Learning Models for Psychiatric Diagnosis and Prognosis

by

Animesh Kumar Paul

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

University of Alberta

Abstract

While it is very difficult to diagnose/prognosis psychiatric disorders reliably, especially in early course, such early diagnosis/prognosis is critical for producing an effective treatment. This necessity has motivated many researchers to apply machine learning approaches to high-dimensional neuro-imaging data, to produce models that can produce accurate diagnoses, and prognoses. The machine learning problems are more challenging in the psychiatric field due to their small sample sizes and large feature sets. These challenges motivated us to explore various novel ways of applying machine learning methods to produce models that can predict the diagnosis and prognosis of psychiatric disorders.

We considered the following 3 tasks: (1) We built a classifier that can distinguish healthy subjects versus Obsessive-Compulsive Disorder (OCD) patients. In the learning pipeline, we incorporated prior neurobiological knowledge by using pre-defined brain atlases for feature extraction. The best model (ensemble logistic regression) achieved 80.3% accuracy. We also demonstrated a way to transfer information across psychiatric diagnoses, e.g., schizophrenia (SCZ) to OCD. (2) We next explored ways to apply a machine-learned schizophrenia diagnostic model to identify first degree relatives (FDRs) with high schizotypy scores. Our empirical results found that FDRs of SCZ patients who were classified as schizophrenia by a diagnosis model, which was learned using only SCZ patients and healthy subjects, had significantly higher 'schizotypal personality scores' than those who were not classified as schizophrenia. (3) We addressed the challenges of building a prognostic model for SCZ patients. Here, we dealt

with two types of SCZ patients based on their treatment: antipsychotic medication versus transcranial direct current stimulation (tDCS) treatment. Our success is limited – achieved 63.77% accuracy from the deep transfer learning model – in predicting treatment response for SCZ patients with antipsychotic treatment. On the other side, our proposed prior neurological knowledge (selected brain regions) based method for SCZ patients with tDCS-treatment was able to provide 77.5% accuracy for predicting the treatment response.

Preface

Most parts of chapter 3 are taken from our submission "Prediction of Obsessive Compulsive Disorder: Importance of neurobiology-aided feature design and cross-diagnosis transfer learning," which is under review by *Nature Scientific Reports* journal. Chapter 4 is also taken from our submission "Extending Schizophrenia diagnostic model to predict Schizotypy in first degree relatives", which is under review by *Nature Schizophrenia* journal. We plan to publish some parts of Section 5.2.

Initial raw data were collected from National Institute of Mental Health Neurosciences (NIMHANS), India. Data preprocessing in Chapter 3 (Section 3.3.1) and Chapter 4 (Section 4.2) were performed by NIMHANS research group. Figures 3.6 and 3.7 in Chapter 3, and Figure 4.1 in Chapter 4 were generated and analyzed by Dr. Sunil Kalmady Vasu.

In Section 5.3.2 (Data Preprocessing), I preprocessed around 40% of the fMRI data, and the rest were pre-processed by NIMHANS research group. Moreover, 4D fMRI data preprocessing in Chapter 5 (Section 5.2.2) was done by NIMHANS research group. Anushree Bose (from NIMHANS research group) and I did the literature review to make Table B.2 in Appendix B.

The rest of the works in this dissertation are my original work, under the supervision of Prof. Russell Greiner and Dr. Sunil Kalmady Vasu.

To my life partner- Priyanka Aich
For being support, inspiration, and the love of my life.

To my parents- Amilan Kumar Paul, Kanina Rani Paul
For teaching me the values and attitudes that guide all my decisions.

To my grandparents- Rabindra Nath Paul, Urmena Rani Paul
For being with me, and supporting throughout my life.

For AI to add the most value and for patients and physicians to embrace it, it needs to support, not supplant, the patient-physician relationship ... AI will be most effective when it enhances physicians' ability to focus their full attention on the patient by shifting the physicians' responsibilities away from transactional tasks toward personalized care that lies at the heart of human healing.

– Steven Lin, Clinical assistant professor of medicine, Stanford University, 2019.

AI is only as good as the humans programming it and the system in which it operates. If we are not careful, AI could not make healthcare better, but instead unintentionally exacerbate many of the worst aspects of our current healthcare system.

– Bob Kocher, Adjunct professor, Stanford University School of Medicine, and Zeke Emanuel, chair of the Department of Medical Ethics and Health Policy, University of Pennsylvania, 2019.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Russell Greiner, and co-supervisor Dr. Sunil Kalmady Vasu, for their endless support of my M.Sc. study and research. Without their persistent help, insightful comments, and suggestions, this dissertation would not have been completed. I have also benefited from their feedback during my thesis writing.

I also want to thank the professors Andrew Greenshaw and Bo Cao, who were the part of my thesis defense committee, for their valuable comments, which helped to improve the quality of the thesis.

I would like to thanks my parents and grandparents for their endless unconditional support throughout my life. Finally, I want to thank my loving wife, Priyanka, who has supported me, uplifted me, comforted me, and carried joy to my life when I felt sad or anxious or hopeless, had lost passion or energy.

This work was supported by the Alberta Machine Intelligence Institute (AMII), and Alberta Innovates Graduate Student Scholarship (AIGSS).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Psychiatric disorders are challenging to diagnosis and treat, due to heterogeneity in the clinical presentation [114], the comorbidity with other psychiatric disorders [150], and the fact that they do not have a precise and reliable set of diagnostic or prognosis features. Moreover, psychiatric symptoms can be hard to elucidate, because the manifestations of psychiatric disorders are often not tangible measurements like body weight or blood sugar level, which makes the diagnosis and treatment even harder. These problems are worse by the weak correspondence between the clinical symptoms and objective measurements like functional magnetic resonance imaging (fMRI), structural MRI (sMRI), and Diffusion MRI (DTI), genetic testing, etc.

In attempts towards evidence-based psychiatry, there has been a surge of fMRI studies of the brain of individuals with various psychiatric conditions in the past couple of decades [39]. Until recently, the goal of such studies was to find 'biomarkers' – features that were, individually, associated with the diagnosis at the group level (e.g., healthy controls versus patients) – using mass univariate or multivariate analyses in a 'hypothesis testing' approach, where a data generating model is assumed (e.g., generalized linear model), and hence fitted coefficients are interpreted as 'effect' on outcome (e.g., presence of OCD) [39], [102]. Generally, these models are validated by some measure of goodness of fit or by examination of residuals [156].

On the other hand, machine learning studies, such as ours, tries to build an effective prediction model (e.g., determine a particular subject has a mental

disorder or not) that does not assume a data generating model, and then use the statistically rigorous validation approach of estimating the prediction accuracy of the learned model based on its performance in labeling instances in a dataset of unlabeled cases (using cross-validation). However, the association fMRI studies in psychiatry are often based only on the training sample (i.e., without considering its generalizability) [16], and it is not surprising that those studies show poor replication in terms of implicated brain regions (this is over and above the other challenges in psychiatric studies such as clinical heterogeneity, etc.) [32].

This dissertation focuses on learning evidence-based prediction models. Many researchers now consider ways to learn a model (predictive study) based on the historical account of experiences and outcomes of patients [116]. The general framework used in this thesis for training and testing a machine learning method is given in Figure 1.1, where the classifier is trained using historical data (e.g., neuroimaging and clinical data) with target class labels that are based on diagnosis or prognosis labels. Finally, the learned classifier is tested on a novel subject. Many recent studies have demonstrated that fMRI of the brain has sufficient information to generate models that can discriminate healthy controls (HC) from patients with various psychiatric illnesses, such as autism [58], schizophrenia (SCZ) [10], [25], [78], [134], depression [40], and Obsessive-Compulsive Disorder (OCD) [17], [54], [132], [137], [145], [162].

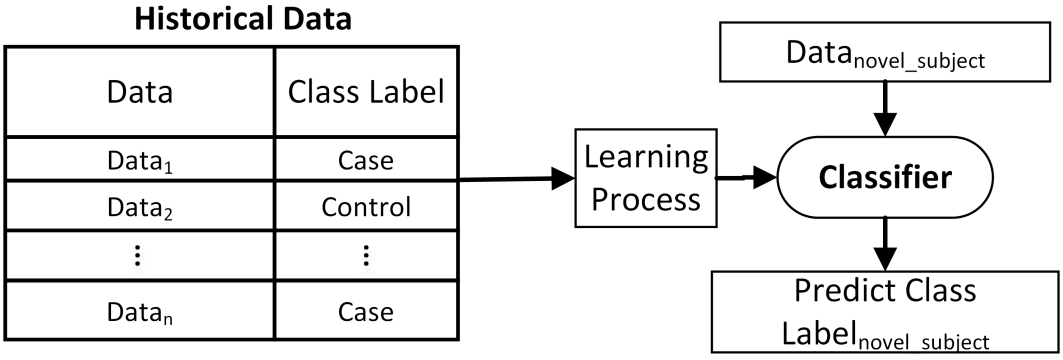


Figure 1.1: General framework used in this dissertation for machine learning and prediction in diagnosis/prognosis.

Inspired by the existing ML methods and the success of earlier studies,

we built different ML models for the diagnosis and prognosis of psychiatric disorders using fMRI data mainly. Some studies also incorporated other neuroimaging and clinical features with fMRI features to investigate the effect of those features.

1.1 Contribution

This dissertation makes the following specific contributions:

1. Empirically, we obtained 80.3% accuracy in predicting OCD against HC using fMRI data. The features of this model extracted from fMRI data are engineered by incorporating prior neurobiological knowledge of brain functioning – here, based on using just some specified pre-defined sets of brain regions. We used several ablation studies to show that by adding more domain knowledge in each step, that feature design based on prior neurobiological knowledge (parcellations) leads to better performance than agnostic and automated feature design (neural nets).

2. To examine the benefit of the transfer learning across psychiatric diagnoses (SCZ to OCD), we first learned a model to distinguish SCZ patients from HC. Then, we learned a model for distinguishing OCD from HC using only those features set that were selected in the learned model of SCZ versus HC prediction. We found that the selection of the feature sets can be transferred from SCZ to OCD prediction model without significant loss in prediction performance.

3. Recent studies have demonstrated that first degree relatives (FDRs) of patients with SCZ are more likely to exhibit intermediate phenotypes of schizophrenia (called 'endophenotypes') than the general population, even when they do not (or do not yet) present with a full set of clinical symptoms [46]. These characteristics of FDRs motivated us to investigate how such populations would be classified by a machine-learned model that is capable of distinguishing SCZ from HC based on resting-brain fMRI activation patterns. We observed that FDRs who were

(mis)classified as SCZ patients had significantly higher 'schizotypal personality scores' than those who were not classified as SCZ patients.

4. It empirically shows that maybe just often incorporating prior neurological knowledge helps to improve model performance. We were able to achieve 63.77% accuracy in predicting the antipsychotic treatments response using a deep transfer learning method, and 77.5% accuracy in predicting the transcranial direct-current stimulation (tDCS) treatments response using the feature concatenation method based on raw features – instead of average features – from some specified regions that were selected based on prior knowledge.

1.2 Dissertation Layout

The rest of this dissertation is organized as follows. Chapter 2 provides the necessary background for working with neuroimaging data, and outlines the preprocessing of raw fMRI, sMRI, and DTI data. Chapter 3 describes the methods, and methods' results for predicting the diagnosis of OCD[1]. Chapter 4 shows the analysis of schizotypal personality disorder for first degree relatives (FDRs) of SCZ patients[2]. Chapter 5 describes the different machine learning models for predicting the prognosis of two types of Schizophrenia patients based on the received treatment (Antipsychotic-treatment based, and Transcranial direct current stimulation (tDCS)-treatment based), and provides the results of each model. Finally, Chapter 6 presents the conclusion.

This dissertation contains two appendices at the end. Appendix A and Appendix B provides some additional information and experimental results of Chapter 3 and Chapter 5, respectively.

[1]Most parts of this chapter are taken from our submitted journal article "Prediction of Obsessive Compulsive Disorder: Importance of neurobiology-aided feature design and cross-diagnosis transfer learning", which is under review by *Nature Scientific Reports* journal.

[2]This chapter is taken from our submitted article "Extending Schizophrenia diagnostic model to predict Schizotypy in first degree relatives", which is under review by *Nature Schizophrenia* journal.

Chapter 2

Background Material

This chapter presents the background material necessary to understand the dissertation. Section 2.1 describes the basics of magnetic resonance imaging. Different types of MRI preprocessing pipelines are given in Section 2.2. Section 2.3 gives a brief description of the treatment response measurements.

2.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology that uses a powerful magnetic field and radio waves to build images of the body's organs, tissues, and structures. Clinicians use this to diagnose (i.e., figure out the current stage of cancer) and treat medical conditions. There are different types of MRI imaging techniques – each serving its own purpose. Section 2.1.1 - 2.1.3 give the basics of the three most common magnetic resonance imaging techniques: structural MRI, functional MRI, and diffusion tension imaging.

2.1.1 Structural Magnetic Resonance Imaging

Structural magnetic resonance imaging (sMRI) is an MRI technique that measures the local differences in water molecules, which helps to determine the shapes of specific physiological parts of the brain and the brain's subregions sizes. By using this technique, we can distinguish between normal (i.e., blood vessels, non-cancerous tissues) and abnormal (i.e., tumor), and between grey matter and white matter [152]. Here, sMRI gives information about the anatomy and pathology of the brain. Moreover, sMRI can

also be used as reference images for other imaging techniques when performing the data preprocessing steps (i.e., co-registration, normalization) [95]. In sMRI, brain structure is computed as 3d structure of voxels (i.e., one at each $1mm \times 1mm \times 1mm$ subregion) whose value is proportional to the intensity of that location, which corresponds to cerebrospinal fluid, gray matter, and white matter. Each voxel represents a 3-D volume of the particular block of the brain under the given slice thickness.

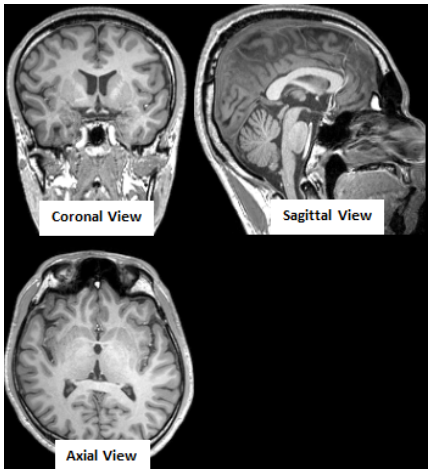


Figure 2.1: Sample of T1-weighted sMRI scan (Dark, light, gray, and bright colors denote cerebrospinal fluid, white matter, cortex, and fat, respectively.).

