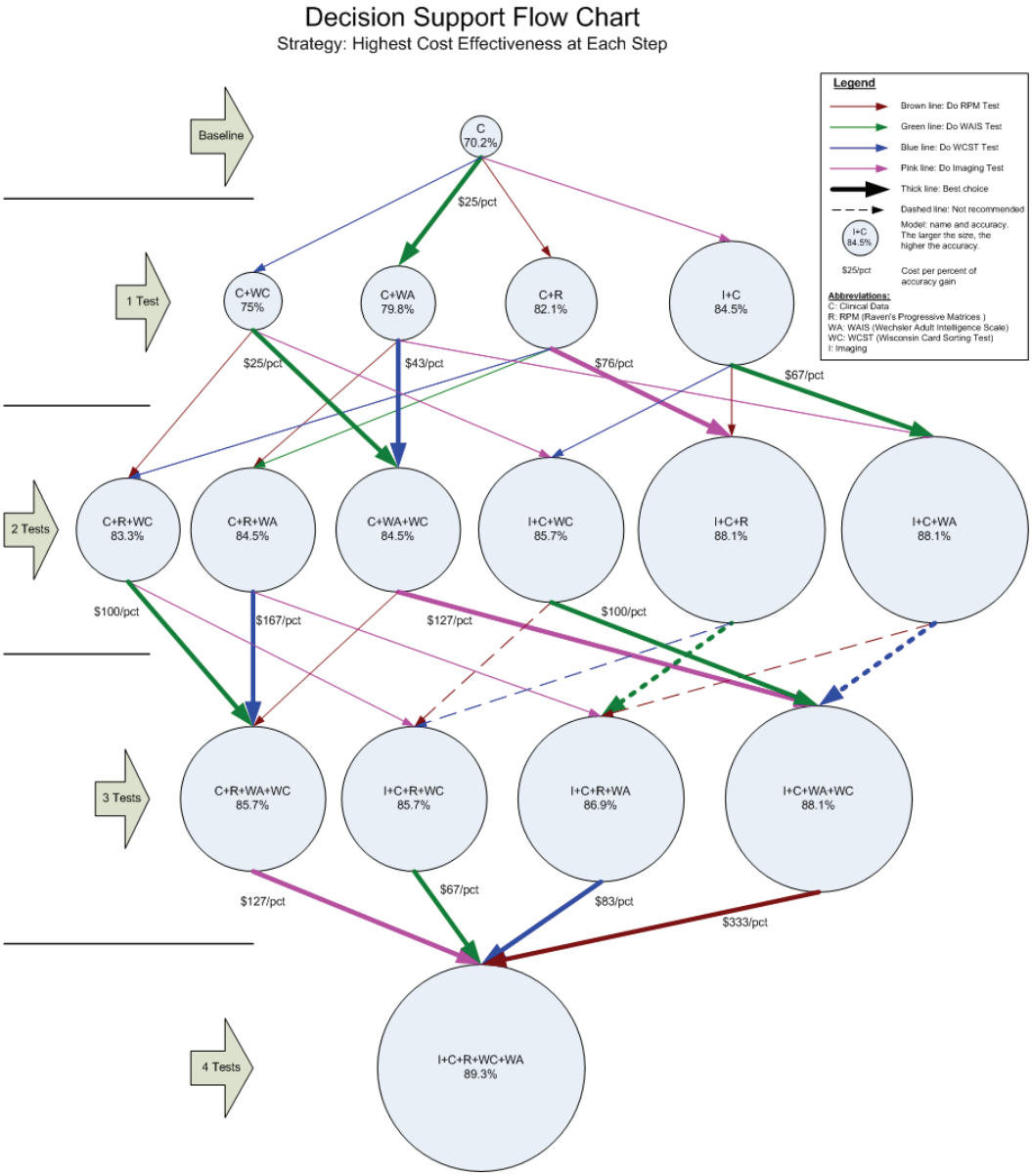

Figure 6.4 Decision support flow chart (strategy: highest cost effectiveness)
Circles represent models, with size representing accuracy. Models are arranged in layers according to the number of tests (baseline: 0 test, other layers: 1 to 4 tests) they contain, and they are sorted by accuracy in each layer. Arrows of different colors represent different tests. The number along a thick arrow is the relative cost, cost per percent of accuracy gain. This flow chart helps clinicians to choose the most cost effective test in different situations by following the thick arrows.

The cost effectiveness based flow chart helps clinicians to select a suitable test at situations when the cost is a concern. For example, from model C+WC, the most cost effective test is WAIS, which flows to model C+WA+WC. A 9.5% of accuracy gain (from 75% to 84.5%) is achieved, and the relative cost is $25/pct.

Compared to the best choice (Imaging test) from previous flow chart (on strategy of highest accuracy gain), a different decision (WAIS test) is suggested.

However, even the flow chart can recommend a most cost effective test among all possible further tests for each step theoretically; the clinicians shall also consider the absolute accuracy increment before a test is performed. For example model C+WA+WC (accuracy 84.5%), the flow chart recommends the imaging test which leads to model I+C+WA+WC (accuracy 88.1%), but the accuracy increment is 3.6% only. Clinicians shall make decision on if such an increment is practical useful according to their needs.

In summary, the decision support flow charts provide useful tools for clinicians in selecting suitable tests on patients after classification models are validated in large scale trials.

6.2.2 Decision Support System Software

We develop the Decision Support System software in pure Java programming language. Java Swing is used to build the GUI part. We use the Weka library (Ver 3.4.13, University of Waikato, New Zealand) to perform functions such as Bayesian Network model loading, and instance classification.

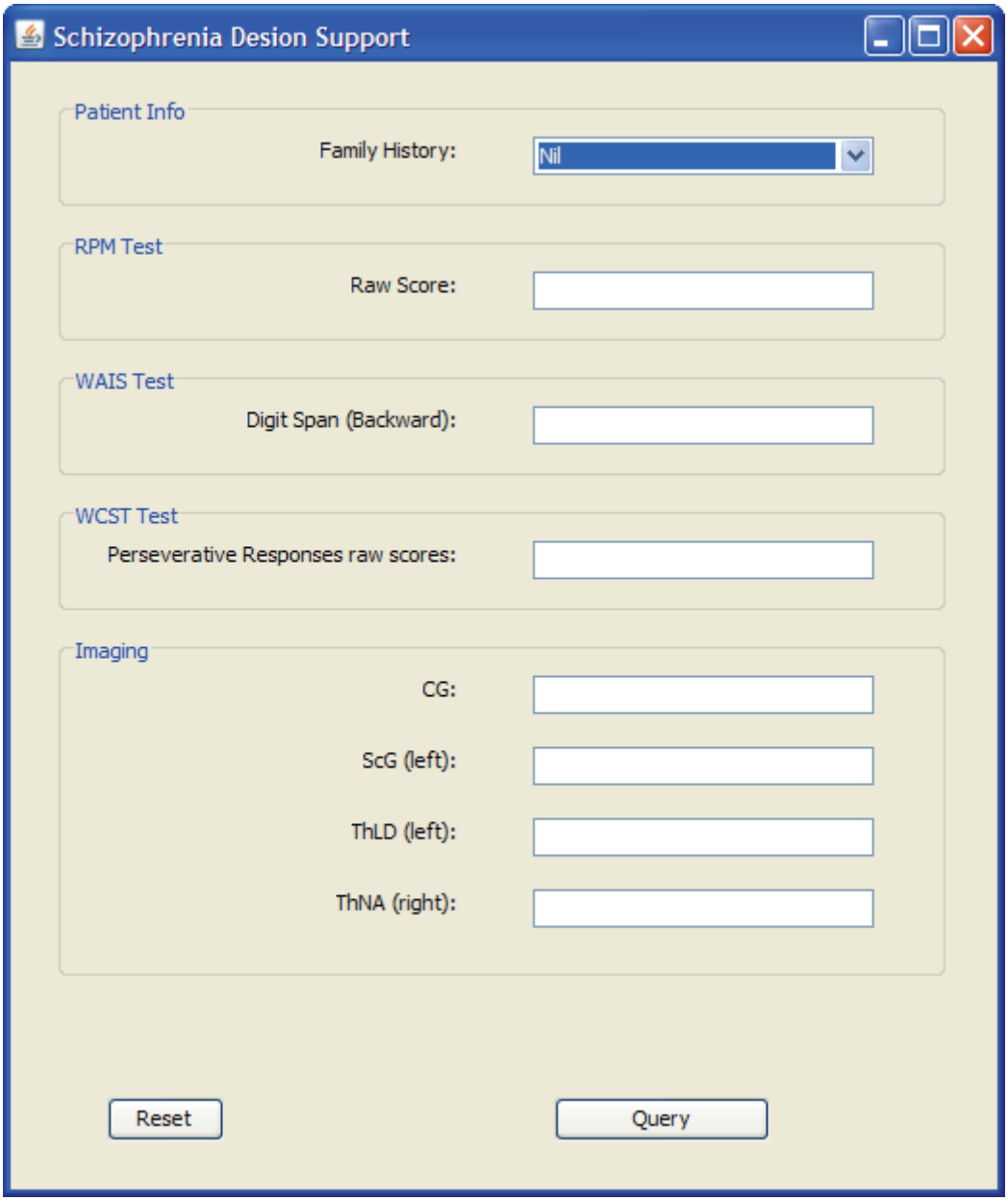

Figure 6.5 Decision support system user input GUI

We have tested run the software on Window XP platform. Figure 6.5 is a screen capture of the data input GUI. Users can select family history from a dropdown list. Only selected data are required to be input for neurocognitive tests, and neuroimaging test. If a test is not done, the relevant fields can be left blank. The Reset button is used to clear the user's input.

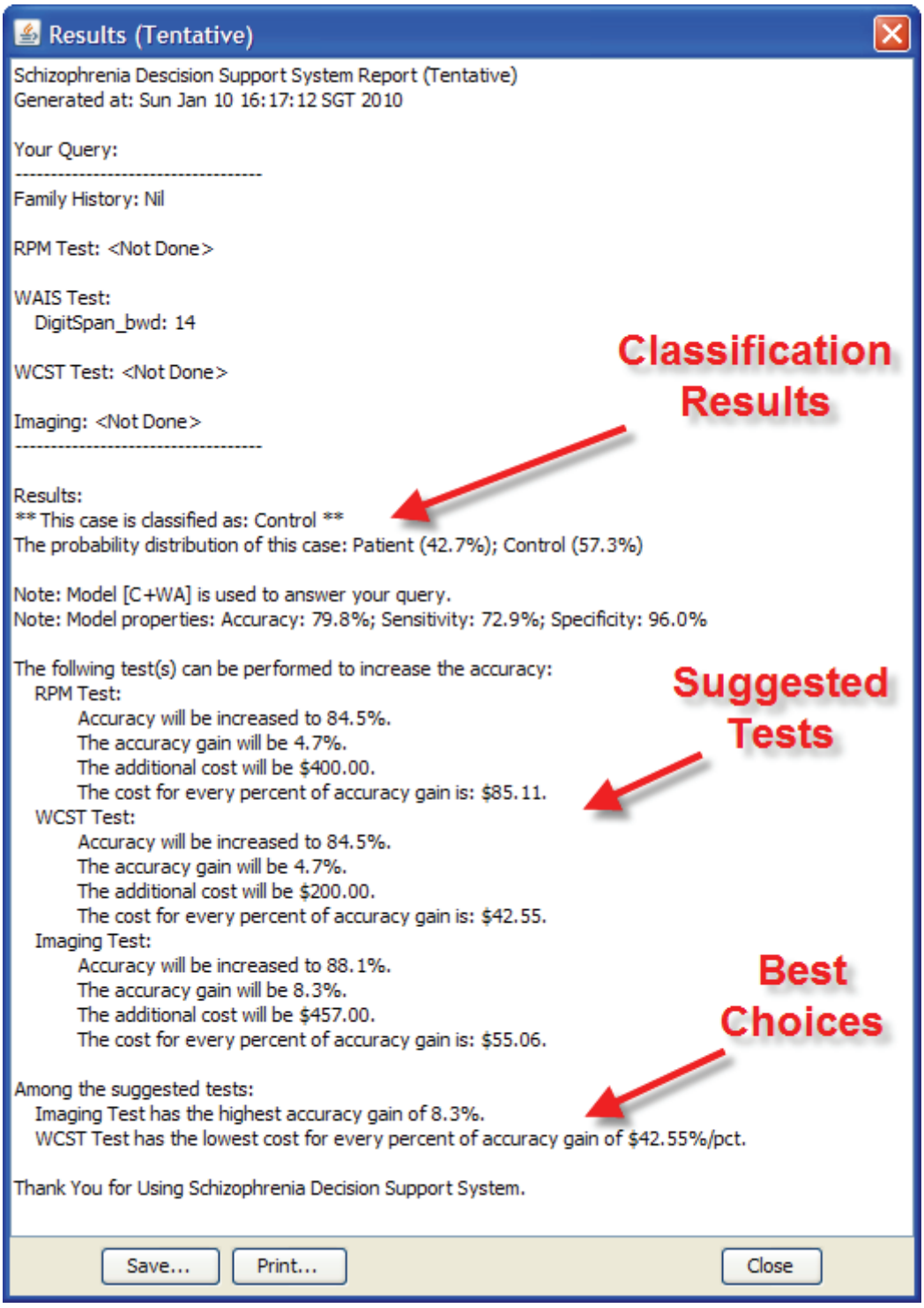

Figure 6.6 Report with classification results and suggested further tests

After inputting necessary information from the Data Input GUI, the user clicks the Query button. The Decision Support Engine uses the user input data to select a suitable schizophrenia classification model to use. Then the selected Bayesian

Network model is used to classify the user's case. A report based on the classification results will be generated and shown to the user. An example of a report is shown as in Figure 6.6. Along with the classification result (in this case, a normal control) and the probability distribution, the decision support system shows three suggested further tests (RPM, WCST and Imaging) that will increase the classification accuracy. Finally it shows the best choices according to different strategies (highest accuracy gain and highest cost effectiveness).

In summary, we made two Decision Support Flow Charts that helps clinicians to choose suitable tests in order to improve the diagnostic accuracy with or without the cost consideration. We also developed a decision support system. It provides support in schizophrenia diagnosis by using objective criteria, such as family history and various quantifiable neurocognitive tests results, and neuroimaging features. In order to achieve higher diagnosing accuracy, it also gives suggestions on what tests should be done, and shows their accuracy gains and cost effectiveness.

6.3 Performance of Decision Support System

We estimate the overall performance of the decision support system. We use average accuracy to represent the accuracy of the decision support system, based on the assumption that all models have the same opportunity to be applied. That is because all neurocognitive tests and neuroimaging tests are independent and we assume there is no preference in selecting any tests. Hence, we define the overall accuracy ($Acc_{overall}$) as the mean accuracy of all models used in the system.

$$Acc_{overall} = \frac{\sum_{m} Acc_m}{N} \quad \textbf{(6.5)}$$

where Acc_m is a model's accuracy; N is the number of models.

Since our decision support system uses 16 models (Acc_{16}) (The 16 models come from clinical data plus all combinations of 4 independent tests, i.e., 3 neurocognitive tests and 1 neuroimaging test, 2^4=16). The overall accuracy $Acc_{overall}$ is calculated as:

$Acc_{overall} = Acc_{16} = 83.8\%$

Model C takes into account of clinical information only (family history). As we have pointed out previously, this model classifies all cases as patients, and so its type I error is 100%. Hence it shall not be applied in practice. If we exclude this model from decision support system (a more realistic situation because diagnosis should not solely depends on family history), the mean accuracy of all the other models becomes 84.8%.

$Acc_{overall} = Acc_{15} = 84.8\%$

In summary, a schizophrenia decision support system is developed by using all models we constructed. The models are based on subjective criteria, such as family history of psychiatric disease, neurocognitive tests results, and neuroimaging

results. The performance of the decision support system can be measured by the overall classification accuracy.

6.4 Performance of Cost Based Decision Support System

Based on the private rated cost of all test (Table 6.1), the cost of all models can be calculated by their component individual tests. The Accuracy Gain (AccGain) is defined as the increased accuracy of a model compared with the baseline model (Model C). Relative Cost (*RC*) for each model is calculated as the cost of the model divided by the *AccGain*. The results are shown in the Table 6.2:

Table 6.2 Accuracy and Cost of models

Model →		**C**	**C+R**	**C+WA**	**C+WC**	**C+R+WA**	**C+R+WC**	**C+WA+WC**	**C+R+WA+WC**
Acc		70.2%	82.1%	79.8%	75.0%	84.5%	83.3%	84.5%	85.7%
AccGain			11.9%	9.6%	4.8%	14.3%	13.1%	14.3%	15.5%
Cost			$400	$240	$200	$640	$600	$440	$840
RC			$33.61	$25.00	$41.67	$44.76	$45.80	$30.77	$54.19
Model →	**I**	**I+C**	**I+C+R**	**I+C+WA**	**I+C+WC**	**I+C+R+WA**	**I+C+R+WC**	**I+C+WA+WC**	**I+C+R+WA+WC**
Acc	77.4%	84.5%	88.1%	88.1%	85.7%	86.9%	85.7%	88.1%	89.3%
AccGain	7.2%	14.3%	17.9%	17.9%	15.5%	16.7%	15.5%	17.9%	19.1%
Cost	$457	$457	$857	$697	$657	$1,097	$1,057	$897	$1,297
RC	$63.47	$31.96	$47.88	$38.94	$42.39	$65.69	$68.19	$50.11	$67.91

Abbreviations: Acc, Accuracy; AccGain, Accuracy Gain; RC, Relative Cost; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test; I, Imaging

We define the overall relative cost ($RC_{overall}$) as the mean relative cost of all models used in the system.

$$RC_{overall} = \frac{\sum_{m} RC_m}{N} \quad \textbf{(6.6)}$$

where RC_m is a model's relative cost; N is the number of models

Since Model C's relative cost is undefined, the overall relative cost of the decision support system will be consist of the rest 15 models (without Model I).

$RC_{overall} = RC_{15} = \$47.02/\%$

Which means, by using the decision support system, people can expect an average 1 percent of accuracy increase compared to the baseline model C for every $47.02 spent on testing.

The relative cost of all models and the decision support system overall is shown in Figure 6.7. Each dot represents a model. The red line is the overall relative cost of the decision support system, which is $47.02/%. From cost effectiveness point of view, the lower the relative cost, the better the model. We can see that although the most comprehensive model I+C+R+WA+WC has the highest accuracy, it almost has the highest relative cost $67.91/%.

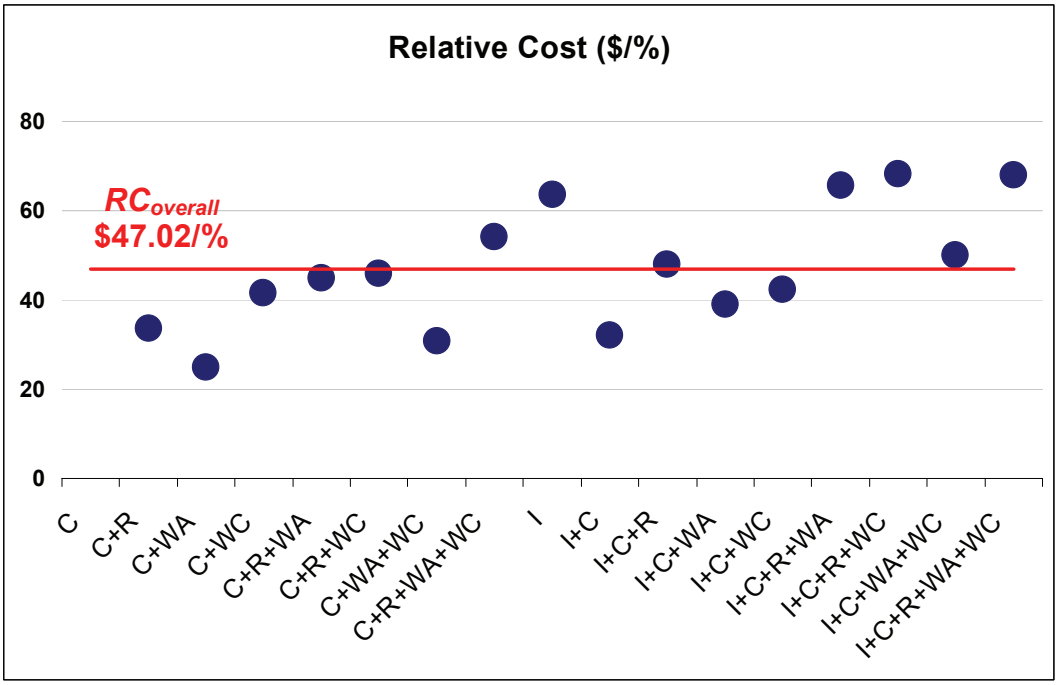

Figure 6.7 Relative Costs of Models and Overall Relative Cost of Decision Support System Abbreviations: RC, Relative Cost; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test; I, Imaging

Chapter 7

Conclusions and Discussion

Schizophrenia is a very common psychiatric disease which affects about 1% of the world's population. It does not only damage to the patient's health, but also is a big economical burden to the patient's family and the whole society. However the existing diagnosis of schizophrenia heavily depends on subjective criteria, such as family member's observation of patient's symptoms (for example, bizarre behaviors). There is no lab test for this disease.

We aim to reveal the relationship between schizophrenia and the objective and quantifiable criteria from neuroinformatics and neuroimaging. 156 study subjects, including 92 schizophrenia patients and 64 healthy normal controls are recruited by our collaborating hospitals. All patients are scanned using MRI and DTI imaging. Some of them (84) has completed all 4 neurocognitive tests: RPM, WAIS, CPT and WCST.

7.1 Conclusions

7.1.1 Neuroinformatics Based Modeling

Significant factors are chosen from over 50 clinical items by the feature selection method (Correlation-based Feature Subset). The only highly relevant (statistically

and clinically) factor selected is the fam_hx (family history). Based on that factor, a baseline Bayesian Network classification model is generated.

Unlike RPM, WAIS and WCST tests, CPT test is found not contributing to the classification accuracy. Bayesian Network models are generated when RPM, WAIS and WCST tests results are included. All Bayesian Networks degenerate to Naive Bayesian Networks as the factors selected are conditionally independent of each other on the target node (pt_ctrl). Their classification accuracies range from about 75% to 85.7%.

7.1.2 Neuroimaging Based Modeling

We use the DTI imaging to study the brain white matter abnormalities of schizophrenia. FA images (which measure the neural connectivity level) are generated based on the DTI image.

We apply a method developed at Biomedical Imaging Lab, Agency for Science Technology, and Research to automatically place ROIs on brain images by registering the Talairach-Tournoux brain atlas to the structural MRI images and DTI images. 48 brain structures are identified. Each brain structure is divided into the left hemisphere, right hemisphere, and as a whole. So the FA values in 144 ROIs (48 x 3) are quantified by their mean values, standard deviations, and volumes.

Four factors are chosen after the feature selection: CG (cingulate gyrus), ScG_left (left subcallosal gyrus), ThLD_left (left thalamus: lateral dorsal nucleus) and ThNA_right (right thalamus: anterior nucleus). The Bayesian Network classification model with the 4 imaging features can achieve an accuracy of 77.4%.

7.1.3 Combined Model

Combined models are also constructed by using neuroimaging features and different neurocognitive tests results together. The accuracies of all models are higher than 85%. Not surprisingly, the most comprehensive model consisting of clinical information, all 3 neurocognitive tests results and the neuroimaging features achieves the highest accuracy (89.3%). Compared to the baseline model (where no neurocognitive test and no neuroimaging are included), the accuracy increases by 19.1%.

This proves our hypothesis that the accuracy diagnosis of schizophrenia can be improved by using objective and quantitative criteria from a wider spectrum of modalities including neuroinformatics and neuroimaging.

The schizophrenia models generated in this work combine both neuroimaging features and neuroinformatics features, which has never been attempted before to our best knowledge. The most comprehensive combined model reveals and quantifies the relationship between schizophrenia and the following factors:

- fam_hx
- RPM_raw

- DigitSpan_bwd
- PersResponses_raw
- CG
- ScG_left
- ThLD_left
- ThNA_right

Family history of psychiatric disease (fam_hx) is one of the important factors in schizophrenia as reported by various studies that we mentioned earlier in this thesis; The deficit in eductive and reproductive functions (as assessed by RPM_raw), deficit in verbal working memory (as assessed by DigitSpan_bwd), undue perseverative responses which is caused by frontal lobe deficit (as assessed by PersResponses_raw), and reduced neural connectivity in cingulate gyrus (CG) (for attention function), subcallosal gyrus (ScG) (for left prefrontal and right prefrontal interhemispheric communication), and thalamus lateral dorsal nucleus (ThLD) and anterior nucleus (ThNA) (somatosensory and visuo-spatial functions and modulation of alertness) are the other significant factors associated with schizophrenia.

Although CPT test has been done, it is found to have little relation with schizophrenia. This suggests that the difference in capability of attention and impulsivity is not significant between schizophrenia patients and normal controls when compared to other factors.

7.1.4 Significant Features

From the baseline model, a feature is added to build a new model, which will increase the classification accuracy. We use baseline model and the new model to classify all 84 cases in order to have a closer look at the data pattern of the correctly and incorrectly classified cases.

A partial classification results are shown in Table 7.1. In this table, each case is classified by the baseline model C and the model C+R. Column pt_ctrl is the case's ground truth. Column "C Predict" is the classification results of model C. Column "C Correct?" tells if model C classification results are correct, and so as "C+R Predict" and "C+R Correct?". The last column "From C to C+R" shows the comparison between the two models classification results. If their classification results are the same, the value is "Same". If model C fails to classify the case, but model C+R successfully classifies it, the value is "Improved". In contrast, if model C+R fails to classify the case, but model C successfully classifies it, the is "Worse".

The comparison results are visualized in Figure 7.1. The top figure shows the distribution of cases in the 2D plot of fam_hx and RPM_raw feature space. Black circles represent correctly classified cases, and red circles represent cases that are classified wrongly. We notice that all failed cases fall within the zone of fam_hx = Nil and RPM_raw range about 45 to 60. The bottom figure shows the comparison of the classification results of the two models. We see that there is no improvement of the new model C+R when RPM_raw is below 50. However, the new model C+R makes many improvements when RPM_raw greater than 50.

RPM_raw is the number of missing patterns correctly identified by a person in the RPM test. It seems the score of 50 is a low limit for model C+R to work well, which means when RPM_raw is greater than 50, it is more useful in contributing to the classification of schizophrenia; below that, it is not very sensitive to the classification.

Table 7.1 Model classification results comparison (partial)

CDNo	pt_ctrl	C Predict	C Correct?	C+R Predict	C+R Correct?	From C to C+R
1	Patient	Patient	Yes	Patient	Yes	Same
2	Patient	Patient	Yes	Patient	Yes	Same
5	Patient	Patient	Yes	Patient	Yes	Same
6	Patient	Patient	Yes	Control	No	Worse
7	Patient	Patient	Yes	Patient	Yes	Same
14	Control	Patient	No	Patient	No	Same
20	Patient	Patient	Yes	Patient	Yes	Same
21	Patient	Patient	Yes	Patient	Yes	Same
23	Control	Patient	No	Control	Yes	Improved
26	Patient	Patient	Yes	Patient	Yes	Same
32	Control	Patient	No	Control	Yes	Improved
33	Control	Patient	No	Control	Yes	Improved
34	Control	Patient	No	Control	Yes	Improved
38	Patient	Patient	Yes	Control	No	Worse
39	Control	Patient	No	Control	Yes	Improved
40	Control	Patient	No	Control	Yes	Improved
42	Patient	Patient	Yes	Patient	Yes	Same

Abbreviations: C, Clinical Data; R, RPM Test; pt_ctrl, patient or control

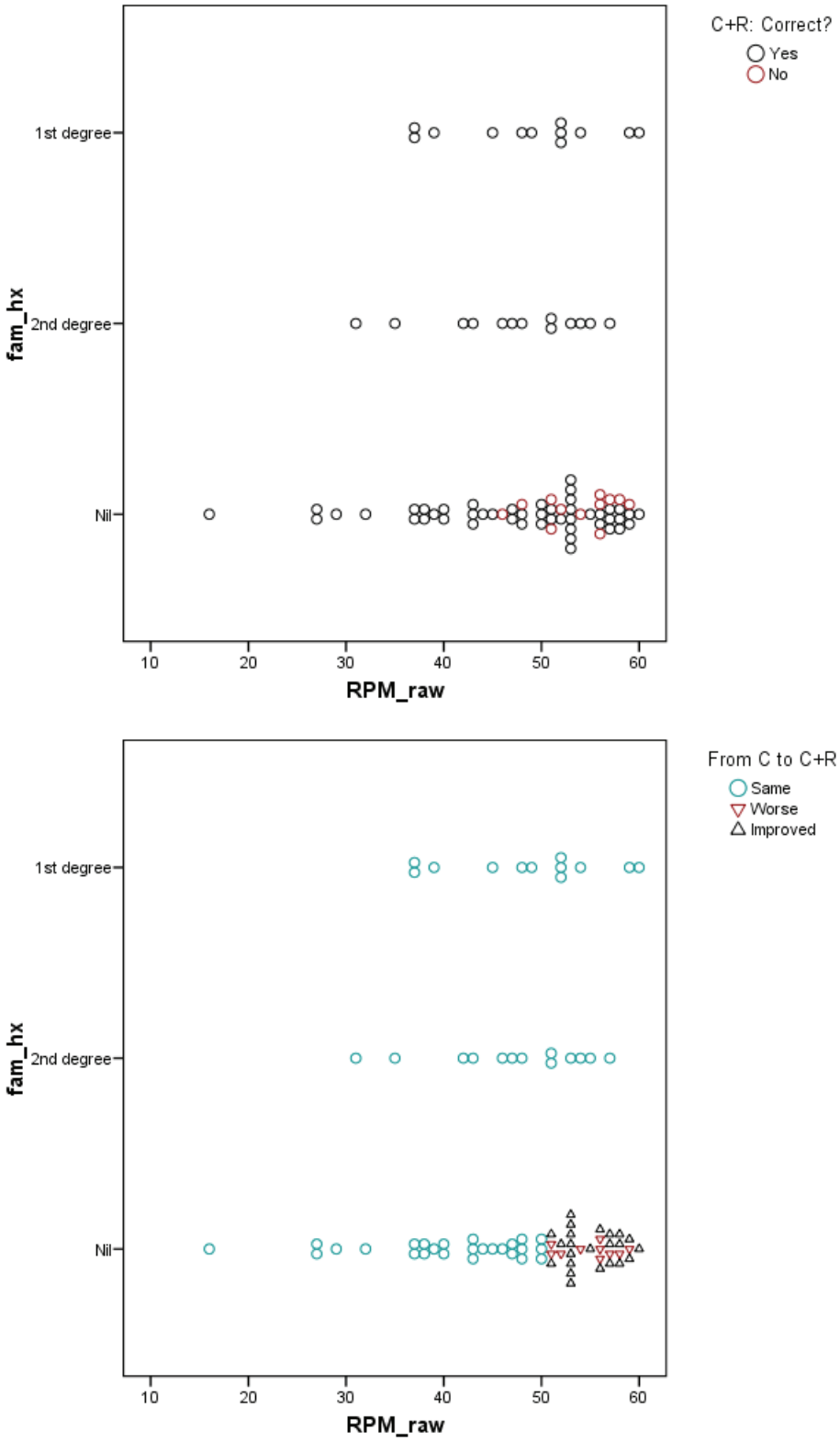

Figure 7.1 Case distribution for model C+R

Top diagram shows if the classification result is correct. Black circle, Yes; Red circle, No; Bottom diagram shows the comparison of the classification results of the old model C and the new model C+R. Cyan circle, same result; Red downwards triangle, Worse (old model is correct, new model is wrong); Black upward triangle, Improved (old model is wrong, new model is correct); Abbreviations: C, Clinical Data; R, RPM Test;

Similarly, the distribution of classification results and the comparison with baseline model C of the model C+WA is visualized in Figure 7.2. The top diagram shows the classification results in the fam_hx and DigitSpan_bwd space. The bottom diagrams shows if the results are improved by adding WAIS test into the baseline model. As we can see that most wrongly classified cases are in the zone of fam_hx = Nil and DigitSpan_bwd score 7, 8, 9. When DigitSpan_bwd is lower than 7, model C+WA performs the same as model C, whereas for score higher than 9, model C+WA outperforms model C in 8 out of 10 cases.