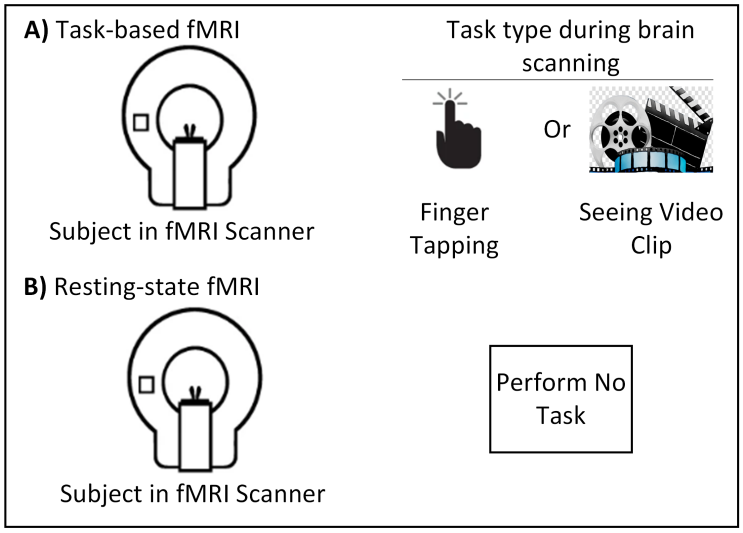


Figure 2.2: Task-based versus resting-state fMRI experiments.

2.1.2 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is a non-invasive process for understanding brain functions. fMRI scans capture the neural activity by measuring the changes of the oxygenated and deoxygenated hemoglobin concentration in the blood at certain locations in the brain, which known as the BOLD signal [75]. When a region of the brain is active, it requires oxygen. With the increase of neural activity, the demand for oxygen is increased. The blood that carries the oxygen fulfills this demand. For achieving the proper temporal resolution (collecting time series of brain information rapidly), fMRI has low spatial resolution (i.e., $3mm \times 3mm \times 3mm$) compared to sMRI (i.e., $1mm \times 1mm \times 1mm$). That is why we need sMRI scanning data with fMRI data of the brain.

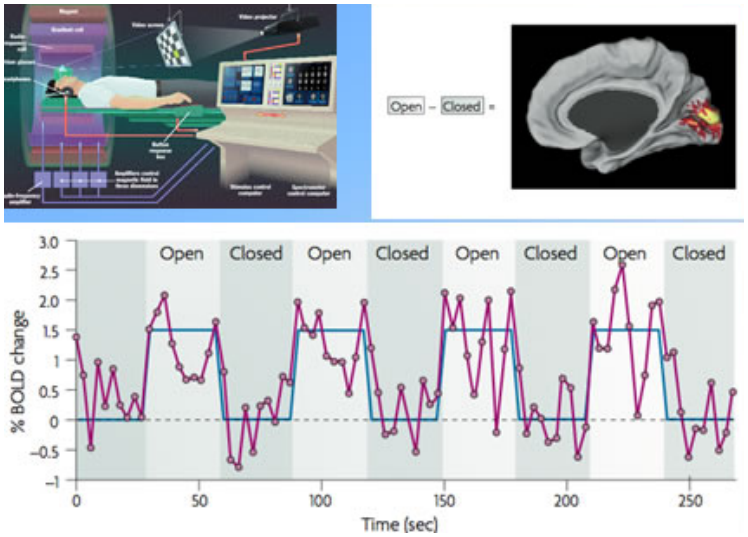


Figure 2.3: Brain activity fluctuation observed using fMRI experiment [153].

There are two types of fMRI experiments: task-based and resting-state (shown in Figure 2.2). In the task-based fMRI, the subject is performing some specific tasks – perhaps looking at the specific point, reading a book, finger tapping, seeing video clip, etc, but in the resting state process, the fMRI scan records the brain of subjects while he/she is not performing any task. Figure 2.3 shows a fMRI experiment in which the subject opens/ closes his/her eyes every 30 seconds – the results show that when the subject opens

eyes (means task performed) for 30 seconds, the signal of some brain regions is high compared to when the eyes are closed.

2.1.3 Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) data is collected for measuring the structural connectivity [130]. This technique helps to understand the white matter structure by measuring the water molecules diffusion in each voxel for all directions.

If we drop ink into a cup of water, water freely diffuses as there are no constraints, and the shape of the diffusion becomes a sphere. The center of the sphere does not move because there is no flow. This type of diffusion is called isotropic (shown in Figure 2.5). On the other hand, when there are constraints, the shape of the ink becomes oval in 2D space or ellipsoid in 3D space. This is called an anisotropic diffusion (shown in Figure 2.5).

In the biological tissues, water diffusion follows the actual path of white matter fibers, leading to "anisotropic diffusion" [29]. As the water diffusion in the brain is not uniform, the underlying tissue orientation can be computed using the diffusion amplitude and direction. The model tensor can be calculated from the amount of diffusion of the water and the direction of the diffusion. Based on the estimated tensors, we can find the structural connectivity (refer to Figure 2.4) by following the longest axis of the diffusion ellipsoid [57].

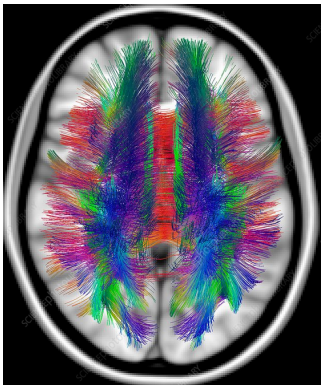


Figure 2.4: Reconstructed white matter fibres of the whole brain [154].

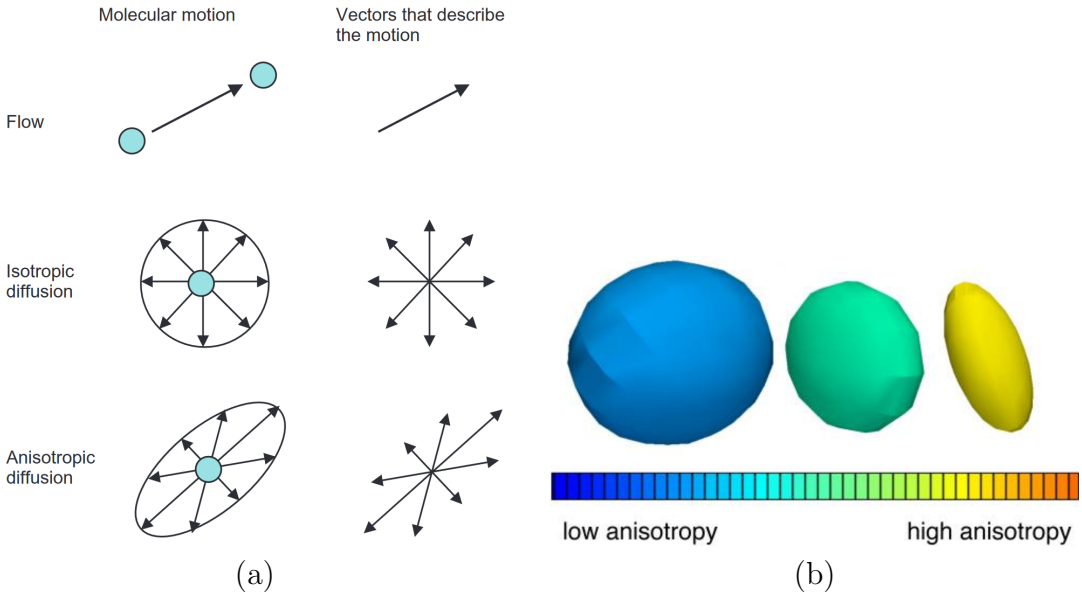


Figure 2.5: Types of diffusion: (a) Comparison between isotropic and anisotropic diffusion [89], (b) Example of three diffusion tensors to show the difference in tensor anisotropy and orientation [96].

2.2 Image Preprocessing Pipeline

2.2.1 fMRI Preprocessing Pipeline

The SPM-based Matlab toolbox DPARSFA (Data Processing Assistant for Resting-State fMRI – Advanced Version) [159] is used for fMRI data preprocessing and feature extraction. The general pipeline is depicted in Figure 2.6. The steps are explained briefly below:

Slice Timing Correction

The fMRI data collection process does not acquire the slices (shown in Figure 2.7) in a volume at a single time. There are different processes for acquiring the data: interleaved, ascending order, and descending order. For handling the time differences between the slice data collection, we need to apply the slice timing correction method, so that we can model that the resultant volume – all slices of the brain – as if they were acquired at a single time.

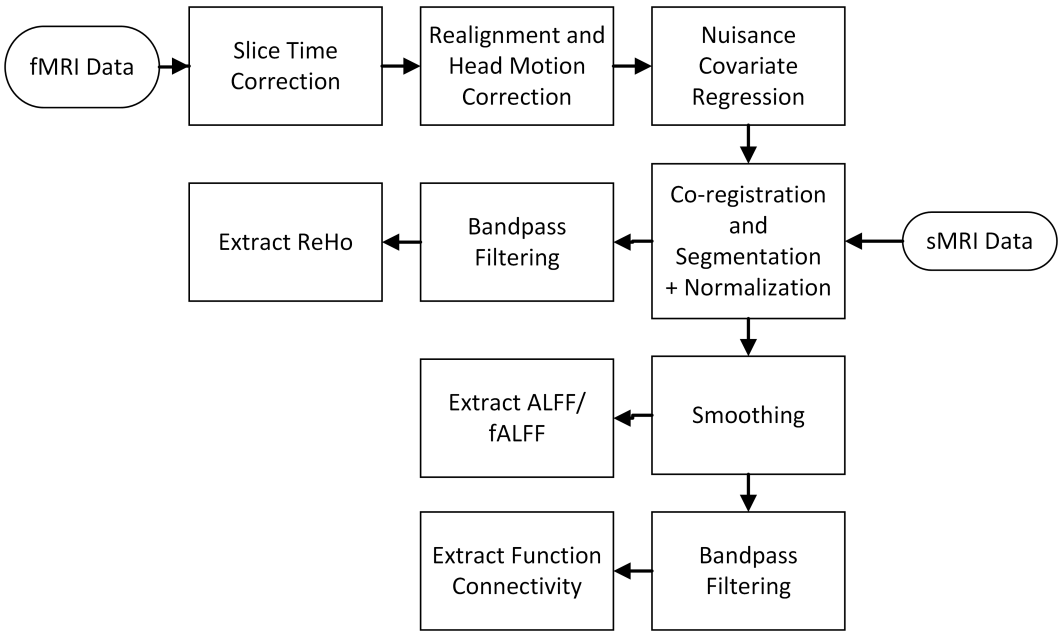


Figure 2.6: Pipeline of fMRI preprocessing and feature extraction.

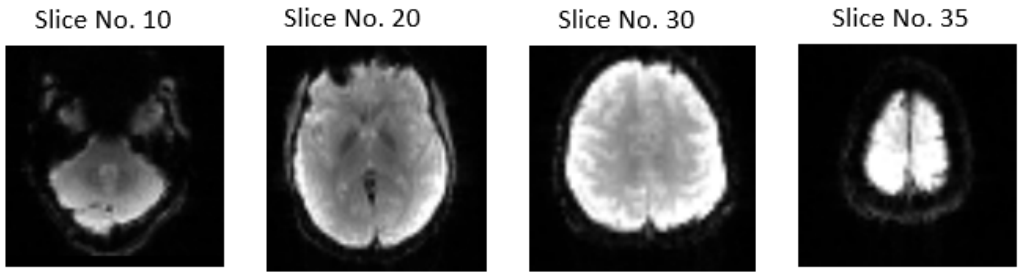


Figure 2.7: Axial slices of a brain volume for a particular time point.

Co-Registration and Segmentation

This co-registration process first superimposes the functional images on the structural image. This step is followed by segmentation, where the brain is divided into different tissue types (grey matter, white matter, and cerebrospinal fluid) using a standard template.

Realignment and Head Motion Correction

During the time of data acquisition, the subject's head position may move slightly with time. So, in the resultant outcomes of the experiment, the location of the functional images varies from time to time, and these unwanted

variations in the signal lead to poor data quality. The realignment and head motion correction method ensures that all the time series volumes are placed in the same location.

Nuisance Covariate regression

This regression is performed to remove mean global signals from cerebrospinal fluid segments (CFS) and white matter signal, and also, denoise signals those were induced as a result of head motion. It is done by using SPM's new segmentation method [37].

Normalization

As the subjects' brains have different shapes and sizes, for making the comparison between the subjects, we need to normalize the fMRI images. Here, we can transform all the images to one standard space/template (e.g., Montreal Neurological Institute (MNI)[1], Statistical Parametric Mapping (SPM)[2] template, etc.), which involves computing a transformation matrix from the original image to the template. Here, we map each normalized fMRI image to MNI space (it is a common coordinate space that is defined by Montreal Neurological Institute using a large series MRI of healthy individuals [85]) resampled to $3 \times 3 \times 3mm^3$.

Smoothing

A 3D Gaussian kernel with a 4mm width is used on all datasets for smoothing purposes. This minimizes noise and effects due to residual differences in functional and gyral anatomy during inter-subject averaging [159].

Bandpass Filtering

We band-pass filtered (0.01–0.08 Hz) the normalized smoothed images with respect to temporal data, which helps to reject frequencies that are outside the range of 0.01–0.08 Hz frequency.

[1] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases`
[2] `https://www.fil.ion.ucl.ac.uk/spm/`

Extracting ALFF and fALFF

ALFF and fALFF are the measurements of the regional spontaneous neuronal activity. These measurements help to quantify the difference in neuronal activity between subjects and between conditions. That is why we explored the possible use of these two measurements in the classification framework.