DigitSpan_bwd is the longest of number of digits that a person can repeat correctly in the reverse order. It assesses the memory capacity of a person. This score contributes to the model classification capability when it is blower than 7, and higher than 9. When it is in the mid-range of 7 to 9, its contribution is mixed. A total of 30 cases (16 improved and 14 worse) fall in this range. A further investigation of the 30 cases reveals that DigitSpan_bwd is close to but not significantly different ($p=0.058$) between the improved cases (mean 7.44, SD 0.629) and worse cases (mean 8.0, SD 0.877). However, by using the two-tailed independent t test, SpatialSpan_bwd is found to be significantly different ($p=0.005$) between them: for the improved cases, the mean score is 8.94 (SD 1.436), and for the worse cases, it is 7.14 (SD 1.703). This suggests SpatialSpan_bwd score (which assesses spatial working memory) can be a good candidate to suplement the DigitSpan_bwd score when it fails to work in the mid-range of 7 to 9. It is reasonable since both tests are for the memory capacity aspects.

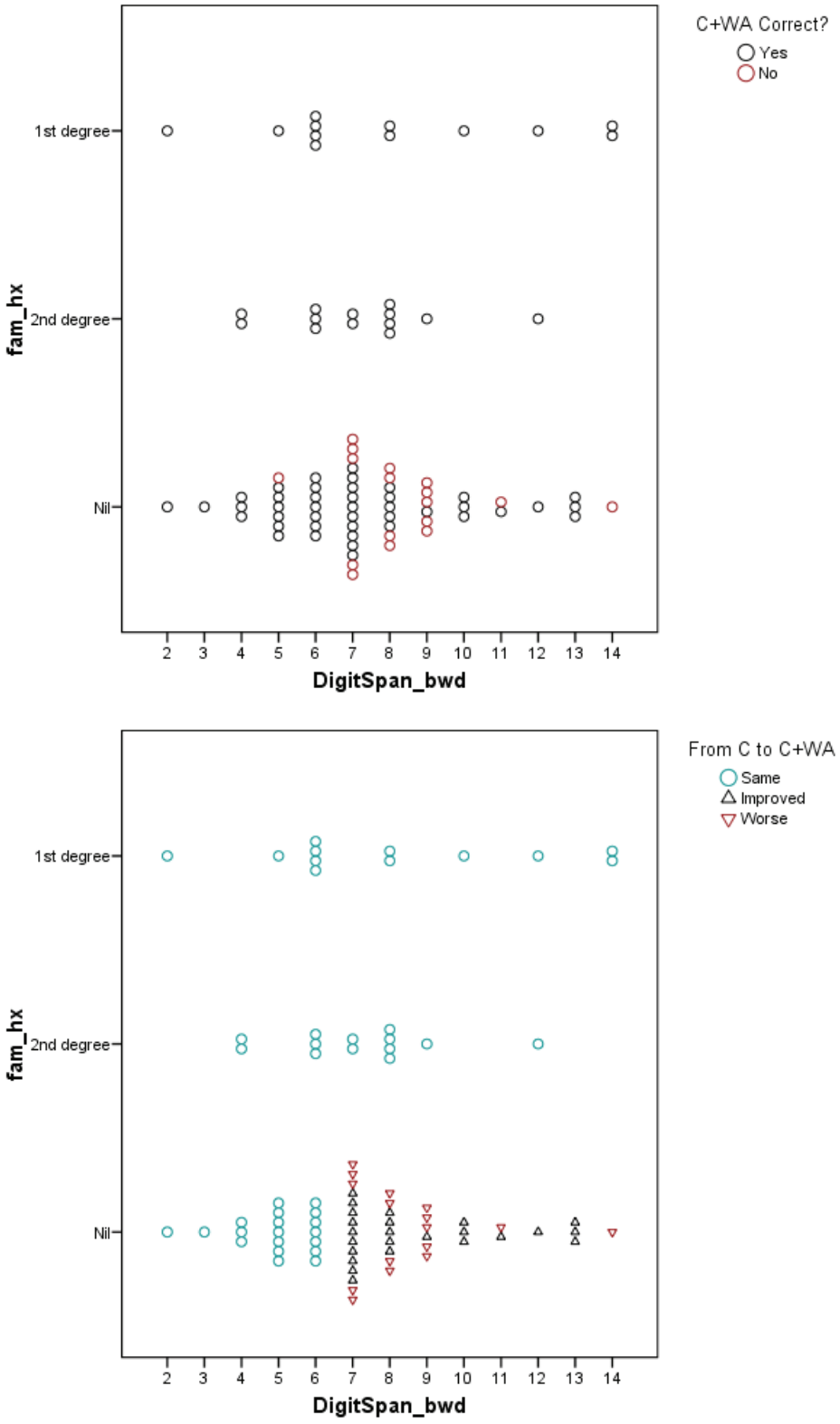

Figure 7.2 Case distribution for model C+WA

Top diagram shows if the classification result is correct. Black circle, Yes; Red circle, No; Bottom diagram shows the comparison of the classification results of the old model C and the new model C+WA. Cyan circle, same result; Red downwards triangle, Worse (old model is correct, new model is wrong); Black upward triangle, Improved (old model is wrong, new model is correct); Abbreviations: C, Clinical Data; WA, WAIS Test;

The results of model C+WC is shown in Figure 7.3. The top part is the distribution of correctly and wrongly classified cases in fam_hx and PersResponses_raw space. The bottom diagrams shows whether the results are improved by adding WCST test into the baseline model. Please note that for the clear visualization purpose only, the PersResponses_raw score is rounded to the nearest 5. Otherwise the indicators (circles and triangles) will be packed together along the horizontal line since PersResponses_raw takes different integer values close to each other, which prevents them from being stacked along the vertical direction.

We observed that almost all wrongly classified cases (20 cases) fall in the zone of fam_hx = Nil and PersResponses_raw below about 30, in contrast, there are 24 correctly identified cases in the same zone.

PersResponses_raw is reported to be associated with the dysfunction of frontal lobe. When it is higher than 30, the model can classify cases very well. When it is lower than 30, the model's performance is mixed. Further investigation (using independent t test) of the these subset of case (PersResponses_raw less than 30) doesn't find any feature from WCST test that is able to significantly improve the classification accuray of this subset.

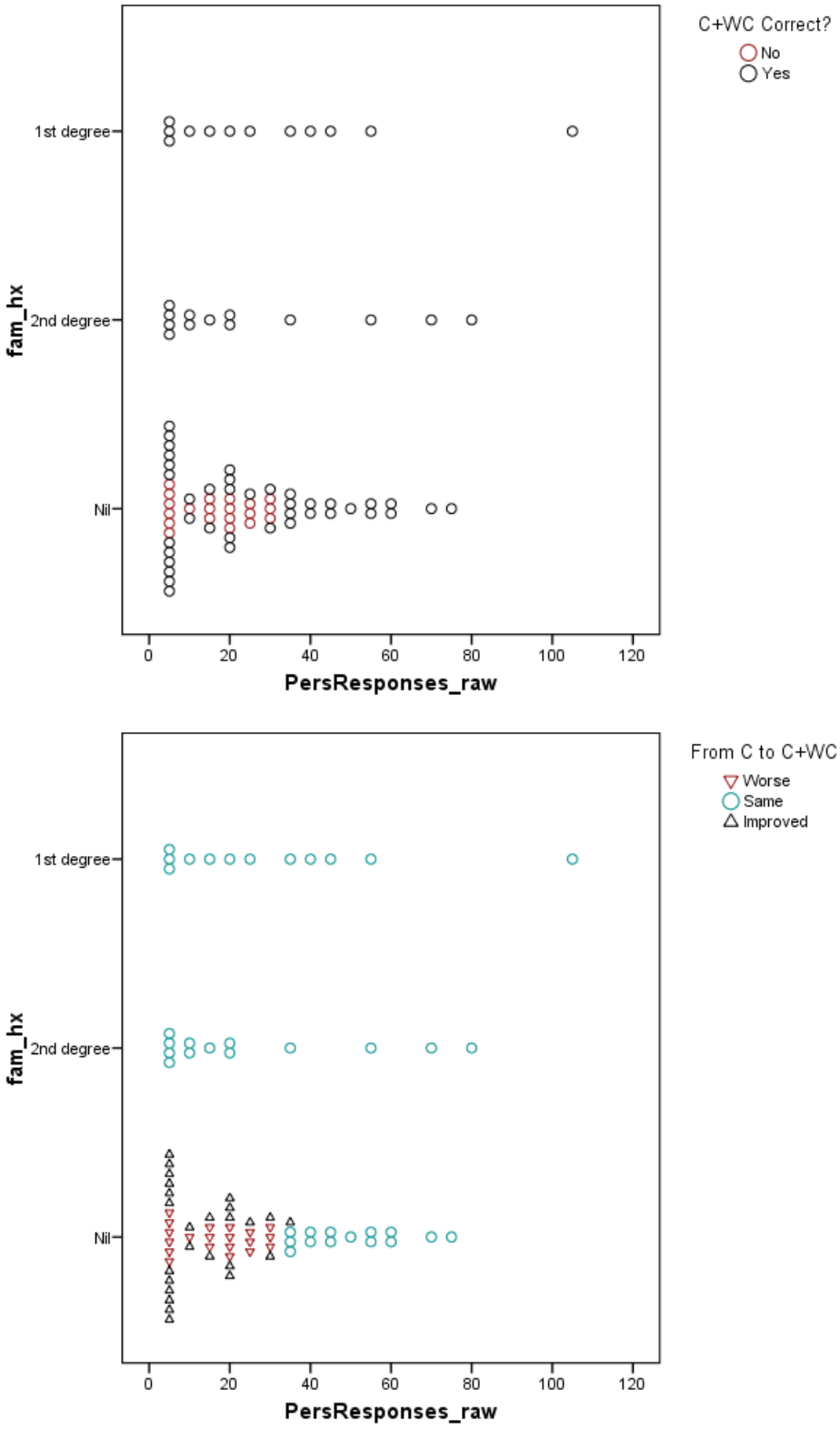

Figure 7.3 Case distribution for model C+WC

Top diagram shows if the classification result is correct. Black circle, Yes; Red circle, No; Bottom diagram shows the comparison of the classification results of the old model C and the new model C+WC. Cyan circle, same result; Red downwards triangle, Worse (old model is correct, new model is wrong); Black upward triangle, Improved (old model is wrong, new model is correct); Abbreviations: C, Clinical Data; WC, WCST Test; Note: PersResponses_raw score is rounded to the nearest 5.

Results of model I+C are illustrated in Figure 7.4 and Figure 7.5. Since there are 4 significant image features, CG, ScG_left, ThLD_left and ThNA_right, the case distributions are plotted in 2 separated 2D plots: one for CG and ScG_left space (Figure part A), the one for ThLD_left and ThNA_right (Figure part B). In each part of the figure, the top diagram shows the distribution of correctly and wrongly classified cases, whereas the bottom diagram depicts if the results are improved by adding imaging test into the baseline model.

It is observed that wrongly classified cases are located within the zone of CG range about 0.25 to 0.275 and ScG_left range about 0.25 to 0.35, as well as ThLD_left range about 0.3 to 0.4 and ThNA_right range about 0.37 to 0.43. Outside of the zone, the model I+C can classify most cases correctly.

Except for CG (where 0.25 to 0.275 is at about its middle range), the other 3 ranges are close to the high end of values. When we place the same ranges in distribution of patients and controls (Figure 7.6), not surprisingly, we can see that these ranges cover the area with mostly mixed patients and controls.

As we know, these 4 image features are FA values that are associated with the strength of neural connectivity. Lower values usually imply defects in the brain connections. We can see this from Figure 7.6 that lower left portions (low FA values) of both diagrams contain mostly patients.

Compared to baseline model, this model I+C outperforms by 24 improved cases, but underperforms by 8 worse cases (which are actually patients, but wrongly classified as controls, where as model C classifies them correctly as patients). A further investigation shows that this group of 8 patients has a mean ThLD_left value of 0.3625 (SD 0.0587) and a mean ThNA_right value of 0.3866 (SD 0.0252). They are even higher than overall mean value of all controls, ThLD_left 0.3497 (SD 0.0857) and ThNA_right 0.3595 (SD 0.0408), although not statistically significantly. Not other difference between the rest patients is found, except for the weight. The mean weight of the 8 patients is 74.65kg (SD 16.94) is significantly higher than the rest patients (mean 61.25kg, SD 10.37). However no other literature reports a directly relationship between body weight and neural connectivity functions in thalamus.

This observation reveals that some schizophrenia patients (8 out of 51, or 15.7% in our study) do not suffer from decreased FA values in the thalamic regions. It also shows that model I+C may not work in patients with high ThLD_left and ThNA_right values.

In conclusion, in terms of increasing classification accuracy, RPM_raw performs better when it is greater than 50, when compared to less than 50. DigitSpan_bwd performs well when it is lower than 7 or greater than 9; and SpatialSpan_bwd score can suplement when DigitSpan_bwd is between 7 to 9. PersResponses_raw is more helpful in increasing the classification accuracy when it is higher than 30. Imaging features contributes more to classification accuracy, when outside the CG range of

about 0.25 to 0.275, ScG_left range of about 0.25 to 0.35, ThLD_left range of about 0.3 to 0.4 and ThNA_right range about 0.37 to 0.43.

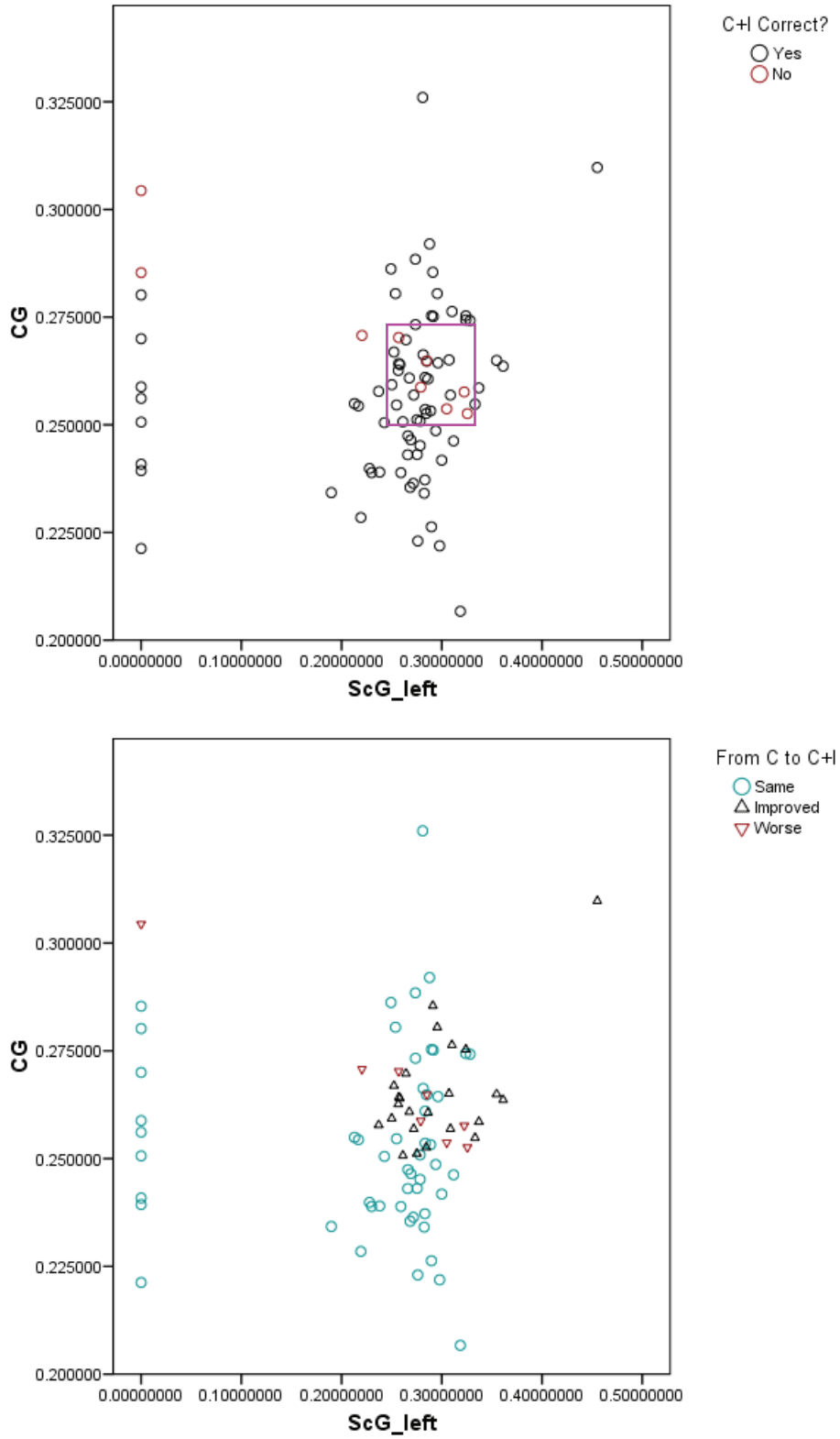

Figure 7.4 Case distribution for model I+C (part A)