ALFF (Amplitude of Low Frequency of Fluctuations) was calculated as total power within the frequency range between 0.01 and 0.08Hz to estimate the strength of low-frequency oscillations [164], whereas fALFF (Fractional ALFF) was calculated as the power within the low-frequency range (0.01–0.08Hz) divided by the total power within the whole frequency range [171]. Computed powers for two measures (ALFF and fALFF) are transformed into Z-scores for each subject. Finally, for each subject, it gives a 3D false image, where the value at (x, y, z) is the intensity of low-frequency oscillations (LFO) for ALFF, and the relative contribution of specific LFO to the whole frequency range for fALFF [7].

Extracting ReHo (Regional Homogeneity)

ReHo is a voxel-based analysis of the brain, which measures the regional synchronization of neural activities. It computes Kendall's coefficient concordance (KCC) to determine the similarity between the time series of the given voxel and its nearest neighbors [65], [165]. The computed similarity values are standardized using Z-scores [112]. Finally, it gives a 3D image for each subject. We can use this data along with the other features in the classification framework as it helps to find the difference in regional homogeneity between healthy subjects and patients.

2.2.2 sMRI Preprocessing Pipeline

We used SPM's CAT12 toolbox[3] that is widely used for the structural MRI data preprocessing. The general preprocessing pipeline involved four steps:

[3] `http://www.neuro.uni-jena.de/cat/`

1. **Normalization**: First, normalize sMRI images to one standard space/ template (e.g., MNI, SPM template, etc.).

2. **Segmentation**: Segment the brain into three parts: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). So, each voxel of the brain must be a member of one of these parts [109].

3. **Modulation**: Modulation tries to maintain the same amount of GM volume in the original image and the segmented GM image [86]. In order to do it, the rate of contraction that has occurred during the spatial normalization step is used for scaling the voxel concentration of GM volume. The value of Jacobian determinant provides a measure of the volume changes, and it is calculated from the deformation field generated from the spatial normalization step [101]. Then, these values are multiplied with the voxel values of GM volume for computing the modulated GM images (shown in Figure 2.8).

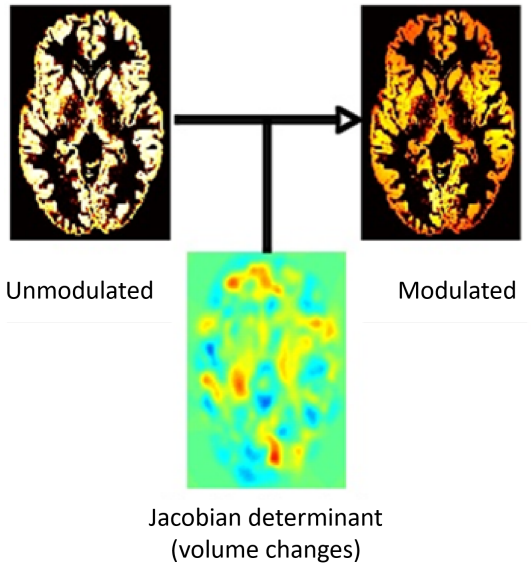


Figure 2.8: Correction of GM volume changes during spatial normalization using Jacobian determinant (yellow = compressed regions, and blue = expanded regions.) [86].

4. **Spatial smoothing**: The 8mm full width half maximum (FWHM) 3D Gaussian kernel is used for smoothing purposes.

The sample outputs from the sMRI preprocessing pipeline are given in Figure 2.9.

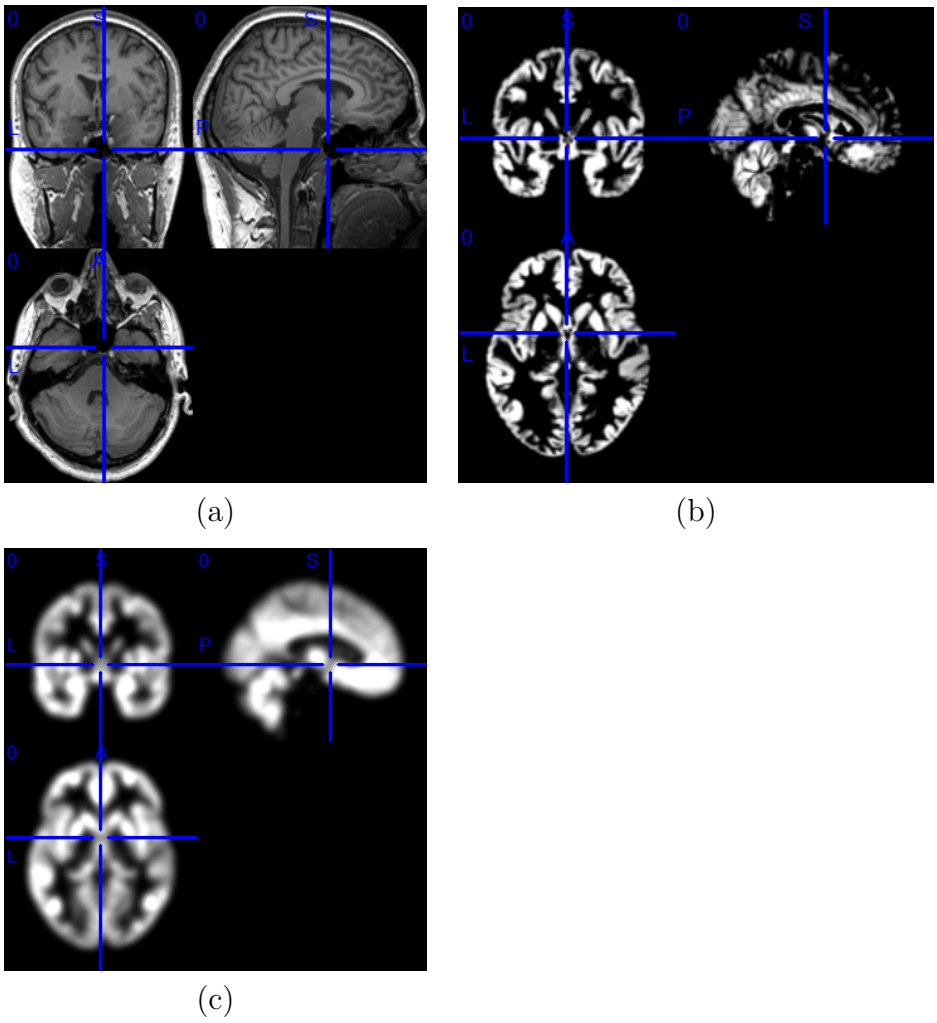


Figure 2.9: Sample outputs of sMRI preprocessing: a) original sMRI image, b) segmented gray matter image after normalization, segmentation and modulation steps, and (c) smoothed image.

2.2.3 DTI Preprocessing Pipeline

DTI Preprocessing and Feature Extraction

We used the Enigma Group's DTI preprocessing protocols followed by their given templates for DTI preprocessing, and feature extraction [33]. Our in-house python code, which is written using Nipype's FSL interfaces [92], is used for performing the preprocessing steps. The general preprocessing pipeline

involved the following steps:

1. **Convert 3-dimensional image to 4-dimensional image**: There are 360 samples in our dataset, It contained 65 three-dimensional images as DTI data for each subject. Initially, we combined all 3D images to make it a 4D DTI image for each subject.

2. **Re-orient the images to standard space**: Re-orient the 4D images to standard space (e.g., MNI152 template) using the FSL reorientation module.

3. **Correct eddy current and head motion**: Diffusion images are distorted based on eddy currents and the motion of the head. Due to the eddy currents, images contain some artifacts (shown in Figure 2.10) like the shear, enhanced background, loss of image intensity, tracking the wrong fiber, etc [146]. These distortions vary with different gradient directions. We have used FSL's eddy current correction method for correcting these distortions, and along with distortions correction, head motion is removed by using the affine co-registration to a reference volume of this specific subject [146].

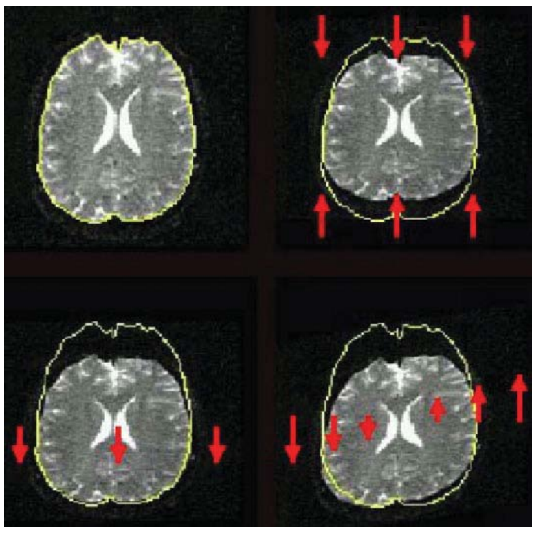


Figure 2.10: Examples of distortion resulting from eddy current: contraction (top right), shift (bottom left), and shear [139].

4. **Brain extraction**: Removing the non-brain tissues and the skull from the data called the "brain extraction". We have applied the brain extraction tool (BET) from FSL that identifies the brain tissue before fitting the tensor modeling. The output of this tool is the binary mask that distinguishes the brain versus non-brain at each voxel [43].

5. **Tensor model fit**: For calculating the tensor model, we need the gradient information of DTI images. There are four ways to measure the critical features of diffusion tensor, viz., mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA). Measurement of these measures depends on the three eigenvalues (i.e., λ_1,λ_2, and λ_3 denote the length of the axis in the tensor) of the tensor. AD is equal to the λ_1, which is the largest eigenvalue, RD is the average of the other eigenvalues ($(\lambda_2 + \lambda_3)/2$), and MD is the average of all eigenvalues, which is considered as a total amount of diffusion in a voxel. FA quantifies the amount of anisotropic diffusion within a voxel, which is the most common measurement of the DTI studies.

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

   The range of the FA values is 0 to 1 (refers to Figure 2.11). If the value is 0, the diffusion of the water molecules is isotropic, meaning diffusion is equal in all directions. If the value is 1, all the molecules are diffused along one direction [141].

6. **Skeletonized image generation**: For generating the skeletonized images, I have used FSL's TBSS method [129] and Enigma template [34]. The standard TBSS procedure is as follows:

   a) Firstly, it erodes images slightly and adds zero to the end slices for removing outliers, which are generated from the tensor fitting process.

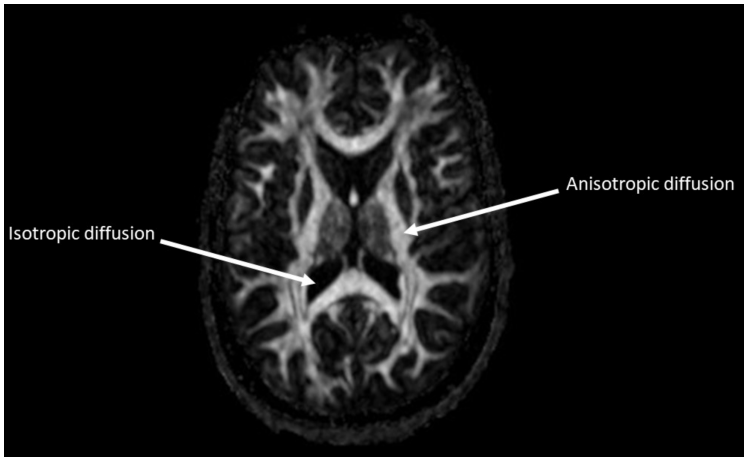


Figure 2.11: Fractional Anisotropy (FA) map – white color (value = 1) denotes anisotropy is high in white matter but black color (value = 0) denotes low in ventricles [31].

b) Secondly, it applies the non-linear registration using Enigma template

c) Thirdly, it makes registration of the images linearly – images are generated from the previous step – to the standard MNI space.

d) Finally, it generates the skeletonized images [140].

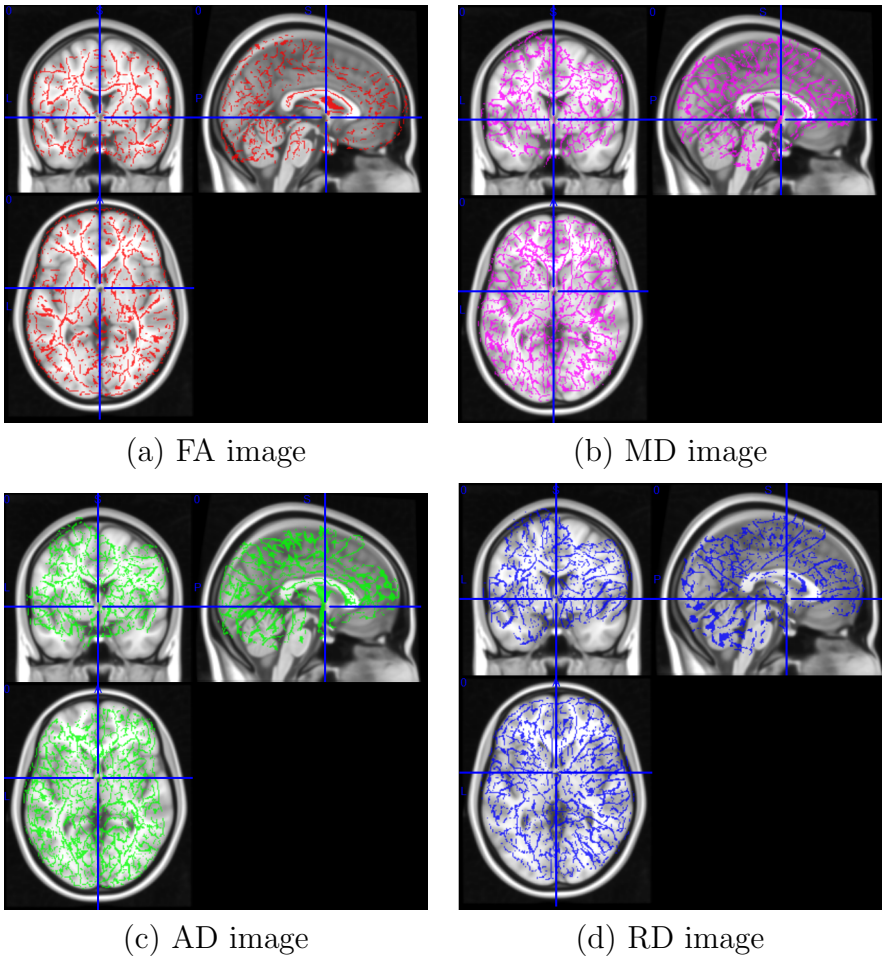
We need to follow these above steps to generate the skeletonized images from FA, MD, AD, RD images. Output results for a single subject (Overlay FA, MD, AD, RD images over MNI152_T1_1mm template[4]) are given in Figure 2.12.

2.3 Treatment Response Measurements

2.3.1 Y-BOCS_O and Y-BOCS_C

We have used Yale–Brown Obsessive Compulsive Scale (Y-BOCS) [77], which consists of 10 questions [158], for measuring the symptom severity of Obsessive-Compulsive Disorder (OCD) that is characterized by intrusive recurrent thoughts (obsessions) and repetitive behaviours (compulsions). Y-BOCS has 10 core questions: 5 questions (time duration, interference, distress, resistance, and

[4] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases`



(a) FA image (b) MD image

(c) AD image (d) RD image

Figure 2.12: Overlay FA, MD, AD, RD images over MNI152_T1_1mm template.

control of obsessions) for rating obsession (Y-BOCS_O) and 5 questions – same as Y-BOCS_O rating items– for rating compulsion (Y-BOCS_C). The ratings of these questions are assigned by clinicians. Each question takes a rating from 0 to 4, where 0 denotes no symptoms, and 4 denotes severe symptoms. Summation of the 5 items of obsession gives total values of Y-BOCS_O, and similarly, the total of the 5 items of compulsion gives total Y-BOCS_C values. Based on these Y-BOCS_O and Y-BOCS_C scores, we categorized the patients into least versus most severe (described in Section 3.7.4).

2.3.2 Clinical Global Impression (CGI)

Clinical Global Impression (CGI) [53] is widely used for quantifying symptom severity, the treatment response, and the efficacy of the given treatment to the mental disorder patients. Experienced clinicians assign the rating (completely based on the clinician's subjective assessment) using this CGI scale to quantify the severity of the patient's illness at the assessment time, considering the particular population with clinicians' total clinical experience. CGI severity of illness (CGI_S) range is 1 (normal) to 7 (among the most extremely ill patients). We have used the baseline and follow-up CGI_S scores for defining the treatment response (Responder or Non-Responder) of schizophrenia patients (described in Section 5.3.1), and used the baseline CGI_S scores for incorporating the transfer learning concept in the deep learning models (described in Section 5.2.3).

2.3.3 SAPS and SANS

SAPS (Scale for the Assessment of Positive Symptom) and SANS (Scale for the Assessment of Negative Symptoms) are used to measure the positive/ negative symptoms of schizophrenia [68]. SAPS has 34 sub-items for which clinicians assign severity levels ranging from 0 (None) to 5 (Severe). These 34 sub-items fall under 4 groups: hallucinations, delusions, bizarre behavior, and positive formal thought disorder. SANS consists of 25 sub-items, which fall under 5 groups: affective flattening or blunting, alogia, avolition-apathy, anhedonia-asiciality, and attention. We defined the treatment response (Responder or Non-Responder) of the patients using baseline and follow-up SAPS and SANS scores, described in Section 5.3.1.

2.3.4 Auditory Hallucination (PsyAH)

Psychotic Symptom Rating Scales (PSYRATS) is widely used to measure the severity level of delusions and hallucinations in schizophrenia patients [157]. Psychotic Symptom Rating Scales- Auditory Hallucination (PsyAH) includes frequency, duration, location, loudness, belief about the origin, amount of

negative content, degree of negative content, amount of distress, intensity of distress, disruption to life, controllability of Auditory Hallucination. Each sub-item of PsyAH takes rating from 0 (no symptoms) to 4 (severe symptoms) [50], and the ratings are assigned by clinicians. So, the total score range for PsyAH is 0 to 44. We have used the baseline and follow-up PsyAH scores for defining treatment response (Responder or Non-Responder), described in Section 5.3.1.

Chapter 3

Obsessive-Compulsive Disorder Diagnosis Prediction and Cross-Diagnosis Transfer Learning

3.1 Introduction

Obsessive-Compulsive Disorder (OCD) is a debilitating condition – the fourth-most common psychiatric illness, which affects millions of people worldwide [115]. Evidence shows that a reliable diagnosis of the condition is crucial for timely intervention and improvement in patients' overall quality of life [61]. However, due to heterogeneity in the clinical presentation [114] and comorbidity with other psychiatric disorders [150], including bipolar disorder [6] and schizophrenia [120], precise diagnosis continues to be a challenge.

In the last few years, there has been an increased use of machine learning methods in psychiatric applications, often to produce models that predict the diagnosis of novel subjects [116]. Many studies have demonstrated that functional Magnetic Resonance Imaging (fMRI) of the brain has sufficient information to generate models that can discriminate healthy controls from patients with various psychiatric illnesses, such as autism [58], schizophrenia [25], depression [40] as well as OCD [17], [48], [54], [122], [125], [137], [162] (see list of OCD studies in Table 3.1). However, most of these studies were conducted on small datasets, that used fewer than one hundred subjects for

both learning and cross-validation of the learned model [17], [48], [122], [125]. Limited samples can severely limit the generalizability of such models, and in fact, many have noted that small sample sizes can lead to overestimated measures of performance [121]. Studies on larger sample size (greater than one hundred) have reported OCD prediction performance ranging from 72 to 79% [54], [137], [162].

While it is important to have sufficient information to learn a generalizable model, it is challenging to conduct studies that sample thousands of psychiatric patients due to pragmatic reasons – such as subjects recruitment challenges, hard to get patients consent for information sharing, etc. Therefore, it is critical to extract as much relevant information possible from limited data at hand, to build a generalizable model – and in particular, to reduce the chance of overfitting. Kalmady et al. [63] explored a way to design features – each a combination of one of several regional connectivity-based measures, and one of the various different parcellation maps (each incorporating a unique source of prior neurobiological knowledge) – then learn a classifier based on these features. They found this framework, called EMPaSchiz (read as 'Emphasis'; standing for 'Ensemble algorithm with Multiple Parcellations for Schizophrenia prediction'), could effectively discriminate drug-naive schizophrenia patients from healthy controls [63]. To date, however, this approach has not been explored to distinguish OCD patients from healthy controls; none of the earlier machine learning studies in OCD [17], [48], [54], [122], [125], [137], [162] have combined the regional and connectivity features, nor have any employed an ensemble approach to jointly learn from multiple parcellations. Moreover, no one has explored whether such knowledge-driven methods can provide performance that is comparable to today's standard machine learning approach of applying neural networks, which are generally known to provide 'state of the art' results [123], albeit being less transparent than simpler models [23] such as EMPaSchiz.

Today, psychiatric disorders are classified based on a consensus about clusters of symptoms that the patients experience currently or in the past. Classificatory systems, such as the Diagnostic and Statistical Manual of Mental

Disorders (DSM), have provided clinicians worldwide with a standard framework to identify, treat, and manage these conditions [5]. However, while most scholars are convinced that pathology underlying symptoms of psychiatric illnesses is attributable to features of neural pathways and their interfacing axes, most clinical decisions in psychiatric practice are based neither on the etiological mechanisms [91] nor on dynamic aspects of brain structure or function [59]. In fact, decades of research towards the understanding of neuropathology underlying these illnesses have shown only weak and unclear correspondence between brain measures and clinical categorization [24]. Interestingly, many brain regions and networks that are implicated in one psychiatric disorder seem to be implicated in other disorders (or at least share a common subset) [44]. This observation can be potentially leveraged when learning models by transferring knowledge gained with a dataset of one condition, to help diagnose another condition, where features that are relevant for predicting one psychiatric disorder can be used to predict another – this is considered a type of 'transfer learning' [21].

In addition to generalizability, it is critical that machine learning models in healthcare demonstrate robustness and trustworthiness [52]. Hence, it is valuable to build models that can provide interpretability in terms of providing (a) auxiliary arguments in favor of model's correctness, (b) transparency with the usage of features explainable by domain experts such as psychiatrists, and (c) reproducibility to ensure the model can be trusted in high-stakes situations such as medical diagnosis.

In this study, we apply the EMPaSchiz learning approach to an OCD-vs-Control dataset, to produce a model that can diagnose OCD, and show empirically that this use of EMPaSchiz demonstrates some of the points mentioned above. We believe that these factors are relevant to the future of neuroimage-based machine learning methods for psychiatric diagnosis, and medical diagnosis in general. This research mainly explores the following questions:

1. Can EMPaSchiz approach produce models that can predict OCD accurately, in a way that is interpretable?

2. Can customized feature design, based on prior neurobiological knowledge (parcellations) produce models that are comparable to knowledge-agnostic automated methods (neural nets)?

3. Can we transfer information obtained by learning models for schizophrenia diagnosis, to OCD diagnosis, by using the high-level features selected from the model learned for predicting schizophrenia, for the task of predicting OCD?

3.2 Dataset

3.2.1 Samples

Our study sample contained 188 patients attending the OCD Clinic of the National Institute of Mental Health & Neurosciences (NIMHANS, India), who fulfilled DSM-IV criteria for OCD. The diagnosis of OCD was established using the Mini International Neuropsychiatric Interview (MINI) Plus [124], which was confirmed by another psychiatrist through an independent clinical interview. Symptoms were measured using the Yale-Brown obsessive-compulsive scale (Y-BOCS) [45]. Healthy Controls (HC) were recruited from among the consenting healthy volunteers from the same locale to match for age and sex. We used 200 age- and sex-matched HC, who were screened to rule out any psychiatric diagnosis using the MINI. Table 3.2 provides details of the demographic and clinical profiles of subjects who qualified to be included in the study.

In addition to OCD and HC, we also used a dataset of drug-naive schizophrenia patients for transfer learning analyses in this study. This cohort has been analyzed in the Kalmady et al. study [63], and basic demographic information of these subjects is provided in Table 3.3. They obtained written informed consent after providing a complete description of the study to all the subjects. The NIMHANS ethics committee reviewed and approved the original research protocol. The Research Ethics Board at the University of Alberta, Edmonton approved the secondary analysis of de-identified, preprocessed data.

Study	Year	Total: Size of classes	Data Acquisition	Features	Feature selection / reduction	Classifier	Validation	Accuracy, %
Soriano-Mas	2007	144: 72 OCD, 70 HC + 60[a]: 30 OCD, 30 HC	sMRI	GMV	Not Specified	Probabilistic classifier	LOO-CV	76.6
Shenas	2013	24: 12 OCD, 12 HC	Resting state fMRI	Functional connectivity	PCA, KPCA, LDA	SVM	LOO-CV	66
Gruner	2014	46: 23 OCD, 23 HC	Resting state fMRI	Functional networks	ICA	Logistic Regression	LOO-CV	80.1
Li	2014	56: 28 OCD, 28 HC	diffusion MRI	FA map	Not specified	SVM	LOO-CV	84
Parrado-Hernánde	2014	172: 86 OCD, 86 HC	sMRI	Voxel-based GM	Multivariate feature selection method	SVM	LOO-CV	73.9
Shenas	2014	24: 12 OCD, 12 HC	Task-fMRI	Functional connectivity	PCA, kernel PCA and linear discriminant analysis (LDA)	SVM	LOO-CV	74
Hu	2016	66: 33 OCD, 33 HC	sMRI	GMV and WMV	Voxel-wise univariate analysis	SVM	LOO-CV	81.8
Sen	2016	29: 16 OCD, 13 HC	Resting state fMRI	Graph-based	Minimum Redundancy Maximum Relevance (mRMR)	SVM	LOO-CV	80
Takagi	2017	108: 56 OCD, 52 HC + 28[a]: 10 OCD, 18 HC	Resting state fMRI	Functional connectivity	L1-norm regularized sparse canonical correlation analysis	Sparse LR	LOO-CV	73
Trambaiolli	2017	74: 38 OCD, 36 HC	sMRI	Region-based	Two multivariate and five univariate FS algorithms	SVM	LOO-CV	71.6
Zhou	2018	93: 48 OCD, 45 HC	sMRI + DTI	GMV,WMV, FA, and MD	Use PRoNTo's "Prepare feature set" programs.	SVM	LOO-CV	61.29-80.65
Yang	2019	136: 68 OCD, 68 HC	Resting state fMRI	fALFF	Not Specified	SVM	LOO-CV	72
Hu	2019	176: 88 OCD, 88 HC	Resting state fMRI	ReHo	Not Specified	SVM	LOO-CV	78.98
Bu	2019	98: 44 OCD, 44 HC	Resting state fMRI	ALFF, fALFF, ReHo, FCS	Not Specified	SVM	LOO-CV	74.07-95.3

All of the studies used single site data
HC= Healthy Controls, sMRI = Structural MRI, fMRI= Functional MRI, GMV = Gray matter volume, WMV = white matter volume, FA = Fractional anisotropy, and MD = Mean diffusivity; a) Dataset is used for test validation

Table 3.1: Description of studies that provided machine learning model for predicting OCD.

Characteristic	OCD	HC	Stat	p
N	175	175		
Sex [M:F]	97:78	105:70	0.57^b	0.449
Age	28.66 ± 6.11	27.93 ± 4.68	1.26^a	0.207
Total Intracranial -Volume (TICV, mL)	1380.16 ± 142.65	1428.79 ± 150.26	3.09^a	0.002
Age at onset	21.36 ± 7.45			
YBOCS-Compulsion	12.55 ± 4.25			
YBOCS-Obsession	13.43 ± 2.92			
YBOCS-Total	25.97 ± 6.45			

a Independent Sample Test [t]
b Chi-Square test [χ^2]
N = 152 for Age at onset, YBOCS-Compulsion, YBOCS-Obsession, YBOCS-Total

Table 3.2: Clinical and Demographic profile of OCD and HC subjects.

Characteristic	Schizophrenia	HC	Stat	p
N	81	93		
Sex [M:F]	53:28	60:33	0.001^b	0.97
Age	30.72 ± 6.16	29.41 ± 5.71	1.45^a	0.15
Total Intracranial -Volume (TICV, mL)	1400 ± 135	1480 ± 144	3.76^a	<0.001
Age at onset (N=152)	26.6 ± 6.17			

a Independent Sample Test [t]
b Chi-Square test [χ^2]

Table 3.3: Demographic and clinical profile of SCZ and HC subjects (previously published in [63]).