Top diagram shows if the classification result is correct. Black circle, Yes; Red circle, No; Pink Rectangle, Range of most wrongly classified cases; Bottom diagram shows the comparison of the classification results of the old model C and the new model I+C. Cyan circle, same result; Red downwards triangle, Worse (old model is correct, new model is wrong); Black upward triangle, Improved (old model is wrong, new model is correct); Abbreviations: C, Clinical Data; I, Imaging Test;

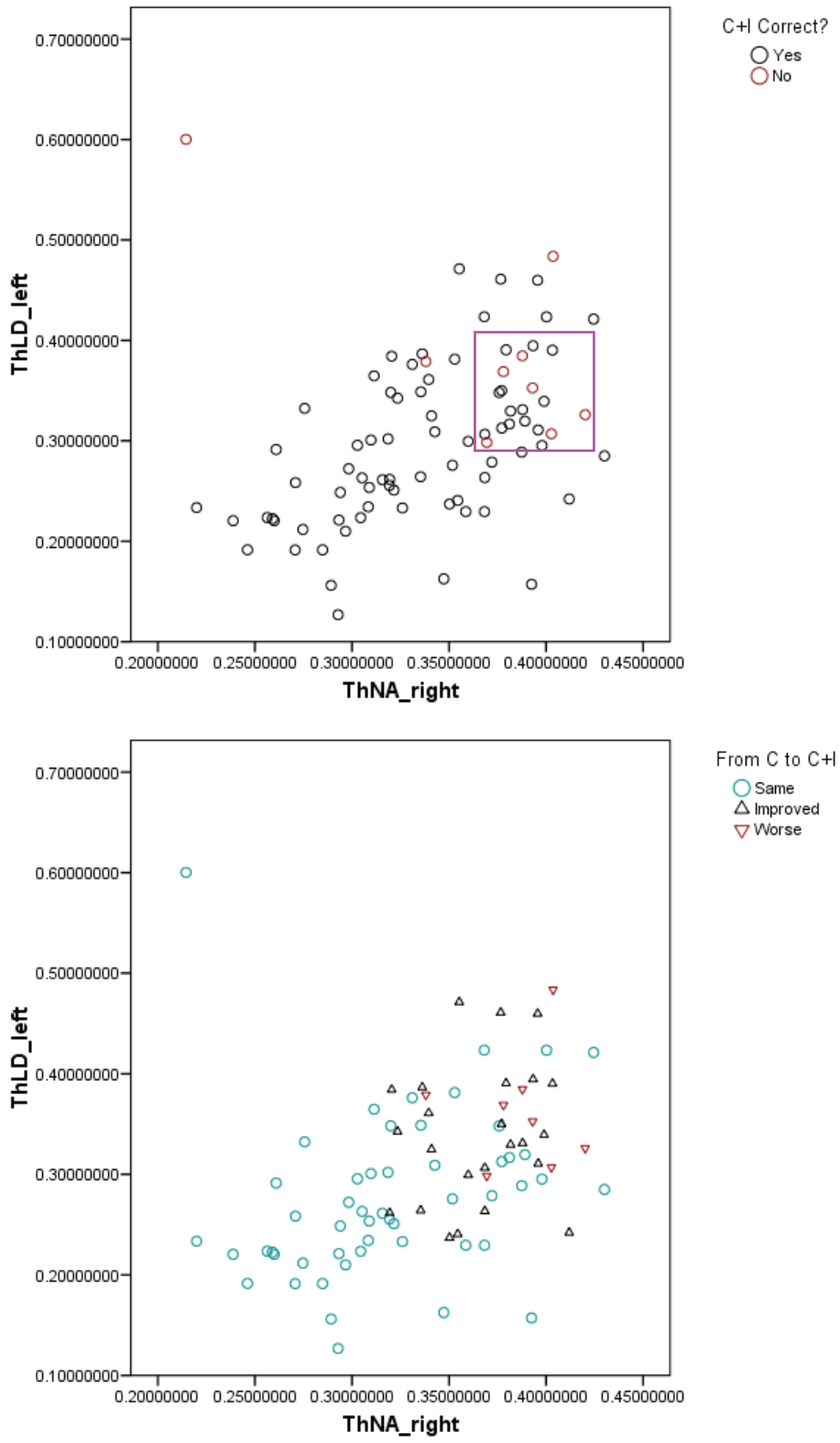

Figure 7.5 Case distribution for model I+C (part B)

Top diagram shows if the classification result is correct. Black circle, Yes; Red circle, No; Pink Rectangle, Range of most wrongly classified cases; Bottom diagram shows the comparison of the classification results of the old model C and the new model I+C. Cyan circle, same result; Red downwards triangle, Worse (old model is correct, new model is wrong); Black upward triangle, Improved (old model is wrong, new model is correct); Abbreviations: C, Clinical Data; I, Imaging Test;

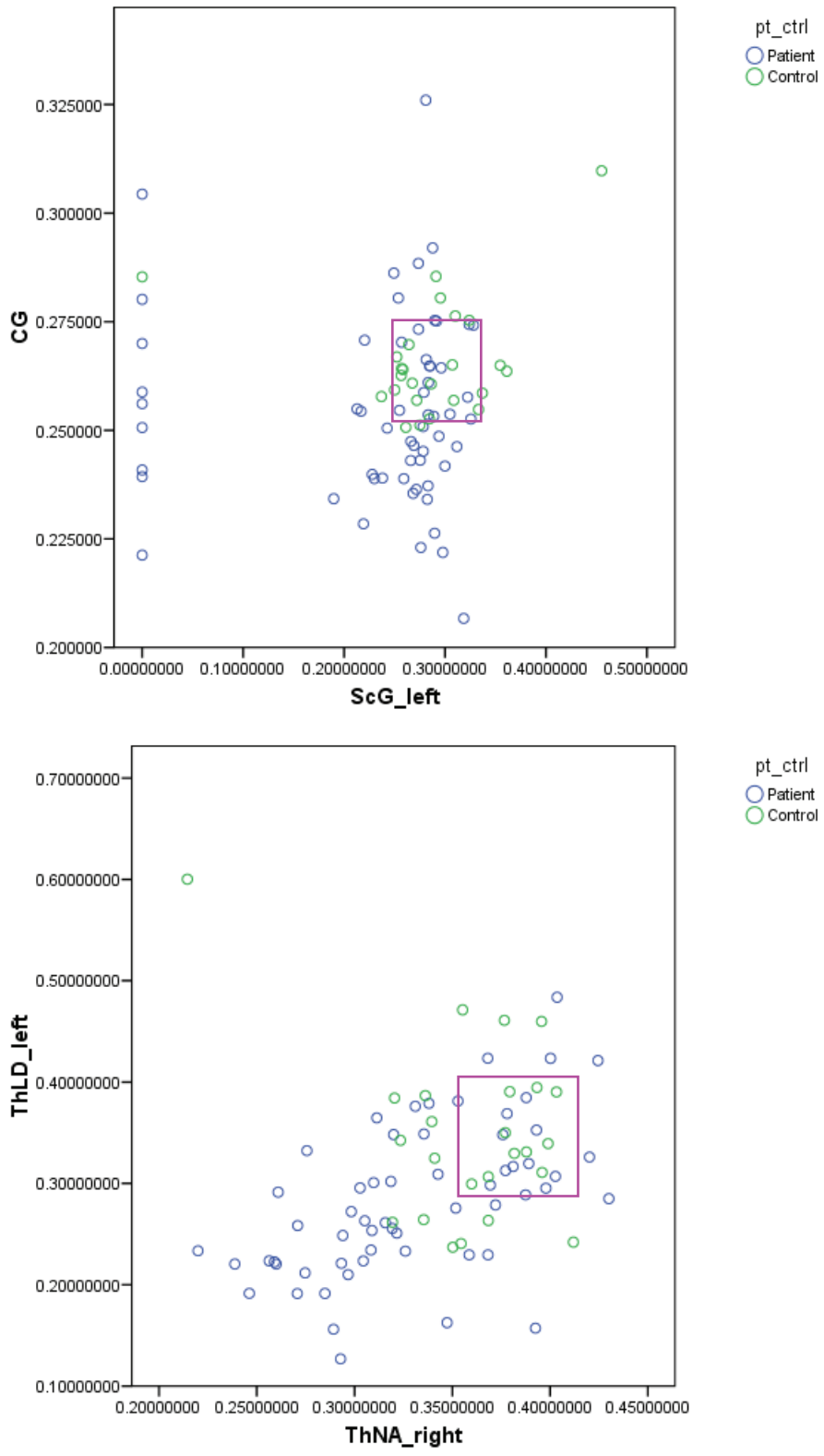

Figure 7.6 Distribution of patients and controls
Top: distribution cases in CG and ScG_left space; Bottom: distribution of cases in ThLD_left and ThNA_right space; Blue circle, Patient; Green circle, Control; Pink Rectangle, Range of most mixed patients and controls;

7.1.5 Decision Support System

Based on all schizophrenia classification models, we make two decision support flow charts to choose suitable tests by using different strategies. One strategy is the highest accuracy gain. At each step, clinicians can follow the flow chart to choose the best further test that leads to a new model with the highest possible accuracy, regardless of the cost. Another strategy is the highest cost effectiveness. At each step, clinicians can follow the flow chart to choose a further test with the least cost for every percentage of accuracy gain. However, at some steps, adding a test can only achieve a small accuracy increment. For example, from model I+C+WC to model I+C+WA+WC, the accuracy increases from 85.7% to 88.1% by 2.4% only. Clinicians may need to make decision on whether such an increment is practically useful to their needs.

The decision support system software based on all models is developed to support the decision making in schizophrenia diagnosis. The system will automatically choose an appropriate model depending on the available case data and classify a case as either patient or normal. Suggestions on what tests should be performed in order to get more accurate classification results will also be given.

Unlike some existing schizophrenia decision support systems, (Razzouk, et al., 2006) and (Yana, et al., 1997), that use patient clinical information only, our decision support system makes use of both neuroinformatics features and neuroimaging features. Since the criteria used in our decision support system are

objective and quantifiable, the classification result will be more reproducible and reliable.

7.1.6 Summary

In summary, schizophrenia classification models can be constructed using objective and quantifiable criteria from neuroinformatics and neuroimaging data. The most comprehensive model can achieve an accuracy of 89.3%. A decision support system based on these models can provide additional objective evidence to clinicians and augment the current diagnostic procedures.

Despite the unique combination of neuroinformatics and neuroimaging data, our models and decision support system are still tentative and limited due to the relatively small sample size and types of data. For example, Type I Error, or False Positive Rate, is still at a noticeable 20% level even for the most comprehensive model including all 8 features. Further refinements need to consider by using more extensive clinical information, other types of neuroimaging data and biological information such as genetic data.

7.2 Discussion

7.2.1 Uniqueness

The number of samples collected in this study is restricted by the budget because the data collection is very costly. For example, MRI and DTI scan costs $457 per

person. For a complete set of all tests (four neurocognitive tests plus neuroimaging), the total cost is $1,497 per person if he/she is charged according to the private rate as discussed at section 3.4. This has not included the administrative cost yet. The actual total cost of data acquisition for this study is about $150,000.

Despite that, a total of 156 study subjects are recruited. Among them, all subjects are scanned to acquire the MRI and DTI images, and only 89 to 95 subjects undergo various neurocognitive tests (see Table 3.6). Finally 84 subjects who have completed all neurocognitive tests and neuroimaging scans are used in our study of schizophrenia.

To our best knowledge, this is the first project that tries to build schizophrenia models based on unique combination of neuroimaging and neuroinformatics data. Although DTI has been widely used to examine the brain white matter abnormalities for schizophrenia, and many qualitative results have shown brain white matter changes in schizophrenia, the quantitative relationship between the FA values in brain anatomical structures and schizophrenia has not been revealed yet.

Furthermore, no other study has been reported to solve the non-quantifiable diagnosis criteria problem in the current standard diagnosis procedures. DSM-IV was published in 1994, and the next version DSM-V is still in the preparation stage, with its publication date being postponed to May 2013 as announced by American Psychiatric Association ("News Release," 2009). Along the way, our work shows interesting results and promising directions in the attempt of improving diagnosis

accuracy by using objective and quantitative neuroimaging and neuroinformatics features.

7.2.2 Model Accuracies

By adding a new test result to an existing model, a new model is generated. Usually the new model will have a higher accuracy. For example, model C+WC has an accuracy of 75%. After adding in WAIS test, the new model C+WA+WC has an accuracy of 84.5%; an accuracy gain of 9.5% is achieved.

However, the following exceptions are observed:

Model I+C+WA has an accuracy of 88.1%. After adding in another test result (WCST), it becomes model I+C+WA+WC, and the accuracy remains unchanged (88.1%). No accuracy gain is achieved. After adding RPM test, the model I+C+R+WA even has a lower accuracy (86.9%): a negative accuracy gain is incurred (-1.2%). New models accuracies do not increase as expected after new tests are added. However, the differences between the models with additional tests (I+C+WA+WC, I+C+R+WA) and the existing model (I+C+WA) are small: 0% and -1.2%, which convert to 0 or 1 case difference for the 84 training datasets. This fluctuation may be caused by the small number of validation cases. In this situations, RPM test's additional contribution to the classification accuracy is already small; when the number of cases is small, the irregularity of the sample data may affect a small number (e.g., 1 or 2 cases) of classification results, and cause the decreased accuracy.

In another case, model I+C+R has an accuracy of 88.1%. After adding in WAIS test, the new model I+C+R+WA has an accuracy of 86.9%, a negative accuracy gain is incurred (-1.2%). After adding WCST test, the new model I+C+R+WC has an accuracy of 85.7%, with a negative accuracy gain of -2.4%. These convert to 1 or 2 cases difference in all the 84 training cases. The fluctuation may also be caused by the reasons as discussed above.

7.2.3 Validation

Each model is tested by using the 10-fold cross-validation when it is constructed. The accuracy of the model may be more optimized than the actual condition. External validation is required to test these models. In future, preferably the testing can be done in the hospital environment.

However, before large scale external validation can be done, we can do some quasi-external validation. We call the validation as quasi-external, because the cases we want to evaluate are not used in model construction, but they have been collected for this study already.

Remember that in this study, we have recruited 156 cases. Clinical information for all cases is collected; neuroimaging (sMRI and DTI) for all cases are also acquired. But not all of them have completed all four neurocognitive tests. Specifically, 93 study subjects have completed RPM test, 95 for WAIS test, 89 for CPT test and 93 for WCST test (Table 3.6). 84 subjects have completed all four tests, and they are

used to build schizophrenia models. The remaining 72 cases are not used in model construction. We can use part of these cases that have completed at least one neurocognitive test for validation purpose.

We validate model C using all 72 cases. As we mentioned before, this model classifies all cases as patient, hence the accuracy is the same as the prevalent patient rate, and there is no accuracy gain.

There are also 72 study subjects that completed the neuroimaging scan; they are used to validate model I+C. Among them, 46 are correctly classified, and so the accuracy is 63.9%.

For the rest of the models, there are only 6 to 12 cases for validation. Their accuracies range from 55.6% to 100%. The results are summarized in Table 7.2 and Figure 7.7 (Accuracy) and Figure 7.8 (Type I and Type II Error). Since the numbers of cases used in validation are small, the validation results are not very reliable. Large scale external validation is required before the decision support system can be applied in clinical practice.