3.2.2 Samples Acquisition

Magnetic Resonance Imaging (MRI) was done in a 3.0 Tesla scanner (Magnetom Skyra, Siemens). Resting-State Functional MRI: BOLD (Blood Oxygen Level Dependent) sensitive echo-planar imaging was obtained using a 32-channel coil for a duration of 5 minutes 14 seconds, yielding 153 dynamic scans. The scan parameters were: TR= 2000 msec; TE = 30 msec; flip angle = 78 degrees; Slice thickness = 3 mm; Slice order: Descending; Slice number = 37; Gap = 25%; Matrix = $64 \times 64 \times 64 mm^3$, FOV = 192×192, voxel size = 3.0 mm isotropic. Subjects were asked to keep their eyes open during the scan. For

intra-subject co-registration, structural MRI: T1-weighted three-dimensional high-resolution MRI was performed (TR = 8.1 msec, TE = 3.7 msec, nutation angle = 8 degree, FOV = 256 mm, slice thickness = 1 mm without inter-slice gap, NEX = 1, matrix = 256×256) yielding 165 sagittal slices.

3.3 Data Preprocessing and Feature Extraction

3.3.1 Data Preprocessing

Preprocessed fMRI data are collected from National Institute of Mental Health Neurosciences (NIMHANS), India, which was preprocessed with DPARSFA (Data Processing Assistant for Resting-State fMRI — Advanced Version) [159], which is an SPM based Matlab toolbox. Initially, they visually inspected the acquired images for artefacts such as incomplete brain coverage or ghosting. After that, they discarded the first ten volumes of each functional time-series before reaching steady magnetization and to allow the participants to adapt to the scanning noise. Then, they followed the fMRI preprocessing pipeline, which is described in Section 2.2.1.

We excluded images for 8 patients and 7 controls from the study based on excessive head movement (translational > 2.0 mm and/or rotational $> 2^{\circ}$) [22] in order to avoid class differences in head motion. In addition, 5 patients and 18 controls were excluded due to incomplete imaging or clinical data. This yielded a total of 350 subjects: 175 controls and 175 patients.

3.3.2 Feature Extraction

For building a good machine learning model for fMRI, we extracted the neurobiologically relevant features of fMRI. Here, we used 14 different brain parcellations schemes that each extracts information from its predefined atlas or set of regions of interests (ROIs). These schemes varied widely in principle: (a) pre-defined ontology of brain structures such as post-mortem cytoarchitecture [147], [168], sulco-gyral anatomy [28], [138], or (b) data-driven modelling of the functional features from resting-state [18], [104], [149]

or task-based fMRI [128], or (c) meta-analyses [30], using analytical techniques such as hierarchical clustering [11], or independent components analysis [81]. In particular, our brain parcellation schemes are power [104], dosenbach [30], yeo [142], aal [147], basc_multiscale_122, basc_multiscale_197, basc _multiscale_325, basc_multiscale _444 [11], destrieux [28], harvard_cort_25[1], harvard_sub_25, smith20, smith70 [128], and msdl [149]. From the fMRI data, using these 14 parcellations, we have extracted two types of fMRI features: Regional-based and connectivity-based features using nilearn python package [2].

Regional Based fMRI features

We used ReHo, ALFF, fALFF measures (described in Section 2.2.1) for extracting the regional fMRI features. Furthermore, a nuisance regression was applied to the features to remove the effect of some confounding variables (age, sex, total intracranial volume, and framewise displacement).

Connectivity-Based fMRI features

We computed functional connectivity between each pair of regional bold signals – averaging the bold signals per region – for each parcellation. We, therefore, used three statistical measures: the inter-regional Pearson correlation (FC-corr), partial correlation (FC-part), and precision (FC-prec) for extracting the connectivity-based fMRI features. For each measure, we considered the lower triangular part of the symmetric matrix and then flattened that to get a feature vector.

3.4 Methods for OCD's Diagnosis Prediction

3.4.1 EMPaSchiz Framework

We used EMPaSchiz (shown in Figure 3.1) to learn a model to predict the diagnosis of OCD. The detailed description of the EMPaSchiz system is provided in the original paper [63]. Briefly, EMPaSchiz extracts 6 resting-state brain

[1] `http://www.cma.mgh.harvard.edu/fsl_atlas.html`

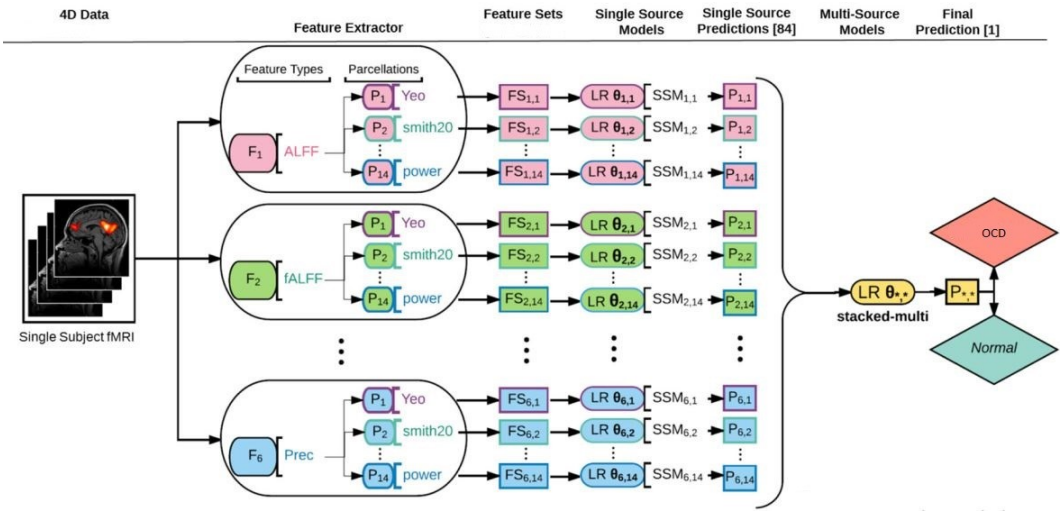


Figure 3.1: EMPaSchiz framework [63].

fMRI features, including 3 regional-based features (ALFF, fALFF, ReHo) and 3 connectivity-based features (FC-corr, FC-part, FC-prec). EMPaSchiz projected each feature extraction onto 14 different parcellations schemes (described in Section 3.3.2). EMPaSchiz first learns 84 (14 parcellation schemes $\times$ (3 + 3) feature types) single-source models (SSMs), each applying L2-regularized logistic regression to learn a classifier, for one [feature-type, parcellation] description of the data. It then applies L2-regularized logistic regression to the 84 prediction probabilities produced by each of these learned models, over the training set, to learn a final ensemble system. At performance time, given a new instance, the learned system will first produce the 84 descriptions of that new instance, then run those 84 SSMs to produce 84 responses, then feed those values into the final learned function, to output the final OCD-vs-HC label.

In Figure 3.1, F_1,.., F_6 denotes 6 feature types, and, P_j denotes the j-th parcellation, where j = 1, 2, ..., 14. Feature sets are shown as $FS_{i,j}$ (e.g., $FS_{1,1}$ to $FS_{1,14}$ each correspond to F_1 feature set), the logistic parameters $\theta_{i,j}$ denotes the SSMs, and $P_{i,j}$ denotes the output of the corresponding SSM, where i denotes the i-th feature type, and j denotes the j-th parcellation. Also, LR $\theta_{*,*}$ denotes the final ensemble model, which outputs probability values ($P_{*,*}$). We used 0.5 as a threshold for determining the final class label

– i.e., return class OCD if and only if the predicted probability was > 0.5.

3.4.2 Neural Network

To examine the importance of incorporating prior neuroanatomical and neurofunctional knowledge in the form of feature design and brain parcellations, we compared the performance of EMPaSchiz with today's standard approach, of using Neural Network (NN). Recent projects [41], [56], [71], [83], [93], [119], [136], [143], [144], [148], [160], [161], [163], [167], [170] using NN models on fMRI data for various psychiatric and neurological tasks have used several consecutive convolutions and max-pooling layers (where the specific network architecture is not fixed, but can depend on the specific task) – which is implicitly extracting high-level features. Most of these studies [41], [83], [119], [136], [143], [144], [148], [161] ran 2D convolutions on the 3D or 2D data that are generated from the original fMRI data, but a few studies [71], [93], [167], [170] show that applying 3D convolution on fMRI data (here, entire brains rather than slices) is more effective than using 2D convolution to produce accurate predictors.

We performed this analysis with three levels, adding more domain knowledge in each step (see Figure 3.2):

NN-1: Neural Network using pre-processed fMRI images without manual feature design

Here, we input 4-dimensional fMRI images that are completely pre-processed, and let the first neural net learner, NN-1, automatically learn relevant features from the data. The learned NN-1 model takes a fixed input size of $143 \times 61 \times 73 \times 61$, for each subject, where 143 is the time dimension, and the other three dimensions represent the 3D brain volume. This model consists of two 3D convolution layers with one max-pooling layer, three layers of Long Short-Term Memory (LSTM), and two fully-connected layers.

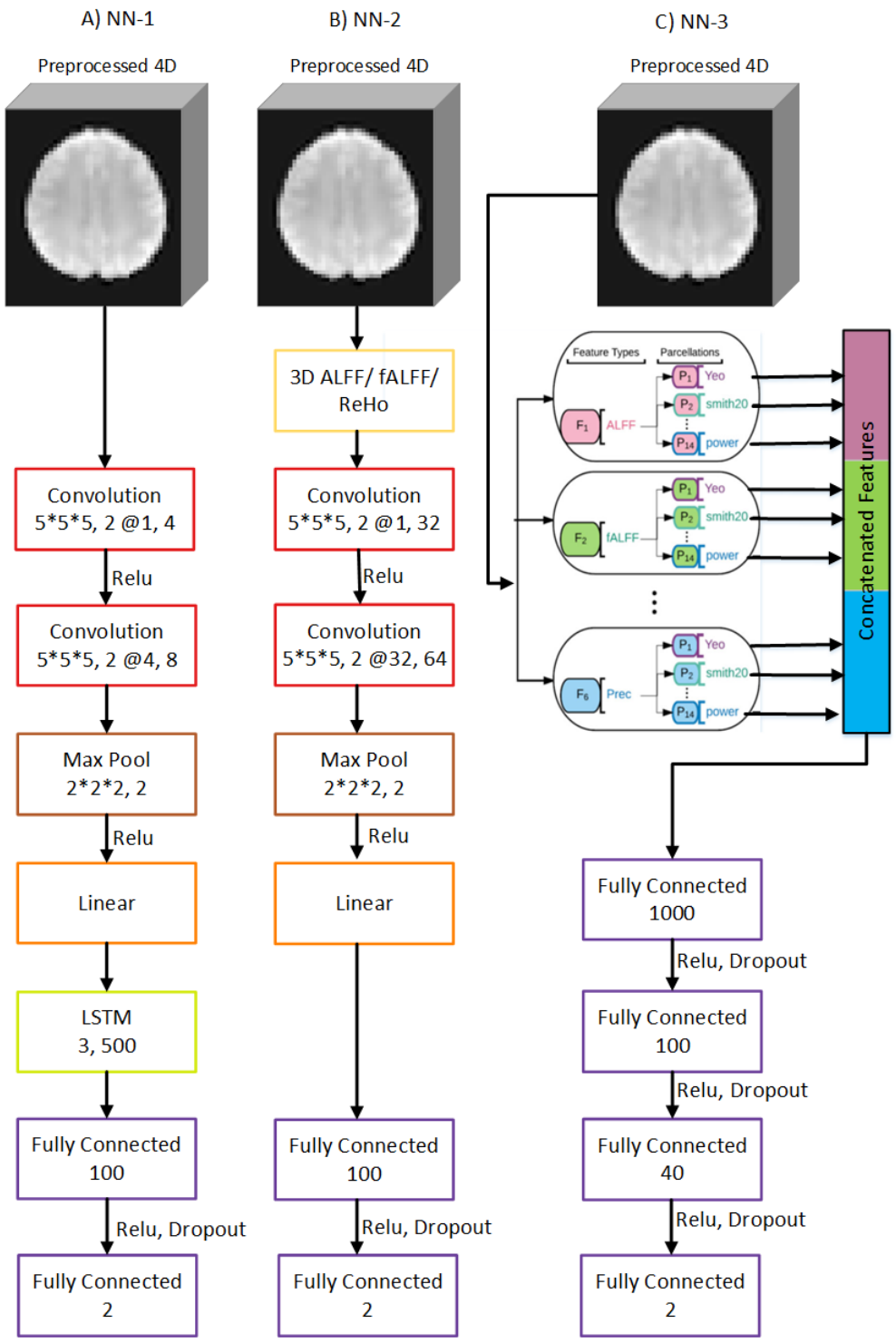


Figure 3.2: The architecture of OCD classification using NN-1, NN-2 and NN-3.

NN-2: Neural Network using designed features, but without parcellation-based aggregation)

Here, we extracted the three different features types (ALFF, fALFF, and ReHo features) using EMPaSchiz framework, but did not aggregate the features based on the brain parcellations. Instead, NN-2 uses a convolutional neural network to predict the OCD label for each of the three feature types individually. The learned NN-2 model takes a fixed input size of $61 \times 73 \times 61$. It consists of two convolution layers with one max-pooling layer and two fully-connected layers.

NN-3: Neural Network using designed features and brain parcellations

Here, we used the feature extractions and parcellation schemes similar to that of EMPaSchiz, however instead of learning with the simplistic logistic regression framework, NN-3 instead uses a more complex NN method for the final step. The key difference is that, while NNs can update the weights in lower layers using backpropagation, EMPaSchiz cannot. For each subject, the learned NN-3 model takes a concatenated feature vector of 84 feature vectors derived from EMPaSchiz framework, and consists of four fully-connected layers of dimensions, of sizes: 1000, 100, 40 and 2.

Figure 3.2 shows the architectures of these 3 NN models. (A) NN-1: Two convolution layers with 4 parameters depicted in order – kernel size, padding, input channel numbers, and the number of filters. This is followed by max-pooling layer with 2 parameters – kernel size and stride. Then, the linearized output of the max-pooling layer before the LSTM layer has 2 parameters – the number of layers and the number of units in each layer. (B) NN-2: Model parameters are depicted in a format similar to NN-1. After max pooling, two fully connected layers with 1 parameter – the number of units. (C) NN-3: Four fully connected layers with 1 parameter – the number of units. For all of the methods, the final output layer has two nodes – OCD versus HC. (For each test subject, the learned system will return the argmax of the values computed here.)

For each NN model, we used Relu as an activation function for each layer and cross-entropy as a loss function. To reduce the risk of overfitting, 50% of layers were dropped out during the training time. We used a maximum of 1000 epochs to train our models, with early-stopping criteria for 100 epochs – i.e., we calculated the validation error after each training epoch, and if the error was found to be not decreasing for a span of 100 epochs, then the training state was reverted back by 100 epochs. Models were implemented in PyTorch (v1.0.1) [99], and trained on a computer with Intel(R) Xeon(R) CPU E5-1660, 16GB RAM and a 12GB NVIDIA TITAN Xp GPU, and we chose a simplistic version of VGG-16 [126] as the full architecture was extremely memory constraining, as it involved training on hundreds of 3 and 4-dimensional tensors [127].

3.5 Cross-Diagnosis transfer learning

As mentioned above, Kalmady et al. [63] used EMPaSchiz to distinguish Schizophrenia patients from Healthy Controls; we considered a version that selected, and used, only a subset of SSMs (This reduced some overfitting). Here, we ask whether these SSMs (that were sufficient for Schizophrenia diagnosis) are sufficient for diagnosing another psychiatric disorder, OCD. In particular, we limited the single-source models (SSMs) of EMPaSchiz – the version of EMPaSchiz that was trying to learn model for distinguishing OCD from HC – to only the ones that were selected in the learned model of schizophrenia prediction. That earlier model [63] was based on 81 schizophrenia patients and 93 HC. Our OCD dataset was also collected from the same site, which initially included 175 OCD and 175 HC. In order to avoid any bias raising due to overlapping subjects in the HC group, we specifically excluded the controls who were included in the schizophrenia study, leaving 88 HC. The EMPaSchiz learner, on 81 schizophrenia patients and 93 HC (results for schizophrenia prediction model is available elsewhere [63]), used L1-regularization techniques at the final layer; this selected only 10 of the 84 SSMs; we then ran that learner with L2-regularization, using only those selected 10 SSMs to learn an OCD

prediction model, based on a dataset of 175 OCD patients and 88 controls, disjoint from the HC used for producing the Schizophrenia model (SCZ_to_OCD transfer model). We also present the results of the transfer model that includes all the healthy controls (SCZ_to_OCD_CommonHC); the sample distributions of these analyses are provided in Table 3.4.

Model	Learning		Prediction	
	Patients	HC	Patients	HC
SCZ_to_OCD	81 (SCZ)	93	175 (OCD)	88
SCZ_to_OCD_CommonHC	81 (SCZ)	181	175 (OCD)	181

Table 3.4: Sample distribution in Transfer learning model.

3.6 Evaluation Phase

We evaluated the learned models in 5 shuffled iterations of a 5-fold balanced cross-validation approach (80% training set, 20% test set; for a total of 25 train-test splits), estimating the model's generalization performance on the held-out fold, in terms of:

1. Accuracy: Fraction of correct predictions (of OCD versus HC) made by the learned model.

2. Precision: The number of correct predictions of OCD = trues divided by the total number of predictions of Class = OCDs.

3. Sensitivity: The number of correct predictions of OCD = trues divided by the total number of OCDs.

4. Specificity: The number of correct predictions of Class = Controls divided by the total number of Controls.

For each variant, we report the mean and standard errors for these metrics overall 25 train-test splits. Also, we report the mean and standard errors for elements of confusion matrices for the 5 iterations.

3.7 Results

3.7.1 OCD Prediction

EMPaSchiz algorithm was able to predict OCD with 80.3% accuracy using the 5 times 5-fold cross-validation. The model was 82.7% sensitive, 79.2% precise and 77.8% specific. Table 3.5 shows the results for the performance of EMPaSchiz algorithm and also for sub-models that stack SSMs only from specific feature extractions. EMPaSchiz's performance was significantly better than any of those subset-stacked models (compared to the best subset-stacked model: stacked-FC-prec at 77.9%, t-test, $p = 0.018$). Figure 3.3 shows a comparative profile of accuracies for various SSM predictors, parcellation-wise stacked models, and EMPaSchiz.

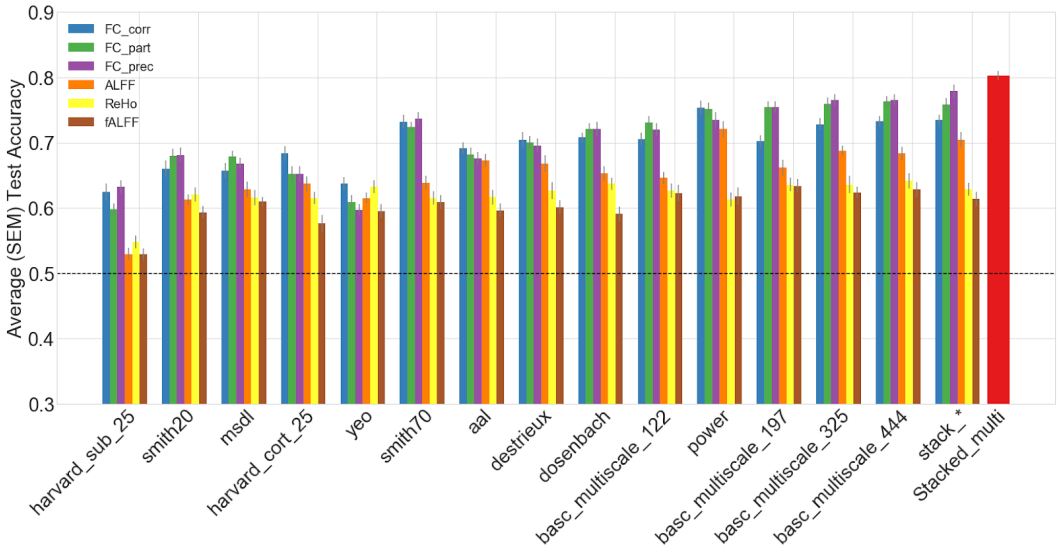


Figure 3.3: Comparison of 5 x 5-fold cross-validation prediction accuracies for single-source models and EMPaSchiz. (The error-bar corresponds to standard error of mean).

3.7.2 Comparison of EMPaSchiz to Neural Networks methods

We compared the performance of EMPaSchiz model to NN techniques that do not use their feature types and/or feature compression (parcellations) methods. Results (refer to Figure 3.4) show that EMPaSchiz model outperforms these

	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
EMPaSchiz	80.3 (0.7)	79.2 (1.0)	82.7 (0.9)	77.8 (1.4)	144.8 (1.5)	136.2 (1.0)	38.8 (1.0)	30.2 (1.5)
Stacked-ALFF	70.5 (1.2)	70.3 (1.2)	71.2 (1.7)	69.7 (1.6)	124.6 (2.8)	122.0 (2.8)	53.0 (2.8)	50.4 (2.8)
Stacked-fALFF	61.4 (1.1)	62.2 (1.2)	59.0 (1.6)	63.5 (1.8)	103.6 (2.0)	111.2 (2.2)	63.8 (2.2)	71.4 (2.0)
Stacked-ReHo	62.9 (1.0)	63.6 (1.2)	61.7 (1.3)	64.0 (1.9)	108.0 (1.5)	112.0 (0.9)	63.0 (0.9)	67.0 (1.5)
Stacked-FC	73.5 (0.8)	72.4 (1.0)	76.6 (1.1)	70.4 (1.4)	134.0 (1.4)	123.2 (1.3)	51.8 (1.3)	41.0 (1.4)
-corr								
Stacked-FC	75.9 (1.0)	74.4 (1.2)	80.2 (1.3)	71.5 (1.9)	140.4 (1.6)	125.2 (1.6)	49.8 (1.6)	34.6 (1.6)
-part								
Stacked-FC	77.9 (1.0)	75.0 (1.1)	84.2 (1.5)	71.5 (1.5)	147.4 (1.3)	125.2 (0.4)	49.8 (0.4)	27.6 (1.3)
-prec								
Baseline	50.00	—	—	—	—	—	—	—

Table 3.5: EMPaSchiz performance on OCD prediction.

NN-1 methods (paired t-test, $p < 0.001$), NN-2 methods for each feature type (ALFF: t-test, p = 0.011 , fALFF: paired-t-test, p < 0.001, ReHo: paired t-test, p = 0.005) and NN-3 (paired t-test, $p < 0.001$). Table 3.6 presents the 5×5-fold cross-validation prediction performance of the different NN models.

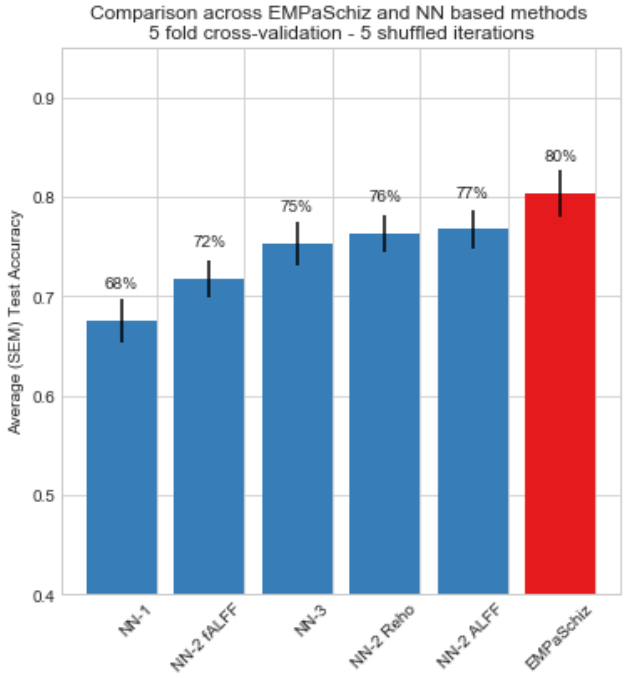


Figure 3.4: Comparison of performance across EMPaSchiz and Neural Network methods.

3.7.3 Transfer learning: feature selection based on Schizophrenia model

To deal with the Schizophrenia task, EMPaSciz's L1-regularization selected only 10 out of 84 SSMs; our SCZ_to_OCD transfer model used only these SSMs when dealing with the OCD task. This included only two feature types - Functional Connectivity Pearson Correlation and Precision. Only 9 of the 14 parcellations included aal, dosenbach, harvard_cort_25, msdl, power, basc_multiscale_122, basc_multiscale_197, basc_multiscale_325, basc_multiscale_444. More information is provided in Table 3.12. Note only 10 of the 9×2=18 possible SSMs were used in learning the OCD diagnosis model.

When we restricted our OCD learning model to ensemble only these 10

Table 3.6: Comparison of performance across EMPaSchiz and Neural Network methods.

	Accuracy (%)	**Precision (%)**	**Sensitivity (%)**	**Specificity (%)**	**True positive**	**True negative**	**False positive**	**False negative**
NN-1	67.5 (1.2)	70.2 (3.0)	66.7 (1.1)	70.4 (2.0)	122.8 (4.1)	113.4 (2.6)	52.2 (4.1)	61.6 (2.6)
NN-2 ALFF	76.7 (1.0)	76.9 (1.3)	77.6 (1.6)	75.9 (1.8)	135.8 (2.1)	132.8 (2.8)	42.2 (2.8)	39.2 (2.1)
NN-2 fALFF	71.7 (0.9)	73.3 (1.2)	69.3 (1.3)	74.2 (1.6)	121.2 (2.1)	129.8 (2.4)	45.2 (2.4)	53.8 (2.1)
NN-2 ReHo	76.3 (0.9)	77.0 (1.0)	76.0 (1.8)	76.7 (1.6)	133.0 (3.6)	134.2 (1.5)	40.8 (1.5)	42.0 (3.6)
NN-3	75.3 (1.1)	75.3 (1.1)	75.4 (1.8)	75.1 (1.4)	132.0 (2.7)	131.4 (1.6)	43.6 (1.6)	43.0 (2.7)
EMPaSchiz	80.3 (0.7)	79.2 (1.0)	82.7 (0.9)	77.8 (1.4)	144.8 (1.5)	136.2 (1.0)	38.8 (1.0)	30.2 (1.5)
Baseline	50.00	—	—	—	—	—	—	—

Table 3.7: EMPaSchiz transfer learning results for OCD diagnosis prediction with feature selection based on Schizophrenia model.

	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
SCZ_to_OCD	93.1 (0.6)	94.9 (0.6)	94.7 (0.5)	89.8 (1.2)	165.8 (0.7)	79.0 (0.8)	9.0 (0.8)	9.2 (0.7)
Original EMPaSchiz	91.8 (0.6)	90.1 (0.7)	98.7 (0.5)	77.9 (1.8)	172.8 (0.4)	68.6 (0.8)	19.4 (0.8)	2.2 (0.4)
N: OCD 175, HC 88; Baseline Accuracy: 66.53								

Table 3.8: EMPaSchiz transfer learning results for OCD diagnosis prediction with feature selection based on Schizophrenia model.

	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
SCZ_to_OCD_CommonHC	76.0 (0.9)	73.6 (1.1)	80.8 (1.3)	71.4 (1.7)	141.4 (1.1)	129.2 (1.4)	51.8 (1.4)	33.6 (1.1)
Original EMPaSchiz	78.8 (1.0)	78.0 (1.3)	80.2 (1.3)	77.3 (1.8)	140.4 (1.8)	140.0 (1.5)	41.0 (1.5)	34.6 (1.8)
SCZ_to_OCD_CommonHC (OCD 175, HC 181) Baseline Accuracy: 50.84								

Table 3.9: EMPaSchiz results for schizophrenia diagnosis prediction with feature selection based on OCD model.