Table 7.2 Summary of validation results

Model→	**C**	**C+R**	**C+WA**	**C+WC**	**C+R+WA**	**C+R+WC**	**C+WA+WC**	**C+R+WA+WC**
Nr	72	9	12	9	9	6	9	6
Cor	33	9	11	5	9	6	8	6
Incor	39	0	1	4	0	0	1	0
Accuracy	**45.8%**	**100.0%**	**91.7%**	**55.6%**	**100.0%**	**100.0%**	**88.9%**	**100.0%**
Err	54.2%	0.0%	8.3%	44.4%	0.0%	0.0%	11.1%	0.0%
Sen	100.0%	100.0%	88.9%	33.3%	100.0%	100.0%	83.3%	100.0%
Type I	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Spe	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Type II	0.0%	0.0%	11.1%	66.7%	0.0%	0.0%	16.7%	0.0%
Pt Nr	33	6	9	6	6	3	6	3
Ctrl Nr	39	3	3	3	3	3	3	3
Model→	**I+C**	**I+C+R**	**I+C+WA**	**I+C+WC**	**I+C+R+WA**	**I+C+R+WC**	**I+C+WA+WC**	**I+C+R+WA+WC**
Nr	72	9	12	9	9	6	9	6
Cor	46	8	10	5	8	5	7	5
Incor	26	1	2	4	1	1	2	1
Accuracy	**63.9%**	**88.9%**	**83.3%**	**55.6%**	**88.9%**	**83.3%**	**77.8%**	**83.3%**
Err	36.1%	11.1%	16.7%	44.4%	11.1%	16.7%	22.2%	16.7%
Sen	69.7%	100.0%	88.9%	50.0%	100.0%	100.0%	83.3%	66.7%
Type I	41.0%	33.3%	33.3%	33.3%	33.3%	33.3%	33.3%	0.0%
Spe	59.0%	66.7%	66.7%	66.7%	66.7%	66.7%	66.7%	100.0%
Type II	30.3%	0.0%	11.1%	50.0%	0.0%	0.0%	16.7%	33.3%
Pt Nr	33	6	9	6	6	3	6	3
Ctrl Nr	39	3	3	3	3	3	3	3

Abbreviations: Nr, Total Number of Instances; Cor, Correctly Classified Instances; Incor, Incorrectly Classified Instances; Acc, Accuracy; Err, Error Rate; Sen, Sensitivity; TPR, True Positive Rate (Patient classified as patient); Type I, Type I Error Rate (Control classified as patient); Spe, Specificity; TNR, True Negative Rate (Control classified as control); Type II, Type II Error Rate (Patient classified as control); Pt Nr, Patient Number; Ctrl Nr, Control Number; I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

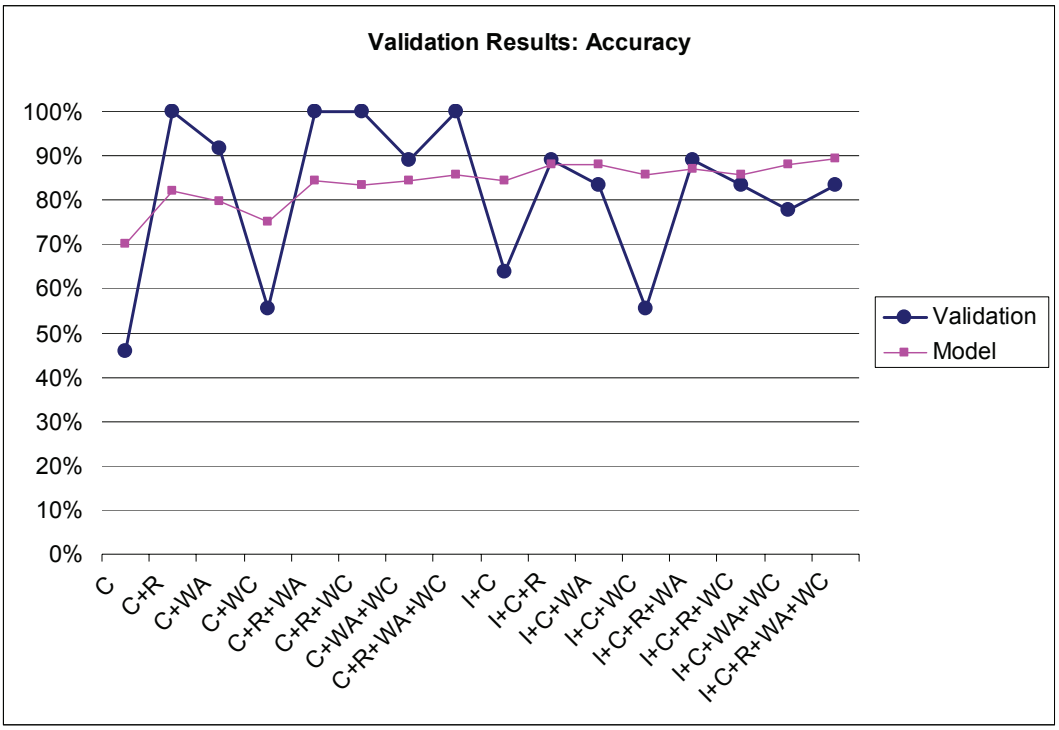

Figure 7.7 Validation results: accuracy
(blue: validation accuracy, pink: model accuracy)

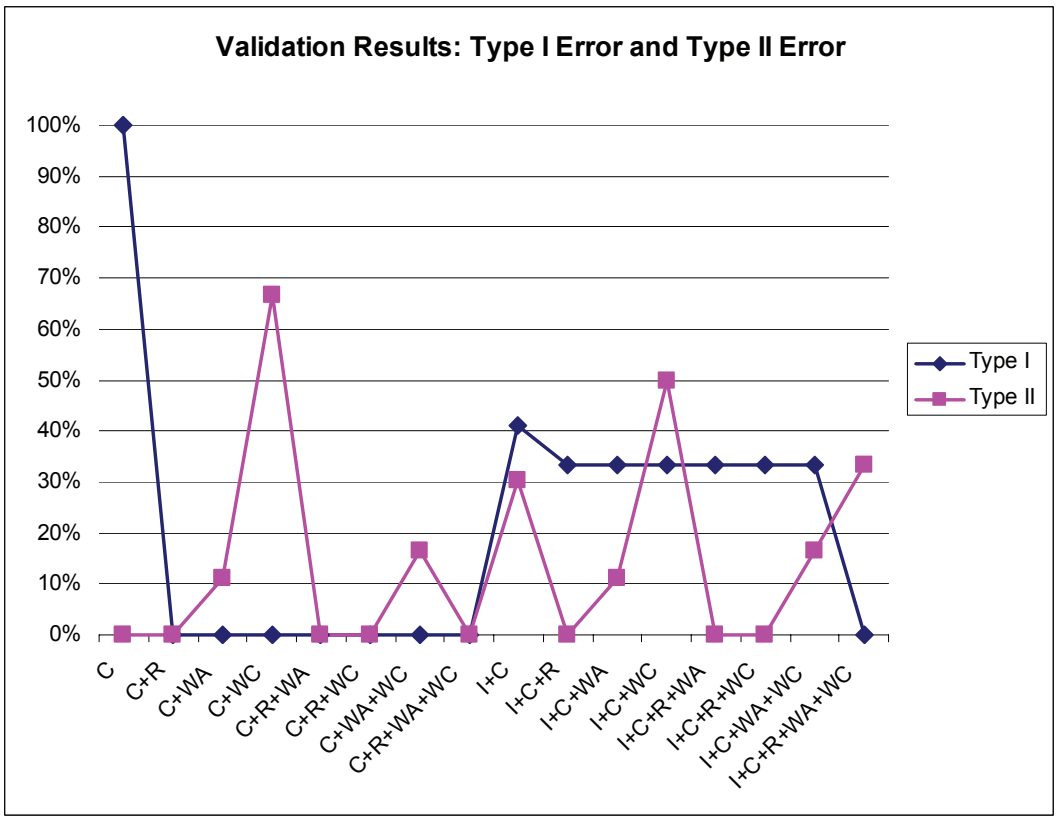

Figure 7.8 Validation results: Type I and Type II error

However, we can estimate the overall accuracy by combining all models as described previously in chapter 6.

$Acc_{overall} = Acc_{16} = 81.7\%$

We notice that the accuracy is 81.7%, which is close to the result of overall accuracy of all models, Acc_{16} (83.8%) as described in chapter 6.

Model C takes into account of clinical information only (family history). As we have pointed out previously, this model classifies all cases as patients, and its type I error is 100%. Hence it shall not be applied in practice. If we exclude this model from decision support system (a more realistic situation because diagnosis should not solely depend on family history), the mean accuracy of all the other models becomes 84.1%, which is also close to the result of Acc_{15} (84.8%) as described in chapter 6.

$Acc_{overall} = Acc_{15} = 84.1\%$

In summary, we validate the decision support system using limited external cases. The overall classification accuracies of the decision support system are 81.7% and 84.1%, for 16 models and 15 models respectively, which are close to the results in chapter 6.

7.2.4 Comparison with Other Decision Support Systems for Diagnosis

We compare our decision support system with the other two schizophrenia diagnosis decision support systems that we have reviewed in chapter 2. The results are summarized in Table 7.3.

We find that: 1) in terms of targeted diseases, our solution tries to diagnose schizophrenia from normal control; Razzouk's solution is used for differentiating schizophrenia from a similar disorder, schizophreniform; and Yana's solution is for the diagnosis of schizophrenia, mood disorders and neurosis. 2) Razzouk's and Yana's solutions rely on symptoms as diagnosis criteria which are subjective and not quantifiable, but our solution uses objective and quantifiable neurocognitive and neuroimaging tests results; 3) Numbers of cases used for building/testing of our solution and Yana's solution are close to each other (84 and 100), and that of Razzouk's solution is only 38; 4) Though the accuracy range of the three solutions overlap, our solution is at the high end. The accuracy of our most comprehensive model (including 8 features) is especially the highest (89.3%) among all solutions.

Overall, our decision support system is the only one based on objective criteria and achieves the highest diagnosis accuracy.

Table 7.3 Comparison of decision support systems for schizophrenia diagnosis

	Targeted Diseases	Criteria Used	Number of Cases	Diagnosis Accuracy
Our solution	Schizophrenia	Family history, neurocognitive tests and neuroimaging	84	**75%-89.3%** Average: 83.8%
Razzouk's solution	Differentiate schizophrenia from schizophreniform disease	Symptoms	38	**66%-82%**
Yana's solution	Schizophrenia, mood disorder, neurosis	80 Yes/no questions including 32 symptoms	100	**73.3%** and **77.3%** for schizophrenia by two different models

7.2.5 Alternative Forms of Models

We also attempt to construct schizophrenia models using other algorithms, such as Alternating Decision Tree (Freund & Mason, 1999) and Logistic Regression (Cessie & Houwelingen, 1992). The results for the most comprehensive model (model I+C+R+WA+WC) using the 2 alternative algorithms are listed and compared with Bayesian Network model in Table 7.4. We find that their accuracies are similar (83.3%) and almost as good as that of the Bayesian Network model (89.3%).

Table 7.4 Models using different algorithms

Model→	Alternating Decision Tree (I+C+R+WA+WC)	Logistic Regression (I+C+R+WA+WC)	Bayesian Network (I+C+R+WA+WC)
Nr	84	84	84
Cor	70	70	75
Incor	14	14	9
Accuracy	**83.3%**	**83.3%**	**89.3%**
Err	16.7%	16.7%	10.7%
Sen	86.4%	86.4%	93.2%
Type I	24.0%	24.0%	20.0%
Spe	76.0%	76.0%	80.0%
Type II	13.6%	13.6%	6.8%

Abbreviations: Nr, Total Number of Instances; Cor, Correctly Classified Instances; Incor, Incorrectly Classified Instances; Acc, Accuracy; Err, Error Rate; Sen, TPR, True Positive Rate (Patient classified as patient); Type I, Type I Error Rate (Control classified as patient); Spe, TNR, True Negative Rate (Control classified as control); Type II, Type II Error Rate (Patient classified as control); I, Imaging; C, Clinical Data; R, RPM Test; WA, WAIS Test; WC, WCST Test

The most comprehensive model based on all 8 significant features in the form of Alternating Decision Tree is illustrated in Figure 7.9. This decision tree has 4 levels, and contains 10 decision nodes (represented by ovals) and 21 prediction nodes (represented by rectangles). A case follows all paths for which all decision nodes are true. The numeric values inside all prediction nodes it traversed are added up together. When it reaches the bottom of the decision tree, if the total of the numeric values along the traversal path is negative, this case is classified as a patient, otherwise, a control.

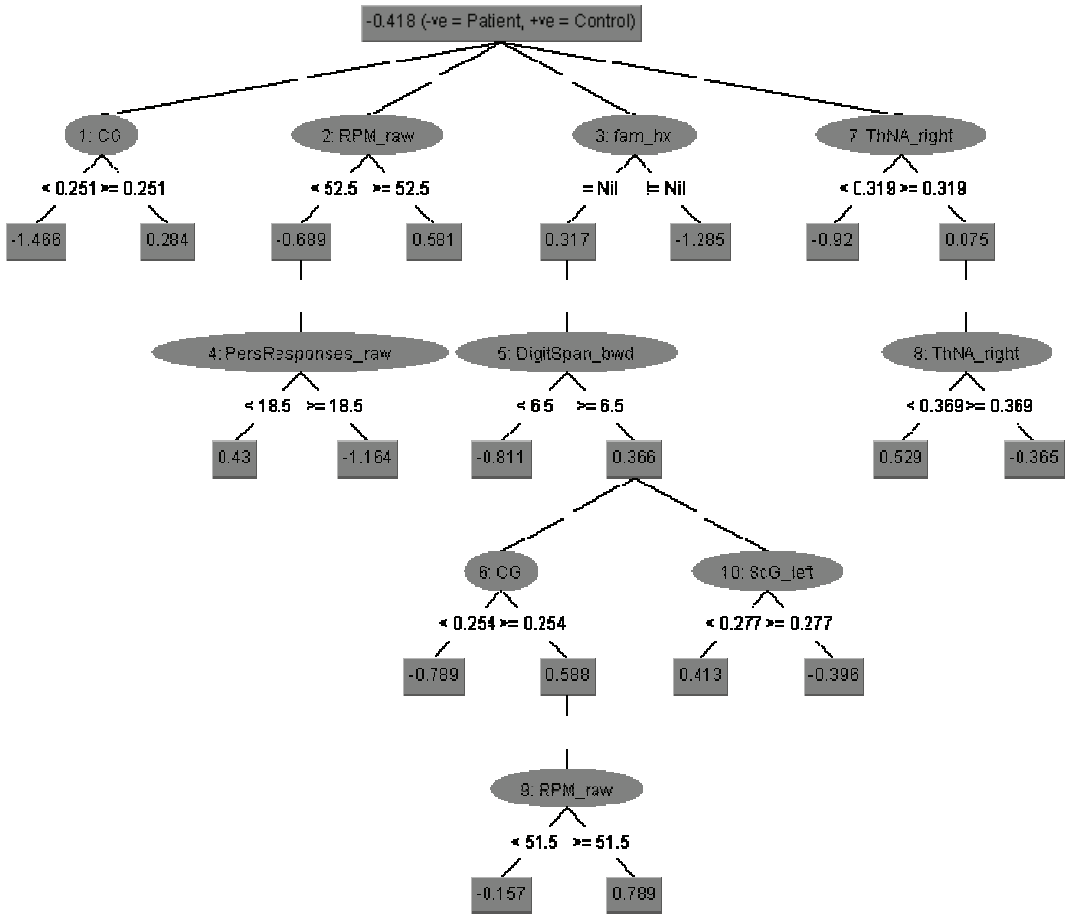

Figure 7.9 Alternating decision tree model on all significant features

7.2.6 Decision Support

The purpose of our decision support system is not meant to replace existing diagnosis procedures. At the current stage, it can only be used in providing additional support to the clinicians. The models underlying the decision support system need to be further verified.

The cost of each neurocognitive test and the neuroimaging may vary from time to time, and from hospital to hospital. The suggestion on the value of tests by the decision support system depends on the particular actual cost of each hospital.

They can be adjusted as the system parameters. In addition, both model structure and model parameters can be enhanced when we have new training datasets.

Currently in our decision support system uses 16 different models. The complexity of the model repository is at $O(2^n)$, where n is the number of tests. The searching for suitable model for classification and suggested further tests is linear, which means it accesses the model repository $O(2^n)$ times. With current computer storage capacity and computing speed, when n is small, the computing performance of the decision support system is still acceptable. However, at the model construction stage, all models are constructed manually. When n is large, constructing of models becomes difficult. Some automated methods for dynamic Bayesian Network model construction need to be applied (Poh, Fehling, & Horvitz, 1994; Xiang & Poh, 2005; S. S. Xu & Poh, 2002).

7.2.7 Limitations of the Image Processing Algorithm

There are some limitations in our ROI selection method. Firstly the resolution of current atlas is not high enough, e.g., typical slice distance is 2 to 5 mm in axial orientation. This restricts the precision of the atlas based ROI placement. This limitation can be overcome by a high resolution atlas which is currently under development.

Secondly, the Fast Talairach-Transformation (FTT) method divides the whole brain into 24 small cuboids. Within each cuboid, linear transformation is used to

warp the atlas. More accurate registration method such as non-linear warping (M. Xu & Nowinski, 2001) should be used in order to get higher registration accuracy.

Furthermore, multiple atlases such as Schaltenbrand and Wahren electronic brain atlas (W. L. Nowinski, et al., 1997; W. L. Nowinski, Liu, & Thirunavuukarasuu, 2006; Schaltenbrand & Wahren, 1977) co-registered with the Talairach atlas and the 3-dimensional detailed brain atlas for structures, vasculatures and tracts (W.L. Nowinski, et al., 2009) can be used for automatically identifying more deep brain structures and brain connections at higher resolution.

7.2.8 Limitations of Study Samples

Among 156 recruited study subjects, only 84 are selected as training datasets. 72 are not used because they do not fulfill the requirements for completion of all neurocognitive tests. If we can have more study subjects, we will be able to achieve better results, in terms of higher precision of model parameters.

Furthermore, our sample data are collected in Singapore. The ethnics are mainly Chinese. This may restrict the generality of our findings in wider geographical and ethnical distribution, though our approach is general and applicable in different distributions.

7.2.9 Future Work Direction

Model Maintenance: Our models are constructed from training datasets consisting of 84 cases. Eight features are selected in total. The ratio of case to feature is about 10 to 1. In future, new datasets need to be collected in order to enhance the Bayesian model structures as well as parameters. And the decision support flow charts can be re-organized automatically according to the new parameters of each model.

Other Imaging Features: We also found the first episode schizophrenia patients had reduced brain white volume in right temporal-occipital region compared to normal controls (Chan, et al., 2010) in another sub-project of this study. In addition to the FA values, the volumetric changes of brain structures shall also be included in the schizophrenia modeling.

Genomic Data: Genetic studies have attempted to identify the genes that are related to certain disease. According to a review report (Lakhan, 2006), genes related to schizophrenia have been found in several chromosomal regions. Since more than 500 genes that have been reported to be associated with schizophrenia, effective feature selection methods such as (Fan, Poh, & Zhou, 2009) and factor grouping technology (Li & Leong, 2005) should be explored to improve the performance of Bayesian Network Models. For example, (J. Sun, Kuo, Riley, Kendler, & Zhao, 2008) uses a combined odds ratio method to ranks the genes and generates a list of highly related genes, including Disrupted-in-Schizophrenia 1

(DISC1), Dystrobrevin-Binding Protein 1 (DTNBP1), Catechol-O-methyl Transferase (COMT), etc.

As such, genomic data shall also be combined with existing source of data in the constructing of more complete models in order to achieve better understanding of the pathophysiology and biological markers related to schizophrenia.

Subtypes and Other Mental Diseases: We have built binary models for classifications of schizophrenia patients and healthy controls. The approach described in this thesis may also be applied to the model construction and decision support in classification of subtypes of schizophrenia, as well as other mental diseases (such as schizoaffective disorder, schizophreniform disorder, bipolar disorder, and unipolar depression) where neurocognitive tests and neuroimaging test are used. Furthermore, classifying of different mental diseases can be combined together to form multi-category classification models and decision support systems.

References

Abbruzzese, M., Bellodi, L., Ferri, S., & Scarone, S. (1995). Frontal lobe dysfunction in schizophrenia and obsessive-compulsive disorder: a neuropsychological study. *Brain Cogn, 27*(2), 202-212.

Agartz, I., Andersson, J. L., & Skare, S. (2001). Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. *Neuroreport, 12*(10), 2251-2254.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci, 22*(3), 425-444; discussion 444-489.

Aggleton, J. P., Desimone, R., & Mishkin, M. (1986). The origin, course, and termination of the hippocampothalamic projections in the macaque. *J Comp Neurol, 243*(3), 409-421.

Amador, X. F., Strauss, D. H., Yale, S. A., Flaum, M. M., Endicott, J., & Gorman, J. M. (1993). Assessment of insight in psychosis. *Am J Psychiatry, 150*(6), 873-879.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision). (2000). (4th ed.): American Psychiatric Publishing, Inc.

Andreasen, N. C. (1999). A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. *Arch Gen Psychiatry, 56*(9), 781-787.

Andreasen, N. C. (2000). Schizophrenia: the fundamental questions. *Brain Res Brain Res Rev, 31*(2-3), 106-112.

Ardekani, B. A., Nierenberg, J., Hoptman, M. J., Javitt, D. C., & Lim, K. O. (2003). MRI study of white matter diffusion anisotropy in schizophrenia. *Neuroreport, 14*(16), 2025-2029.

Arrindell, W. A., & Ende, J. v. d. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement, 9*, 165-178.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *Neuroimage, 11*(6 Pt 1), 805-821.

Barch, D. M., Csernansky, J. G., Conturo, T., & Snyder, A. Z. (2002). Working and long-term memory deficits in schizophrenia: is there a common prefrontal mechanism? *J Abnorm Psychol, 111*(3), 478-494.

Bartlett, James, E. I., Kotrlik, Joe, W., Higgins, & Chadwick, C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal, 19*(1), 43-50.

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed, 15*(7-8), 435-455.

Begre, S., Federspiel, A., Kiefer, C., Schroth, G., Dierks, T., & Strik, W. K. (2003). Reduced hippocampal anisotropy related to anteriorization of alpha EEG in schizophrenia. *Neuroreport, 14*(5), 739-742.

Belin, C., Faure, S., & Mayer, E. (2008). Hemispheric specialisation versus inter-hemispheric communication. *Rev Neurol (Paris), 164 Suppl 3*, S148-153.

Bentall, R. (1992). *Reconstructing Schizophrenia*. London: Routledge.

Bishop, C. (1996). *Neural Networks for Pattern Recognition*: Oxford: Oxford University Press.

Bouckaert, R. R. (2004). Bayesian Network Classifiers in Weka. Retrieved from http://weka.sourceforge.net/manuals/weka.bn.pdf

Boyle, M. (2002). *Schizophrenia: A Scientific Delusion*. London: Routledge.

Broman, J. (1994). Neurotransmitters in subcortical somatosensory pathways. *Anat Embryol (Berl), 189*(3), 181-214.

Buchsbaum, M. S., Friedman, J., Buchsbaum, B. R., Chu, K. W., Hazlett, E. A., Newmark, R., et al. (2006). Diffusion tensor imaging in schizophrenia. *Biol Psychiatry, 60*(11), 1181-1187.

Buchsbaum, M. S., Tang, C. Y., Peled, S., Gudbjartsson, H., Lu, D., Hazlett, E. A., et al. (1998). MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. *Neuroreport, 9*(3), 425-430.

Burns, J., Job, D., Bastin, M. E., Whalley, H., Macgillivray, T., Johnstone, E. C., et al. (2003). Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *Br J Psychiatry, 182*, 439-443.

Caldwell, C. B., & Gottesman, II. (1990). Schizophrenics kill themselves too: a review of risk factors for suicide. *Schizophr Bull, 16*(4), 571-589.

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proc Natl Acad Sci U S A, 97*(4), 1944-1948.

Cessie, S. l., & Houwelingen, J. C. v. (1992). Ridge Estimators in Logistic Regression. *Applied Statistics, 41*(1), 191-201.

Chan, W. Y., Yang, G. L., Chia, M. Y., Lau, I. Y., Sitoh, Y. Y., Nowinski, W. L., et al. (2010). White matter abnormalities in first-episode schizophrenia: A combined structural MRI and DTI study. *Schizophr Res, Accepted for publication in 2009*.

Cheniaux, E., Landeira-Fernandez, J., & Versiani, M. (2009). The diagnoses of schizophrenia, schizoaffective disorder, bipolar disorder and unipolar depression: interrater reliability and congruence between DSM-IV and ICD-10. *Psychopathology, 42*(5), 293-298.

Cheung, V., Cheung, C., McAlonan, G. M., Deng, Y., Wong, J. G., Yip, L., et al. (2008). A diffusion tensor imaging study of structural dysconnectivity in never-medicated, first-episode schizophrenia. *Psychol Med, 38*(6), 877-885.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research, 3*, 507-554.

Cipolotti, L., Husain, M., Crinion, J., Bird, C. M., Khan, S. S., Losseff, N., et al. (2008). The role of the thalamus in amnesia: a tractography, high-resolution MRI and neuropsychological study. *Neuropsychologia, 46*(11), 2745-2758.

Conners, C. (2000). *Conner's Continuous Performance Test II for Windows (CPT II)*. Toronto.: Multi Health Systems, .

Cooper, G., & Herskovitz, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning, 9*, 330-347.

Cozman, F. G. (1998). The Interchange Format for Bayesian Networks from http://www.cs.cmu.edu/afs/cs/user/fgcozman/www/Research/InterchangeFormat/

David, A. S., Malmberg, A., Brandt, L., Allebeck, P., & Lewis, G. (1997). IQ and risk for schizophrenia: a population-based cohort study. *Psychol Med, 27*(6), 1311-1323.

Davidson, L., & McGlashan, T. H. (1997). The varied outcomes of schizophrenia. *Can J Psychiatry, 42*(1), 34-43.

Doty, R. W. (1989). Schizophrenia: a disease of interhemispheric processes at forebrain and brainstem levels? *Behav Brain Res, 34*(1-2), 1-33.

Edlund, M. J. (1986). Causal models in psychiatric research. *Br J Psychiatry, 148*, 713-717.

Everitt, B. S. (1975). Multivariate analysis: the need for data, and other problems. *Br J Psychiatry, 126*, 237-240.

Fan, L., Poh, K. L., & Zhou, P. (2009). A sequential feature extraction approach for naive Bayes classification of microarray data. *Expert Systems with Applications, 36*(6), 9919.

Filbey, F. M., Russell, T., Morris, R. G., Murray, R. M., & McDonald, C. (2008). Functional magnetic resonance imaging (fMRI) of attention processes in presumed obligate carriers of schizophrenia: preliminary findings. *Ann Gen Psychiatry, 7*, 18.

Foong, J., Maier, M., Clark, C. A., Barker, G. J., Miller, D. H., & Ron, M. A. (2000). Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *J Neurol Neurosurg Psychiatry, 68*(2), 242-244.

Foong, J., Symms, M. R., Barker, G. J., Maier, M., Miller, D. H., & Ron, M. A. (2002). Investigating regional white matter in schizophrenia using diffusion tensor imaging. *Neuroreport, 13*(3), 333-336.

Freund, Y., & Mason, L. (1999). *The alternating decision tree learning algorithm.* Paper presented at the Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia.

Giuliani, N. R., Calhoun, V. D., Pearlson, G. D., Francis, A., & Buchanan, R. W. (2005). Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophr Res, 74*(2-3), 135-147.

Green, M. F., Lee, J., Cohen, M. S., Engel, S. A., Korb, A. S., Nuechterlein, K. H., et al. (2009). Functional neuroanatomy of visual masking deficits in schizophrenia. *Arch Gen Psychiatry, 66*(12), 1295-1303.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389-422.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1995). *Multivariate Data Analysis* (4th ed.). Saddle River, NJ: Prentice Hall College Div.

Hall, J., Harris, J. M., McKirdy, J. W., Johnstone, E. C., & Lawrie, S. M. (2007). Emotional memory in schizophrenia. *Neuropsychologia, 45*(6), 1152-1159.

Hall, M. A. (1998). *Correlation-based Feature Subset Selection for Machine Learning*. Hamilton, New Zealand. .

Hansen, K., Lancon, C., & Toumi, M. (2006). Pharmacoeconomic modelling in schizophrenia. Trap or support for decision makers? *Eur J Health Econ, 7*(1), 19-29.

Harding, A., Halliday, G., Caine, D., & Kril, J. (2000). Degeneration of anterior thalamic nuclei differentiates alcoholics with amnesia. *Brain, 123 (Pt 1)*, 141-154.

Harding, C. M., Brooks, G. W., Ashikaga, T., Strauss, J. S., & Breier, A. (1987). The Vermont longitudinal study of persons with severe mental illness, II: Long-term outcome of subjects who retrospectively met DSM-III criteria for schizophrenia. *Am J Psychiatry, 144*(6), 727-735.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.): Prentice Hall PTR.

Heaton, R., Chelune, G., Talley, J., Kay, G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test, Manual*. Odessa, Florida: Psychological Assessment Resources.

Herbener, E. S. (2008). Emotional memory in schizophrenia. *Schizophr Bull, 34*(5), 875-887.

Herz MI, M. S. (2002). *Schizophrenia comprehensive treatment and management*: Lippincott Williams & Wilkins.

Hiller, W., Dichtl, G., Hecht, H., Hundt, W., & von Zerssen, D. (1994). Testing the comparability of psychiatric diagnoses in ICD-10 and DSM-III-R. *Psychopathology, 27*(1-2), 19-28.

Hoffman, R. E., & McGlashan, T. H. (2001). Neural network models of schizophrenia. *Neuroscientist, 7*(5), 441-454.

Holden, J. M., Forno, G., Itil, T., & Hsu, W. (1973). Echoencephalographic patterns in chronic schizophrenia (relationship to therapy resistance). *Biol Psychiatry, 6*(2), 129-141.

Holmes, G., Donkin, A., & Witten, I. H. (1994). Weka: A machine learning workbench. *Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia*.

Honea, R., Crow, T. J., Passingham, D., & Mackay, C. E. (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry, 162*(12), 2233-2245.

Hoptman, M. J., Nierenberg, J., Bertisch, H. C., Catalano, D., Ardekani, B. A., Branch, C. A., et al. (2008). A DTI study of white matter microstructure in individuals at high genetic risk for schizophrenia. *Schizophr Res, 106*(2-3), 115-124.

Hoptman, M. J., Volavka, J., Johnson, G., Weiss, E., Bilder, R. M., & Lim, K. O. (2002). Frontal white matter microstructure, aggression, and impulsivity in men with schizophrenia: a preliminary study. *Biol Psychiatry, 52*(1), 9-14.

Howard, R. (1990). From Inferecne to Relevance to Knowledge. In R. Oliver & J. Smith (Eds.), *Inference Daigrams, Belief Nets and Decision Analysis* (pp. 3-23): John Wiley.

Hubl, D., Koenig, T., Strik, W., Federspiel, A., Kreis, R., Boesch, C., et al. (2004). Pathways that make voices: white matter changes in auditory hallucinations. *Arch Gen Psychiatry, 61*(7), 658-668.

Jacobi, W., & Winkler, H. (1927). Encephalographische studien an chronisch schizophrenen. *Archive fur Psychiatrie unde Nervenkrankheiten, 81*, 299-332.

Jensen, F. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag.

Jiang, H., van Zijl, P. C., Kim, J., Pearlson, G. D., & Mori, S. (2006). DtiStudio: resource program for diffusion tensor computation and fiber bundle tracking. *Comput Methods Programs Biomed, 81*(2), 106-116.

Johnstone, E. C., Crow, T. J., Frith, C. D., Husband, J., & Kreel, L. (1976). Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *Lancet, 2*(7992), 924-926.

Jones, D. K., Catani, M., Pierpaoli, C., Reeves, S. J., Shergill, S. S., O'Sullivan, M., et al. (2006). Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Hum Brain Mapp, 27*(3), 230-238.

Kalus, P., Buri, C., Slotboom, J., Gralla, J., Remonda, L., Dierks, T., et al. (2004). Volumetry and diffusion tensor imaging of hippocampal subregions in schizophrenia. *Neuroreport, 15*(5), 867-871.

Kay, S., Opler, L., & Fiszbein, A. (1986.). *Positive and Negative Syndrome Scale (PANSS) Manual*. North Tonawanda, NY: Multi-Health Systems.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull, 13*(2), 261-276.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2.

Kubicki, M., Westin, C. F., Maier, S. E., Frumin, M., Nestor, P. G., Salisbury, D. F., et al. (2002). Uncinate fasciculus findings in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Am J Psychiatry, 159*(5), 813-820.

Kubicki, M., Westin, C. F., Nestor, P. G., Wible, C. G., Frumin, M., Maier, S. E., et al. (2003). Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Biol Psychiatry, 54*(11), 1171-1180.

Kumra, S., Ashtari, M., Cervellione, K. L., Henderson, I., Kester, H., Roofeh, D., et al. (2005). White matter abnormalities in early-onset schizophrenia: a voxel-based diffusion tensor imaging study. *J Am Acad Child Adolesc Psychiatry, 44*(9), 934-941.