	Accuracy (%)	**Precision (%)**	**Sensitivity (%)**	**Specificity (%)**	**True positive**	**True negative**	**False positive**	**False negative**
OCD_to_SCZ[a]	85.8 (1.2)	89.1 (1.6)	79.8 (1.9)	91.0 (1.4)	64.6 (0.5)	84.6 (1.5)	8.4 (1.5)	16.4 (0.5)
Original EMPaSchiz[a]	85.7 (1.0)	90.9 (1.4)	77.8 (2.0)	92.7 (1.2)	63.0 (0.4)	86.2 (0.5)	6.8 (0.5)	18.0 (0.4)
OCD_to_SCZ_CommonHC[b]	86.9 (0.7)	82.2 (1.2)	73.9 (2.0)	92.7 (0.6)	59.8 (0.4)	167.8 (1.0)	13.2 (1.0)	21.2 (0.4)
Original EMPaSchiz[b]	83.2 (0.7)	84.8 (2.0)	56.8 (1.9)	95.0 (0.8)	46.0 (1.1)	172.0 (0.7)	9.0 (0.7)	35.0 (1.1)

a OCD_to_SCZ (N: SCZ 81, HC 93) Baseline Accuracy: 53.44
b OCD_to_SCZ_CommonHC (N: SCZ 81, HC 181) Baseline Accuracy: 69.08

Table 3.10: Symptoms severity prediction results using EMPaSchiz.

	Baseline (%)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
Obsession (70 subjects)	51.43	58.57	56.41	64.71	52.78	22	19	17	12
Compulsion (78 subjects)	52.56	64.10	62.85	59.46	68.29	22	28	13	15

Table 3.11: Symptoms prediction results focusing on the cortico-striato-thalamo-cortical (CSTC) circuit using EMPaSchiz.

	Baseline (%)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
Obsession (70 subjects)	51.43	58.57	61.29	52.78	64.71	19	22	12	17
Compulsion (78 subjects)	52.56	53.84	51.28	54.05	53.66	20	22	19	17

Model Name	Feature type	Parcellation type
SCZ_to_OCD	FC- Pearson Correlation	basc_multiscale_122, basc_multiscale_197, basc_multiscale_325, basc_multiscale_444.
	FC- Precision	aal, basc_multiscale_325, dosenbach, harvard_cort_25, msdl, power.
SCZ_to_OCD_ CommonHC	FC- Pearson Correlation	basc_multiscale_197, basc_multiscale_325, basc_multiscale_444, dosenbach, power
	FC- Precision	aal,basc_multiscale_122, basc_multiscale_197, basc_multiscale_325, basc_multiscale_444, dosenbach, harvard_cort_25, smith70

FC = Functional Connectivity

Table 3.12: Selected models for transfer learning from SCZ to OCD diagnosis prediction.

SSMs, the accuracy of OCD model was 93.1% (Table 3.7; note, however, we expect higher baseline performance 66.5% due to class imbalance: OCD 175, HC 88). Interestingly, we found that this SCZ_to_OCD model provided performance that is comparable to original EMPaSchiz that is re-trained on this dataset (accuracy: 91.8%) within statistical significance margin (2-sided t-test, p = 0.15) in OCD prediction. Table 3.8 provides results for SCZ_to_OCD_CommonHC transfer model. Also, for the sake of completion, the results (selected models and model performance) for transfer learning models from OCD to schizophrenia are also provided (Tables 3.13, 3.9). All of the above results are obtained from the deterministic training process. The results for the stochastic process during training are given in Appendix A.4.

Model Name	Feature type	Parcellation type
OCD_to_SCZ	RF- ALFF	basc_multiscale_444
	FC- Pearson Correlation	basc_multiscale_197, basc_multiscale_325, basc_multiscale_444, power, dosenbach, harvard_cort_25, smith70.
	FC- Precision	aal,basc_multiscale_122, basc_multiscale_197, basc_multiscale_325, basc_multiscale_444, destrieux,dosenbach, harvard_cort_25, msdl, power, smith70.
OCD_to_SCZ_ CommonHC	RF- ReHo	basc_multiscale_444.
	FC- Pearson Correlation	basc_multiscale_197, basc_multiscale_444, dosenbach, smith70.
	FC- Precision	basc_multiscale_444, power.
RF = Regional Feature, FC = Functional Connectivity		

Table 3.13: Selected models for transfer learning from OCD to SCZ diagnosis prediction.

3.7.4 Symptom Severity Prediction

OCD patients in our sample ranged widely in their symptom severity, which was measured using Y-BOCS scale for obsessions (integer values from 0 to 20) and compulsions (integer values from 0 to 20). For each, we used the first and last quartile of these scales to categorize the least, versus the most severely, symptomatic patients. We then used EMPaSchiz in leave-one-out cross-validation setup to predict the high-symptomatic patients against low-symptomatic ones (majority class baseline accuracy were close to 50%). We used leave-one-out cross-validation (rather than 5-fold) to deal with a low number of subjects that were available for this analysis. Prediction accuracy was

58.6% for obsessions and 64.1% for compulsions of OCD psychopathology. In view of subpar performance, we compared this performance to the model that focused on features derived only from the brain regions that are consistently implicated in OCD by meta-analysis of functional and structural neuroimaging studies [26], [84], [107], which included areas from the cortico-striato-thalamo-cortical (CSTC) circuit, namely, the orbitofrontal cortex, the anterior cingulate cortex, prefrontal cortex and the ventral striatum (specific regions are listed in Table 3.14 and Figure 3.5). Prediction accuracy with CSTC was 58.6% for obsessions and 52.6% for compulsions. Details are provided in Tables 3.10 and 3.11.

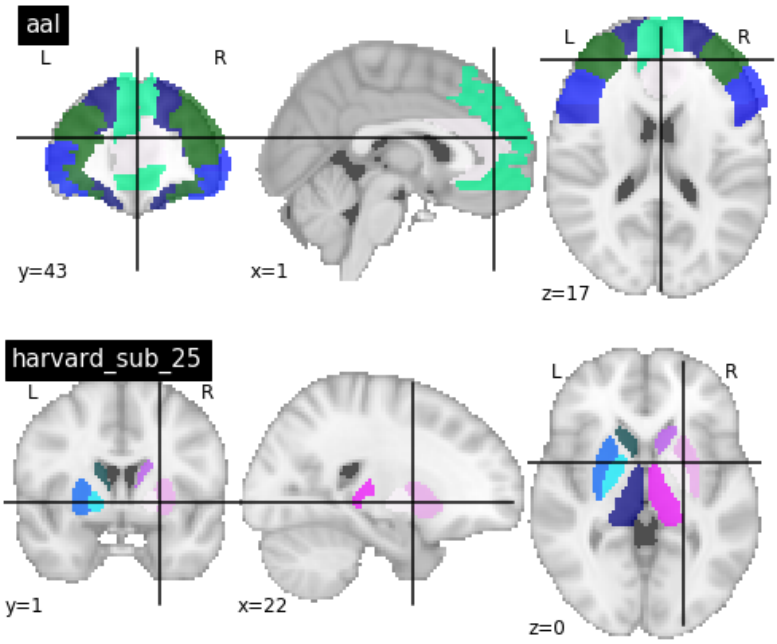


Figure 3.5: Regions selected for cortico-striato-thalamo-cortical (CSTC) circuit using AAL and Harvard-Oxford subcortical atlas.

Atlas	Regions of Interest	Labels
AAL	Prefrontal Cortex	Frontal_Sup_L,Frontal_Sup_R, Frontal_Mid_L, Frontal_Mid_R, Frontal_Inf_Oper_L, Frontal_Inf_Oper_R, Frontal_Inf_Tri_L, Frontal_Inf_Tri_R, Frontal_Sup_Medial_L, Frontal_Sup_Medial_R
	Orbitofrontal Cortex	Frontal_Sup_Orb_L, Frontal_Sup_Orb_R, Frontal_Mid_Orb_L, Frontal_Mid_Orb_R, Frontal_Inf_Orb_L, Frontal_Inf_Orb_R, Frontal_Med_Orb_L, Frontal_Med_Orb_R
	Anterior cingulate cortex	Cingulum_Ant_L, Cingulum_Ant_R
Harvard-Oxford subcortical atlas	Striatum	Left Thalamus, Right Thalamus, Left Putamen, Right Putamen, Left Caudate, Right Caudate, Left Pallidum, Right Pallidum, Left Accumbens, Right Accumbens

Table 3.14: Brain regions selected for the cortico-striato-thalamo-cortical (CSTC) circuit.

3.7.5 Ante-hoc Model Interpretability

To delineate key pathological alterations in OCD, we estimated the reliability of a regional feature's importance for diagnostic prediction, by sorting features by their respective mean logistic regression weight divided by the standard error for each feature in a particular learned SSM generated during 25 folds of cross-validation. The idea of this approach was taken from an earlier neuroimaging study [63].

Figure 3.6 (respectively Figure 3.7) highlights some of the top-most ($>$ 98 or 99th percentile) reliable features using representative atlases for regional resting state measures (respective connectivity). However, since our ensemble model is composed of 84 SSMs, these depictions should be considered just representative in nature, and cannot be claimed as the 'only' important features for OCD prediction.

Figure 3.6 panels show the top 98th percentile of top regional features

overlaid on glass brain: (A) Lower ALFF in L anterior cingulate, L rectus and R paracentral lobule (destrieux); (B) Higher fALFF in bilateral inferior parietal lobule and lower fALFF in L middle frontal gyrus, R insula and L paracentral lobule (basc_multiscale_197); (C) Higher ReHo in L middle frontal gyrus and lower ReHo in bilateral putamen, L middle temporal gyrus (basc_multiscale_122).

Figure 3.7 shows representative alterations in OCD as compared to HC that are suggested by the top most reliable features. Network edges show elevated (red) and suppressed (blue) changes in functional connectivity. Panels show the top 99th percentile of top functional connectivity features using the Dosenbach and msdl atlases. Within the Dosenbach atlas, we found decreased functional connectivity between the following pairs of regions: R anterior prefrontal cortex and L angular gyrus, R occipital gyrus and L posterior occipital gyrus, R frontal gyrus and both R intraparietal sulcus and R superior parietal lobule. Increased functional connectivity between the pairs of regions: interhemispheric occipital gyri, R dorsal anterior cingulate cortex and R parietal, L basal ganglia and L posterior parietal cortex. These differences suggest aberrations in fronto-parietal, cingulo-opercular, occipital and sensorimotor networks. Further, with the msdl atlas, we found decreased functional connectivity between regions: L superior frontal sulcus and L as well as medial default mode network (DMN), and R temporo-parietal junction and R parietal cortex. Increased functional connectivity between L inferior parietal sulcus and bilateral lateral occipital cortex as well as L visual cortex, R anterior insula and motor cortex, L insula and L auditory cortex. These alterations might imply aberrations in several distributed networks such as DMN, language, attention, visual, auditory, motor as well as salience networks.

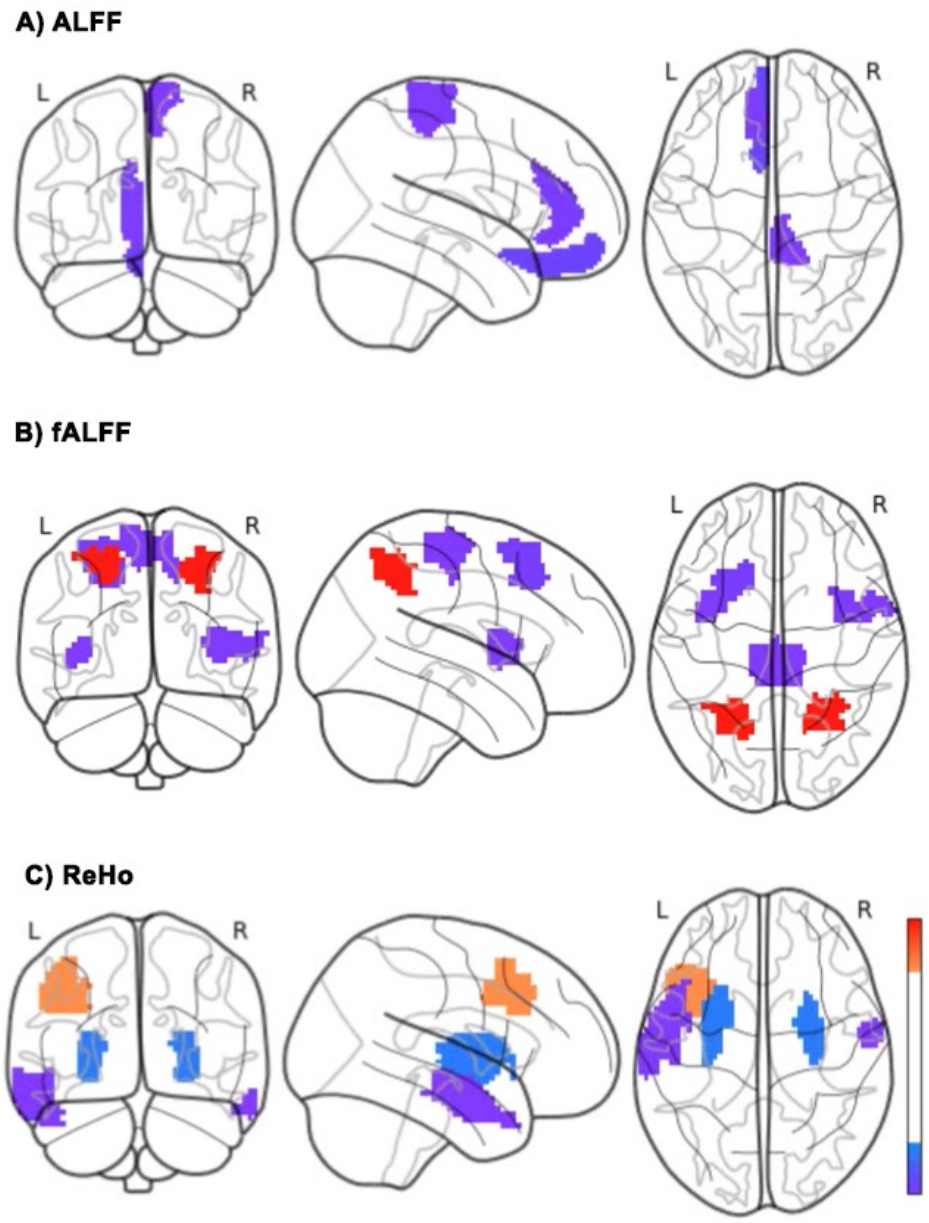


Figure 3.6: Key pathological alterations in OCD suggested by top-most reliable features – elevated (red/orange) and suppressed (blue/purple) changes in regional activity.