Kumra, S., Ashtari, M., McMeniman, M., Vogel, J., Augustin, R., Becker, D. E., et al. (2004). Reduced frontal white matter integrity in early-onset schizophrenia: a preliminary study. *Biol Psychiatry, 55*(12), 1138-1145.

Kyriakopoulos, M., Vyas, N. S., Barker, G. J., Chitnis, X. A., & Frangou, S. (2008). A diffusion tensor imaging study of white matter in early-onset schizophrenia. *Biol Psychiatry, 63*(5), 519-523.

Lakhan, S. E. (2006). Schizophrenia proteomics: biomarkers on the path to laboratory medicine? *Diagn Pathol, 1*, 11.

Lawrie SM, W. D., Johnstone EC (Ed.). (2004). *Schizophrenia: from neuroimaging to neuroscience*: Oxford University Press.

Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., et al. (2001). Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging, 13*(4), 534-546.

Lezak, M. (1995). *Neuropsychological Assessment* (3rd ed.). New York.: Oxford University Press.

Li, G. L., & Leong, T. Y. (2005). A framework to learn Bayesian Networks from changing, multiple-source biomedical data. . *Proceedings of the 2005 AAAI*

Spring Symposium on Challenges to Decision Support in a Changing World, Stanford University., 66-72.

Lieberman, J. A., Koreen, A. R., Chakos, M., Sheitman, B., Woerner, M., Alvir, J. M., et al. (1996). Factors influencing treatment response and outcome of first-episode schizophrenia: implications for understanding the pathophysiology of schizophrenia. *J Clin Psychiatry, 57 Suppl 9*, 5-9.

Lim, K. O., Hedehus, M., Moseley, M., de Crespigny, A., Sullivan, E. V., & Pfefferbaum, A. (1999). Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. *Arch Gen Psychiatry, 56*(4), 367-374.

Lysaker, P. H., Bell, M. D., Greig, T. C., & Bryson, G. J. (2000). Emotional discomfort and impairments in verbal memory in schizophrenia. *Psychiatry Res, 97*(1), 51-59.

Mahurin, R. K., Velligan, D. I., & Miller, A. L. (1998). Executive-frontal lobe cognitive dysfunction in schizophrenia: a symptom subtype analysis. *Psychiatry Res, 79*(2), 139-149.

Mangalore, R., & Knapp, M. (2007). Cost of schizophrenia in England. *J Ment Health Policy Econ, 10*(1), 23-41.

McGorry, P. D., Mihalopoulos, C., Henry, L., Dakis, J., Jackson, H. J., Flaum, M., et al. (1995). Spurious precision: procedural validity of diagnostic assessment in psychotic disorders. *Am J Psychiatry, 152*(2), 220-223.

McIntosh, A., & Lawrie, S. (2004). Structural magnetic resonance imaging. In S. Lawrie, E. Johnstone & D. Weinberger (Eds.), *Schizophrenia: from neuroimaging to neuroscience* (pp. 32): Oxford University Press.

Milner, A. D. (1982). Simple reaction times to lateralized visual stimuli in a case of callosal agenesis. *Neuropsychologia, 20*(4), 411-419.

Milner, A. D., & Lines, C. R. (1982). Interhemispheric pathways in simple reaction time to lateralized light flash. *Neuropsychologia, 20*(2), 171-179.

Minami, T., Nobuhara, K., Okugawa, G., Takase, K., Yoshida, T., Sawada, S., et al. (2003). Diffusion tensor magnetic resonance imaging of disruption of regional white matter in schizophrenia. *Neuropsychobiology, 47*(3), 141-145.

Mitchell, T. (1997). *Machine Learning*. New York: McGraw-Hill.

Miyamoto, S., LaMantia, A. S., Duncan, G. E., Sullivan, P., Gilmore, J. H., & Lieberman, J. A. (2003). Recent advances in the neurobiology of schizophrenia. *Mol Interv, 3*(1), 27-39.

Moriya, J., Kakeda, S., Abe, O., Goto, N., Yoshimura, R., Hori, H., et al. (2009). Gray and white matter volumetric and diffusion tensor imaging (DTI) analyses in the early stage of first-episode schizophrenia. *Schizophr Res*.

Murray, C., & Lozpe, A. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Boston: Harvard University Press.

Murray, R., Jones, P., Susser, E., Os, J. v., & Cannon, M. (2003). *The epidemiology of schizophrenia*. Cambridge: Cambridge University Press.

Neapolitan, R. (2004). *Learning Bayesian Networks*: Prentice Hall.

News Release. (2009). Retrieved January, 2010, from http://www.psych.org/MainMenu/Newsroom/NewsReleases/2009NewsReleases/DSM-5-Publication-Date-Moved-.aspx

Nowinski, W. L. (2005). The cerefy brain atlases: continuous enhancement of the electronic talairach-tournoux brain atlas. *Neuroinformatics, 3*(4), 293-300.

Nowinski, W. L., Chua, B. C., Qian, G. Y., Volkau, I., Knopp, M., Runge, V. M., et al. (2009). The human brain in 1,002 pieces: A comprehensive and interactive 3d atlas of structure, vasculature, and connections from multiple 3T and 7T acquisitions. [Abstract]. *Processing, RSNA 95th Scientific Assembly and Annual Meeting*, 1023.

Nowinski, W. L., Fang, A., Nguyen, B. T., Raphel, J. K., Jagannathan, L., Raghavan, R., et al. (1997). Multiple brain atlas database and atlas-based neuroimaging system. *Comput Aided Surg, 2*(1), 42-66.

Nowinski, W. L., Liu, J., & Thirunavuukarasuu, A. (2006). Quantification and visualization of the three-dimensional inconsistency of the subthalamic nucleus in the Schaltenbrand-Wahren brain atlas. *Stereotact Funct Neurosurg, 84*(1), 46-55.

Nowinski, W. L., & Prakash, K. N. (2005). Dorsoventral extension of the talairach transformation and its automatic calculation for magnetic resonance neuroimages. *J Comput Assist Tomogr, 29*(6), 863-879.

Nowinski, W. L., Qian, G., Bhanu Prakash, K. N., Hu, Q., & Aziz, A. (2006). Fast Talairach Transformation for magnetic resonance neuroimages. *J Comput Assist Tomogr, 30*(4), 629-641.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Okugawa, G., Nobuhara, K., Minami, T., Tamagaki, C., Takase, K., Sugimoto, T., et al. (2004). Subtle disruption of the middle cerebellar peduncles in patients with schizophrenia. *Neuropsychobiology, 50*(2), 119-123.

Okugawa, G., Nobuhara, K., Sugimoto, T., & Kinoshita, T. (2005). Diffusion tensor imaging study of the middle cerebellar peduncles in patients with schizophrenia. *Cerebellum, 4*(2), 123-127.

Pae, C. U., Choe, B. Y., Joo, R. H., Lim, H. K., Kim, T. S., Yoo, S. S., et al. (2004). Neuronal dysfunction of the frontal lobe in schizophrenia. *Neuropsychobiology, 50*(3), 211-215.

Palmer, C. S., Brunner, E., Ruiz-Flores, L. G., Paez-Agraz, F., & Revicki, D. A. (2002). A cost-effectiveness clinical decision analysis model for treatment of Schizophrenia. *Arch Med Res, 33*(6), 572-580.

Park, H. J., Westin, C. F., Kubicki, M., Maier, S. E., Niznikiewicz, M., Baer, A., et al. (2004). White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. *Neuroimage, 23*(1), 213-223.

Parker, G. J. (2004). Analysis of MR diffusion weighted images. *Br J Radiol, 77 Spec No 2*, S176-185.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: Morgan Kaufmann.

Peralta, V., & Cuesta, M. J. (2000). Clinical models of schizophrenia: a critical approach to competing conceptions. *Psychopathology, 33*(5), 252-258.

Pierpaoli, C., & Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med, 36*(6), 893-906.

Poh, K. L., Fehling, M. R., & Horvitz, E. J. (1994). Dynamic construction and refinement of utility-based categorization models. *Systems, Man and Cybernetics, IEEE Transactions on, 24*(11), 1653-1663.

Price, G., Bagary, M. S., Cercignani, M., Altmann, D. R., & Ron, M. A. (2005). The corpus callosum in first episode schizophrenia: a diffusion tensor imaging study. *J Neurol Neurosurg Psychiatry, 76*(4), 585-587.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*: Morgan Kaufmann Publishers.

Radomsky, E. D., Haas, G. L., Mann, J. J., & Sweeney, J. A. (1999). Suicidal behavior in patients with schizophrenia and other psychotic disorders. *Am J Psychiatry, 156*(10), 1590-1595.

Raiffa, H. (1968). *Decision tree analysis*. New York: McGraw Hill.

Rametti, G., Junque, C., Falcon, C., Bargallo, N., Catalan, R., Penades, R., et al. (2009). A voxel-based diffusion tensor imaging study of temporal white matter in patients with schizophrenia. *Psychiatry Res, 171*(3), 166-176.

Ramirez Basco, M., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., et al. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry, 157*(10), 1599-1605.

Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive.* MSc Thesis, University of London.

Razzouk, D., Mari, J. J., Shirakawa, I., Wainer, J., & Sigulem, D. (2006). Decision support system for the diagnosis of schizophrenia disorders. *Braz J Med Biol Res, 39*(1), 119-128.

Robinson, D. G., Woerner, M. G., McMeniman, M., Mendelowitz, A., & Bilder, R. M. (2004). Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder. *Am J Psychiatry, 161*(3), 473-479.

Rotarska-Jagiela, A., Oertel-Knoechel, V., DeMartino, F., van de Ven, V., Formisano, E., Roebroeck, A., et al. (2009). Anatomical brain connectivity

and positive symptoms of schizophrenia: a diffusion tensor imaging study. *Psychiatry Res, 174*(1), 9-16.

Sadock BJ, S. V. (Ed.). (2003). *Synopsis of Psychioatry* (9th Edition ed.): Lippincott Williams and Wilkins.

Saunders, R. C., Mishkin, M., & Aggleton, J. P. (2005). Projections from the entorhinal cortex, perirhinal cortex, presubiculum, and parasubiculum to the medial thalamus in macaque monkeys: identifying different pathways using disconnection techniques. *Exp Brain Res, 167*(1), 1-16.

Schaltenbrand, G., & Wahren, W. (1977). *Atlas for stereotaxy of the human brain* (2nd ed.). Stuttgart: Thieme.

Schlosser, R. G., Nenadic, I., Wagner, G., Gullmar, D., von Consbruch, K., Kohler, S., et al. (2007). White matter abnormalities and brain activation in schizophrenia: a combined DTI and fMRI study. *Schizophr Res, 89*(1-3), 1-11.

SCID-I. (2007). SCID-I/P (for DSM-IV-TR). Retrieved January, 2010, from http://www.scid4.org/revisions/pdf/module_c.pdf

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J., 27*, 379-423, 623-656

Sharp, D. J., Scott, S. K., Mehta, M. A., & Wise, R. J. (2006). The neural correlates of declining performance with age: evidence for age-related changes in cognitive control. *Cereb Cortex, 16*(12), 1739-1749.

Shenton, M. E., Dickey, C. C., Frumin, M., & McCarley, R. W. (2001). A review of MRI findings in schizophrenia. *Schizophr Res, 49*(1-2), 1-52.

Shergill, S. S., Kanaan, R. A., Chitnis, X. A., O'Daly, O., Jones, D. K., Frangou, S., et al. (2007). A diffusion tensor imaging study of fasciculi in schizophrenia. *Am J Psychiatry, 164*(3), 467-473.

Sim, K. (2005). Parietal lobe changes in schizophrenia with passivity. Singapore: National Healthcare Group.

Sim, K., Yang, G. L., Loh, D., Poon, L. Y., Sitoh, Y. Y., Verma, S., et al. (2009). White matter abnormalities and neurocognitive deficits associated with the passivity phenomenon in schizophrenia: a diffusion tensor imaging study. *Psychiatry Res, 172*(2), 121-127.

Singapore. (2009). *Singapore Population Trends 2009*. Department of Statistics: Singapore.

Smith, R. C., Calderon, M., Ravichandran, G. K., Largen, J., Vroulis, G., Shvartsburd, A., et al. (1984). Nuclear magnetic resonance in schizophrenia: a preliminary study. *Psychiatry Res, 12*(2), 137-147.

Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120 (Pt 11)*, 1997-2011.

Steel, R. M., Bastin, M. E., McConnell, S., Marshall, I., Cunningham-Owens, D. G., Lawrie, S. M., et al. (2001). Diffusion tensor imaging (DTI) and proton magnetic resonance spectroscopy (1H MRS) in schizophrenic subjects and normal controls. *Psychiatry Res, 106*(3), 161-170.

Steyerberg, E. W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer.

Sun, J., Kuo, P. H., Riley, B. P., Kendler, K. S., & Zhao, Z. (2008). Candidate genes for schizophrenia: a survey of association studies and gene ranking. *Am J Med Genet B Neuropsychiatr Genet, 147B*(7), 1173-1181.

Sun, Z., Wang, F., Cui, L., Breeze, J., Du, X., Wang, X., et al. (2003). Abnormal anterior cingulum in patients with schizophrenia: a diffusion tensor imaging study. *Neuroreport, 14*(14), 1833-1836.

Sussmann, J. E., Lymer, G. K., McKirdy, J., Moorhead, T. W., Munoz Maniega, S., Job, D., et al. (2009). White matter abnormalities in bipolar disorder and schizophrenia detected using diffusion tensor magnetic resonance imaging. *Bipolar Disord, 11*(1), 11-18.

Szeszko, P. R., Ardekani, B. A., Ashtari, M., Malhotra, A. K., Robinson, D. G., Bilder, R. M., et al. (2005). White matter abnormalities in obsessive-compulsive disorder: a diffusion tensor imaging study. *Arch Gen Psychiatry, 62*(7), 782-790.

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotactic Atlas of the Human Brain*. Stuttgart / New York: Georg Thieme Verlag / Thieme Medical Publishers.

Tsuang, M. T., Stone, W. S., & Faraone, S. V. (2000). Toward reformulating the diagnosis of schizophrenia. *Am J Psychiatry, 157*(7), 1041-1050.

Wang, F., Sun, Z., Cui, L., Du, X., Wang, X., Zhang, H., et al. (2004). Anterior cingulum abnormalities in male patients with schizophrenia determined through diffusion tensor imaging. *Am J Psychiatry, 161*(3), 573-575.

Wang, F., Sun, Z., Du, X., Wang, X., Cong, Z., Zhang, H., et al. (2003). A diffusion tensor imaging study of middle and superior cerebellar peduncle in male patients with schizophrenia. *Neurosci Lett, 348*(3), 135-138.

Wciorka, J., Anczewska, M., Bembenek, A., Golebiewska, M., Hochlewicz, A., Nurowska, K., et al. (1998). Psychopathological profile of acute schizophrenic syndromes diagnosed according to ICD-10 and DSM-IV criteria. *Psychiatr Pol, 32*(3), 251-264.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III*. San Antonio, Texas: The Psychological Corporation.

Whittle, S., Allen, N. B., Lubman, D. I., & Yucel, M. (2006). The neurobiological basis of temperament: towards a better understanding of psychopathology. *Neurosci Biobehav Rev, 30*(4), 511-525.

WHOQOL. (1998). *Development of the World Health Organization WHOQOL-BREF quality of life assessment* (1998/06/17 ed. Vol. 28): The WHOQOL Group.

Wikipedia. (2011a). Raven's Progressive Matrices. Retrieved 2011, from http://en.wikipedia.org/wiki/Raven%27s_Progressive_Matrices

Wikipedia. (2011b). Wisconsin card sort. Retrieved 2011, from http://en.wikipedia.org/wiki/Wisconsin_card_sort

Witten, a. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.

Wolkin, A., Choi, S. J., Szilagyi, S., Sanfilipo, M., Rotrosen, J. P., & Lim, K. O. (2003). Inferior frontal white matter anisotropy and negative symptoms of schizophrenia: a diffusion tensor imaging study. *Am J Psychiatry, 160*(3), 572-574.

Wright, I. C., McGuire, P. K., Poline, J. B., Travere, J. M., Murray, R. M., Frith, C. D., et al. (1995). A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *Neuroimage, 2*(4), 244-252.

Wright, I. C., Rabe-Hesketh, S., Woodruff, P. W., David, A. S., Murray, R. M., & Bullmore, E. T. (2000). Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatry, 157*(1), 16-25.

Wu, E. Q., Birnbaum, H. G., Shi, L., Ball, D. E., Kessler, R. C., Moulis, M., et al. (2005). The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry, 66*(9), 1122-1129.

Xiang, Y. P., & Poh, K. L. (2005). *Constructing Bayesian network in a changing world.* Paper presented at the 2005 AAAI Spring Symposium on Challenges to Decision Support in a Changing World.

Xu, M., & Nowinski, W. L. (2001). Talairach-Tournoux brain atlas registration using a metalforming principle-based finite element method. *Med Image Anal, 5*(4), 271-279.

Xu, S. S., & Poh, K. L. (2002). Value of information analysis in dynamic influence diagrams. In A. Abraham & M. Koeppen (Eds.), *Hybrid Information Systems* (pp. 597-610). Heidelberg: Physica Verlag.

Yana, K., Kawachi, K., Iida, K., Okubo, Y., Tohru, M., & Okuyama, F. (1994). *A neural net screening of psychiatric patients.* Paper presented at the Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE.

Yana, K., Mizuta, H., Kawachi, K., Yoshida, H., Iida, K., Okubo, Y., et al. (1997). Decision support for psychiatric diagnosis based on a simple questionnaire. *Methods Inf Med, 36*(4-5), 349-351.

Yang, G. L., Lim, C. C. T., Hou, Z. J., Gan, R. N., Poh, K. L., & Nowinski, W. L. (2005). A practical computer-aided diagnosis system for intracranial hemorrhage detection in acute stroke. [Abstract]. *Proceeding, Radiological Society of North America, 91st Scientific Assembly and Annual Meeting*, 856-857.

Yang, Y., & Pedersen, J. P. (1997). *A Comparative Study on Feature Selection in Text Categorization.* Paper presented at the International Conference on Machine Learning (ICML'97).

Yeterian, E. H., & Pandya, D. N. (1988). Corticothalamic connections of paralimbic regions in the rhesus monkey. *J Comp Neurol, 269*(1), 130-146.

Zammit, S., Allebeck, P., David, A. S., Dalman, C., Hemmingsson, T., Lundberg, I., et al. (2004). A longitudinal study of premorbid IQ Score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. *Arch Gen Psychiatry, 61*(4), 354-360.

Appendix A

Collected Data Items and Descriptions

Data Item	Description
age	Age
age_onset	Age of First Onset of Illness
alcohol	Alcohol Use (past or current)
antichol	Anticholinergics (Type/ Dose)
antidepr	Antidepressants (Type/ Dose)
atypical_antipsy1	Atypical Antipsychotics 1
atypical_antipsy2	Atypical Antipsychotics 2
benzo	Benzodiazepines (Type/ Dose)
BlockDesign_raw	Block design raw score
BlockDesign_scaled	Block design scaled score
broug_sp	Brought By (specify)
brought	Brought By
Categories_percentiles	Categories completed percentiles
Categories_raw	Categories completed raw scores
CDNo	CD no: Compact Disc Number
Comments	Comments on the data entry
Commissions_percentile	Number of Commissions percentile
Commissions_tscore	Number of Commissions t score
ConceptualLevel	Conceptual Level responses
dadmisn	Date of Admission to Ward
depot_antipsy	Depot Antipsychotics
Detect_percentile	Detectability percentile
Detect_tscore	Detectability t score
DigitSpan_bwd	Digit span backward score
DigitSpan_fwd	Digit span forward score

Data Item	Description
DigitSpan_total	Digit span total score
DigitSpan_total_scaled	Digit span total scaled score
dob	Date of Birth
drug_use	Drug Use (past or current)
dsmaxis1	Diagnosis Axis 1 (DSM IV)
dup_yrs	Duration of Untreated Psychosis (in years)
dur_psyc	Duration of Psychiatric Illness (years)
edulevel	Educational Level
ethnic	Ethnicity
Failure_percentiles	Failure to maintain set percentiles
Failure_raw	Failure to maintain set raw scores
fam_hx	Family History of Mental Illness
fam_hxsp	Family History of Mental Illness (specify)
father	Father's Ethnicity
first_ep	First Episode?
gaf_disa	Global Assessment of Functioning Scale - Disability
gaf_symp	Global Assessment of Functioning Scale - Symptoms
gaf_tot	Global Assessment of Functioning Scale Total
handed	Handedness
height	Height
HitRT_percentile	Hit RT percentile
HitRT_StdError_percentile	Hit RT std error percentile
HitRT_StdError_tscore	Hit RT std error t score
HitRT_tscore	Hit RT t score
Learning_percentiles	Learning to learn percentiles
Learning_raw	Learning to learn raw scores
liv_spec	Living Arrangements (specify)
livingar	Living Arrangements
marital	Marital Status
mdstabil	Mood Stabilizers (Type/ Dose)
med_prob	Medical Problems (past or current)
med_spec	Medical Problems (specify)

Data Item	Description
mgfather	Maternal Grandfather's Ethinicity
mgmother	Maternal Grandmother's Ethnicity
mother	Mother's Ethnicity
mri_date	Date of MRI appt
mri_done	MRI Done?
neurocog	Neurocog Done?
no_hosps	Number of Hospitalizations
nohosp12	Number of Hospitalizations in Last 12 Months
NonpersErrors_percentiles	Nonperseverative Errors percentiles
NonpersErrors_raw	Nonperseverative Errors raw scores
NonpersErrors_standard	Nonperseverative Errors standard scores
NonpersErrors_tscores	Nonperseverative Errors t scores
occ_dad	Father's Occupation
occ_mum	Mother's Occupation
occupatn	Occupation
Omissions_percentile	Number of Omissions percentile
Omissions_tscore	Number of Omissions t score
others1	Other Medications 1 (Type/ Dose)
others2	Other Medications 2 (Type/ Dose)
pangps1	PANSS GPS 1
pangps10	PANSS GPS 10
pangps11	PANSS GPS 11
pangps12	PANSS GPS 12
pangps13	PANSS GPS 13
pangps14	PANSS GPS 14
pangps15	PANSS GPS 15
pangps16	PANSS GPS 16
pangps2	PANSS GPS 2
pangps3	PANSS GPS 3
pangps4	PANSS GPS 4
pangps5	PANSS GPS 5
pangps6	PANSS GPS 6

Data Item	Description
pangps7	PANSS GPS 7
pangps8	PANSS GPS 8
pangps9	PANSS GPS 9
panssn1	PANSS Negative 1
panssn2	PANSS Negative 2
panssn3	PANSS Negative 3
panssn4	PANSS Negative 4
panssn5	PANSS Negative 5
panssn6	PANSS Negative 6
panssn7	PANSS Negative 7
panssp1	PANSS Positive 1
panssp2	PANSS Positive 2
panssp3	PANSS Positive 3
panssp4	PANSS Positive 4
panssp5	PANSS Positive 5
panssp6	PANSS Positive 6
panssp7	PANSS Positive 7
Passivity	Case of Passivity?
PercentConceptualLevel_percentiles	% Conceptual Level responses percentiles
PercentConceptualLevel_raw	% Conceptual Level responses raw scores
PercentConceptualLevel_standard	% Conceptual Level responses standard scores
PercentConceptualLevel_tscores	% Conceptual Level responses t scores
PercentErrors_percentiles	% errors percentiles
PercentErrors_raw	% errors raw scores
PercentErrors_standard	% errors standard scores
PercentErrors_tscores	% errors t scores
PercentNonpersErrors_percentiles	% Nonperseverative Errors percentiles
PercentNonpersErrors_raw	% Nonperseverative Errors raw scores
PercentNonpersErrors_standard	% Nonperseverative Errors standard scores
PercentNonpersErrors_tscores	% Nonperseverative Errors t scores
PercentPersErrors_percentiles	% Perseverative Errors percentiles
PercentPersErrors_raw	% Perseverative Errors raw scores

Data Item	Description
PercentPersErrors_standard	% Perseverative Errors standard scores
PercentPersErrors_tscores	% Perseverative Errors t scores
PercentPersResponses_percentiles	% Perseverative Responses percentiles
PercentPersResponses_raw	% Perseverative Responses raw scores
PercentPersResponses_standard	% Perseverative Responses standard scores
PercentPersResponses_tscores	% Perseverative Responses t scores
PersErrors_percentiles	Perseverative Errors percentiles
PersErrors_raw	Perseverative Errors raw scores
PersErrors_standard	Perseverative Errors standard scores
PersErrors_tscores	Perseverative Errors t scores
Persev_percentile	Perseverations percentile
Persev_tscore	Perseverations t score
PersResponses_percentiles	Perseverative Responses
PersResponses_raw	Perseverative Responses raw scores
PersResponses_standard	Perseverative Reponses standard scores
PersResponses_tscores	Perseverative Reponses t scores
pgfather	Paternal Grandfather's Ethnicity
pgmother	Paternal Grandmother's Ethnicity
pt_ctrl	Patient or Control
Response_percentile	Response Style percentile
Response_tscore	Response Style t score
RPM_percentile	Raven's percentile
RPM_raw	Raven's raw score
sapp_tot	Scale for the Assessment of Passivity Phenomena Total Score
	Scale for the Assessment of Passivity Phenomena 1- Made
sapp1	Emotions
sapp1a	Scale for the Assessment of Passivity Phenomena 1a - Time Frame
	Scale for the Assessment of Passivity Phenomena 2 - Made
sapp2	Movements
	Scale for the Assessment of Passivity Phenomena 3 - Made
sapp3	Impulses/ Decisions to Act
sapp4	Scale for the Assessment of Passivity Phenomena 4 - Somatic

Data Item	Description
	Passivity
sex	Sex
SpatialSpan_bwd	Spatial span backward raw score
SpatialSpan_bwd_scaled	Spatial span backward scaled score
SpatialSpan_fwd	Spatial span forward raw score
SpatialSpan_fwd_scaled	Spatial span forward scaled score
SpatialSpan_total	Spatial span total score
study_no	Study Number (corresponds to Excel document)
sumd1	SUMD 1 - Awareness of Mental Disorder
sumd2	SUMD 2 - Awareness of Consequences of Mental Disorder
sumd3	SUMD 3 - Awareness of Effects of Medication
sumd4	SUMD 4 - Awareness of Hallucinatory Experiences
sumd5	SUMD 5 - Awareness of Delusions
sumd6	SUMD 6 - Awareness of Thought Disorder
sumd7	SUMD 7 - Awareness of Flat or Blunt Affect
sumd8	SUMD 8 - Awareness of Anhedonia
sumd9	SUMD 9 - Awareness of Asociality
sur_prob	Surgical Problems (past or current)
sur_spec	Surgical Problems (specify)
tcu_reg	Regularity of Outpatient Attendance in Last 12 Months
Total_correct	Total correct
TotalErrors_percentiles	Total errors percentiles
TotalErrors_raw	Total errors raw scores
TotalErrors_standard	Total errors standard scores
TotalErrors_tscores	Total errors t scores
Trials_administered	Trials administered
Trials_percentiles	Trails to complete 1st category percentile
Trials_raw	Trials to complete 1st category raw scores
typical_antipsy1	Typical Antipsychotics 1
typical_antipsy2	Typical Antipsychotics 2
Variability_percentile	Variability percentile
Variability_tscore	Variability t score

Data Item	Description
weight	Weight
whoqol1	WHO QOL-BREF 1 (World Health Organization Quality of Life)
whoqol10	WHO QOL-BREF 10
whoqol11	WHO QOL-BREF 11
whoqol12	WHO QOL-BREF 12
whoqol13	WHO QOL-BREF 13
whoqol14	WHO QOL-BREF 14
whoqol15	WHO QOL-BREF 15
whoqol16	WHO QOL-BREF 16
whoqol17	WHO QOL-BREF 17
whoqol18	WHO QOL-BREF 18
whoqol19	WHO QOL-BREF 19
whoqol2	WHO QOL-BREF 2
whoqol20	WHO QOL-BREF 20
whoqol21	WHO QOL-BREF 21
whoqol22	WHO QOL-BREF 22
whoqol23	WHO QOL-BREF 23
whoqol24	WHO QOL-BREF 24
whoqol25	WHO QOL-BREF 25
whoqol26	WHO QOL-BREF 26
whoqol3	WHO QOL-BREF 3
whoqol4	WHO QOL-BREF 4
whoqol5	WHO QOL-BREF 5
whoqol6	WHO QOL-BREF 6
whoqol7	WHO QOL-BREF 7
whoqol8	WHO QOL-BREF 8
whoqol9	WHO QOL-BREF 9
yrsedu	Years of Education
yrsedu_dad	Years of Edu Dad
yrsedu_mum	Years of Edu Mum

Appendix B

Brain Anatomical Structures and Full Names

Brain Structure	Full Name
AB	Amygdaloid body
AC	Anterior commissure
AGIPL	Angular gyrus and inferior parietal lobule
BA	Brodmann's area
C	Cortical areas
CA	Cerebral aqueduct
CC	Corpus callosum
CG	Cingulate gyrus
Ci	Cingulum
Cl	Claustrum
CN	Caudate nucleus
CSTF	Corticospinal tract: Face
CSTIL	Corticospinal tract: Inferior limb
CSTMC	Corticospinal tract: Motor cortex
CSTSL	Corticospinal tract: Superior limb
Cu	Cuneus
FG	Fusiform gyrus
Fo	Fornix
FOF	Fronto-occipital fasciculus
GPL	Globus pallidus lateral segment
GPM	Globus pallidus medial segment
HG	Hippocampal gyrus
Hi	Hippocampus
HyD	Hypothalamus: Dorsal nucleus
HyL	Hypothalamus: Lateral nucleus

Brain Structure	Full Name
HyLPO	Hypothalamus: Lateral preoptic nucleus
HyMPO	Hypothalamus: Medial preoptic nucleus
HyP	Hypothalamus: Posterior nucleus
HyPaV	Hypothalamus: Paraventricular nucleus
HyPV	Hypothalamus: Periventricular nucleus
HySO	Hypothalamus: Supra-optic nucleus
HyVM	Hypothalamus: Ventromedial nucleus
IA	Interthalamic adhesion
IFG	Inferior frontal gyrus
ILF	Inferior longitudinal fasciculus
Ins	Insula
IOG	Inferior occipital gyrus
IPL	Inferior parietal lobule
ITG	Inferior temporal gyrus
LG	Lingual gyrus
LGB	Lateral geniculate body
MB	Mamillary body
MeFG	Medial frontal gyrus
MiFG	Middle frontal gyrus
MF	Major forceps
MGB	Medial geniculate body
MiFG	Middle frontal gyrus
MOG	Middle occipital gyrus
MT	Motor tract
MTG	Middle temporal gyrus
NA	Nucleus accumbens
OC	Optic chiasm
OF	Olfactory fasciculus
OG	Occipital gyri
OlT	Olfactory tract
ON	Optic nerve
OpT	Optic tract

Brain Structure	Full Name
ORad	Optic radiations
OrG	Orbital gyri
PB	Pineal body
PC	Posterior commissure
Pcu	Precuneus
PHG	Parahippocampal gyri
PL	Paracentral lobule
PoCG	Postcentral gyrus
PrCG	Precentral gyrus
PrCOG	Precentral opercular gyrus
Pu	Putamen
RNB	Red nucleus: Bottom
RNT	Red nucleus: Top
ScG	Subcallosal gyrus
SFG	Superior frontal gyrus
SG	Straight gyrus
SLF	Superior longitudinal fasciculus
SmG	Supramarginal gyrus
SmGIPL	Supramarginal gyrus and Inferior parietal lobule
SN	Substantia nigra
SOG	Superior occipital gyrus
SPL	Superior parietal lobule
SpR	Suprapineal recess
STG	Superior temporal gyrus
STN	Subthalamic nucleus
T	Tapetum
ThCM	Thalamus: Centromedian nucleus
ThDM	Thalamus: Dorsomedial nucleus
ThLD	Thalamus: Lateral dorsal nucleus
ThLP	Thalamus: Lateral posterior nucleus
ThNA	Thalamus: Anterior nucleus
ThO	Thalamus: Other structures

Brain Structure	Full Name
ThP	Thalamus: Pulvinar nucleus
ThVA	Thalamus: Ventral anterior nucleus
ThVL	Thalamus: Ventral lateral nucleus
ThVPL	Thalamus: Ventral posterolateral nucleus
ThVPM	Thalamus: Ventral posteromedial nucleus
TTG	Transverse temporal gyri
U	Uncus
UF	Uncinate fasciculus
Ven	Ventricle(s)