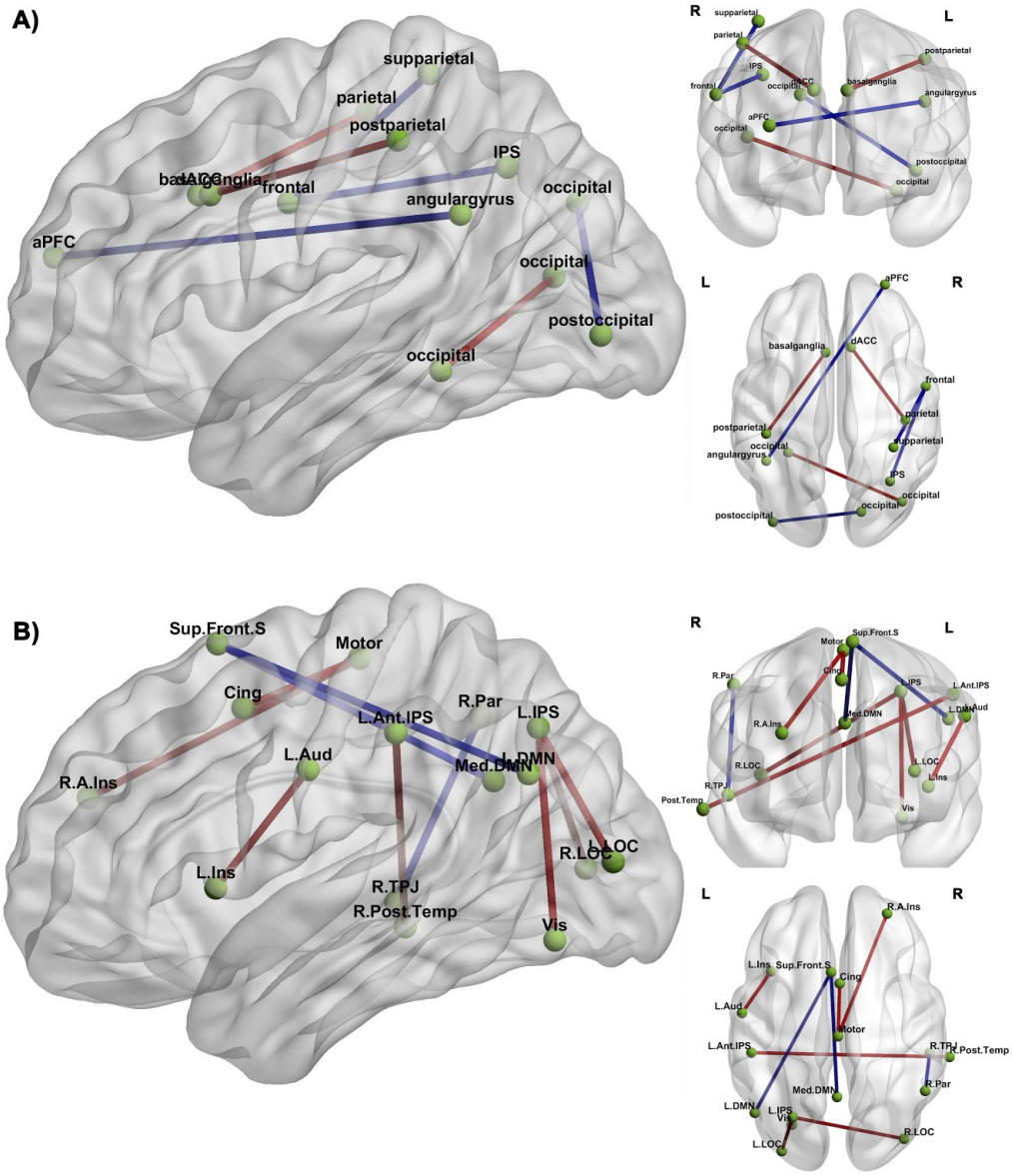


Figure 3.7: Key pathological alterations in OCD suggested by top-most reliable features – network edges show elevated (red) and suppressed (blue) changes in functional connectivity.

3.8 Other Experiments

We performed different experiments on OCD dataset, but unfortunately, we were unable to achieve the desired level of performance from these experiments. The brief descriptions of these experiments are given below, and the details of these experiments are given in Appendix A (Sections A.1 to A.3).

- **OCD Diagnosis Prediction:** For OCD diagnosis prediction, we added sMRI and DTI features in EMPaSchiz framework with fMRI features. We also used the neural network version of EMPaSchiz model.

- **Multi-class Classification Model:** We tried to build a multi-class classification model that could differentiate between healthy subject, SCZ, and first-degree relatives (FDRs) of SCZ, or healthy subject, SCZ, and OCD.

- **Symptom Severity Prediction:** We used different scales for measuring the severity, and based on each scale, we tried to learn a model using EMPaSchiz framework or Deep Probabilistic Canonical Correlation, which could distinguish between the least versus the most severely, symptomatic patients. We also aimed to learn a model that could distinguish OCD patients between washers and checkers. Moreover, we tried a multi-task regression model to predict symptom severity values.

3.9 Discussion

From the study observations, we conclude that our EMPaSchiz algorithm can predict OCD with 80.3% accuracy, which outperforms base models that uses any individual feature type or parcellation scheme. In this case, feature design based on prior neurobiological knowledge (parcellations) leads to better performance than agnostic and automated feature design (neural nets). Selection of single-source feature sets can be transferred from Schizophrenia to OCD prediction model without significant loss in prediction performance. EMPaSchiz provides a generalizable yet fairly interpretable linear model that uses human-expert understandable features and model structure.

As many psychiatric disorders usually manifest with a myriad of overlapping symptoms, reliable clinical diagnosis is a challenging task [1]. Here, we demonstrate a computational framework that (1) builds on prior neurobiological information that is derived from numerous neuroanatomical and neurophysiological studies, and (2) provides a diagnostic performance that matches

the trained psychiatrists about 8 out of 10 times. Earlier studies have shown that diagnostic tools such as DSM / ICD is not always reliable, and psychiatrists do not always agree on the diagnosis, with reported joint rater agreement as low as 0.2 to 0.4 (intraclass kappa) in some circumstances [110], [111]. Given this, one might argue that a machine with predictive accuracy of 80% in OCD diagnosis is 'close enough' for human-like performance, and hence the technology is worth exploring further with even bigger datasets and more sophisticated algorithms. However, this is not an apples-to-apples comparison, since clinician decisions are based on a consensus about clusters of clinical symptomatology, while machine learning models are based on fMRI images – which clinicians are not able to discriminate visually. However, it is desirable to base a clinical diagnosis on objective laboratory measurements (such as an fMRI image), as it strengthens the reproducibility of the disease entities [24]. Unfortunately, this is difficult here, given the earlier evidence of that such brain-derived measurements (like sMRI, fMRI) only weakly correspond with clinical classifications of psychiatric conditions [59].

With cross-validated prediction performance of 80.3%, EMPaSchiz ranks highest among all earlier OCD prediction studies, either using fMRI or other any neuroimaging modalities, with sample size greater than one-hundred subjects [54], [98], [132], [137], [162]. None of the earlier fMRI-based OCD prediction studies have combined the regional and connectivity features, nor have any employed an ensemble approach such as ours to jointly learn from multiple parcellations. Most used a single feature type (such as ALFF, fALFF, ReHo or Pearson Correlation-based functional connectivity [17], [48], [54], [125], [137], [162]), and only one of the pre-defined atlases (e.g. BrainVISA Sulci Atlas, Anatomical Automatic labeling atlas, Harvard-Oxford atlas) [122], [137]. A few other studies tried different approaches to handle the OCD diagnosis prediction problem (See Table 3.1 for details specific to each study).

While high performance on diagnostic prediction is encouraging, note that discriminating OCD patients from normal individuals is perhaps an easy task for trained psychiatrists, especially compared to identifying subtypes based on symptom dimensions or severity. Our experiments with learning models for

identifying severe cases were only mildly successful. Further, our experiments that used only on the CSTC brain regions (which are most commonly implicated in OCD literature [26], [84], [107]) showed performance that is inferior to EMPaSchiz's (which considered all brain regions from the 14 parcellations). Collectively, these results suggest that fMRI data might not have enough information to perform more complicated psychiatric decisions, at least within the scope of our training size and algorithms.

Our study faces the general limitation of most machine learning studies in psychiatry: the ground truth itself might be ill-defined, in terms of validity of current psychiatric classifications as unitary disease constructs. Further, a challenge in clinical practice is differentiating OCD from other disorders. Also, our data was collected using single MRI scanner and from individuals of fairly homogeneous ethnicity. We do not know whether it will perform similarly in multi-centric data collected across mixed population groups and comorbidities as well.

Machine learned models produce more accurate predictions as the size of the training dataset increases, hence it is critical to have a sufficient sample size [35]. Our study used 350 subjects (175 OCD + 175 HC), which is the largest OCD report dataset that has been employed in machine learning until now (the previously largest sample size is 204 (102 OCD + 100 HC), which reported prediction performance of 76.6% [132]). Our study provides two other novel observations. First, it shows that a simple linear model with neurobiology-informed features outperforms complex neural network models, even though those models can automatically design new features that can potentially exploit non-linear interactions. Hence, the EMPaSchiz model is more accurate. Secondly, we demonstrate cross-diagnostic transfer learning in psychiatry applications. We show this can considerably reduce the complexity of data processing, in terms of feature engineering, by pre-training the model with subjects affected by a different psychiatric condition – schizophrenia in our case – and still retain the earlier performance accuracy. We hypothesize that such transfer learning works in this context because the underlying brain abnormalities are common across multiple psychiatric disorders [44].

Chapter 4

Analysis of Schizotypy in First-degree Relatives of Schizophrenia

4.1 Introduction

Genetic inheritance plays a strong role in the etiology of schizophrenia, representing approximately 80% of the liability for the illness, based on numerous twin and adoption studies [78], [80], [134]. Recent studies demonstrated that first degree relatives (FDRs) of patients with schizophrenia are more likely to exhibit associated intermediate phenotypes or 'endophenotypes', than the general population, even when they do not (or do not yet) present with a full set of clinical symptoms [46]. Numerous endophenotypes have been proposed in schizophrenia, including brain structural or functional patterns, sensory processing measures, neuromotor and neuropsychological measures, minor physical anomalies [4], [47]. In this context, it is interesting to investigate how such populations would be classified by a machine-learned model that is capable of distinguishing schizophrenia (SCZ) patient from healthy control (HC) based on resting-brain activation patterns. This study examines whether a schizophrenia diagnosis model, learned using schizophrenia and normal fMRI datasets, can identify higher schizotypal scores in first degree relatives without schizophrenia.

Schizophrenia spectrum disorders (SSD) present a challenge in categorizing disease phenotypes, due to a wide range of overlapping symptoms and a

heterogeneous illness course at the individual level. Schizotypal personality traits, aka schizotypy, which resemble the signs and symptoms of schizophrenia in the general population, provide a spectral rather than categorical view, due to a continuous range of traits and symptoms [27]. In recent years, there is increased interest in learning models from functional neuroimaging to predict schizotypy – with unfortunately limited generalizability due to small training samples and lack of independent validation (for review, see [76]).

The current study explores an alternative approach for diagnostic prediction with unaffected FDR subjects by applying the machine-learned model that EMPaSchiz [63] developed, when it was run on an independent resting-state fMRI dataset of 81 antipsychotic-naive schizophrenia patients and 93 healthy controls (the basic demographic information of these subjects is provided in Table 3.3, and in this chapter, we named this dataset as "SCZ_HC" dataset). Given strong evidence for familial aggregation of schizotypy in schizophrenia spectrum disorders [131], we hypothesize that FDR subjects that this model (mis)predicted have 'schizophrenia' status will have significantly higher schizotypal scores, versus those who are predicted as non-schizophrenia status by machine learning.

4.2 Dataset and Preprocessing

The process of sample acquisition for FDR dataset is the same as Section 3.2.2. This study examined 57 FDRs (M:F = 42:15) based on the following inclusion and exclusion criteria. We included siblings or children of schizophrenia patients, without any axis-1 disorder as evaluated by the Mini International Neuropsychiatric Interview (MINI) Plus [124]. Probands of these FDRs were patients attending the clinical services of the National Institute of Mental Health & Neurosciences (NIMHANS), India, who fulfilled DSM-IV criteria for schizophrenia. The Structured Interview for Psychosis-risk Syndromes (SIPS) scale [79] was administered to ascertain that FDRs were unaffected by active psychosis. SIPS impression for these subjects was 'Genetic Risk and Deterioration Psychosis-Risk Syndrome'. NIMHANS research group recruited only

right-handed subjects to avoid potential confounds of differential handedness. No study subjects had contraindications to MRI or medical illness that could significantly influence brain structure / function, such as seizure disorder, cerebral palsy, or history suggestive of delayed developmental milestones. There was no history suggestive of DSM-IV psychoactive substance dependence or of head injury associated with loss of consciousness longer than 10 min. No subjects had abnormal movements as assessed by the Abnormal Involuntary Movements Scale. Pregnant or postpartum females were not included. The age range was 17 to 38 years (27.2 $\pm$ 5.25 years). A 22-item self-reported screening measure of schizotypal personality traits - "Schizotypal Personality Questionnaire – Brief" (SPQ-B) [108] – was used to assess the schizotypal personality score for each subject. The catchment area for the subject recruitment involved the southern states of India. NIMHANS research group obtained informed written consent after providing a complete description of the study to all the subjects. The NIMHANS ethics committee reviewed and approved the original research protocol. The Research Ethics Board at the University of Alberta, Edmonton approved the secondary analysis of archived data.

Initially, they (NIMHANS research group) visually inspected the acquired images for artefacts such as incomplete brain coverage or ghosting, and no subject was removed from this step. After that, they discarded the first ten volumes of each functional time-series before reaching steady magnetization and to allow the participants to adapt to the scanning noise. Then, they followed the fMRI preprocessing pipeline that is described in Section 2.2.1.

4.3 Methods

Initially, we did the feature extraction for each dataset (FDR and SCZ_HC datasets) by following the procedure that is described in Section 3.3.2. After that, we trained EMPaSchiz model (briefly described in Section 3.4.1) [63] on SCZ_HC dataset (81 SCZs + 93 HCs). Then, we applied the learned EMPaSchiz model to classify each FDR either as schizophrenia patient or

healthy individual, and examined if there is a class difference, between the patients and healthy individuals, in distribution of SPQ-B [108] scores. Note that none of the FDR subjects in this study were in the training set used to produce the EMPaSchiz model (as a schizophrenia patient or healthy control).

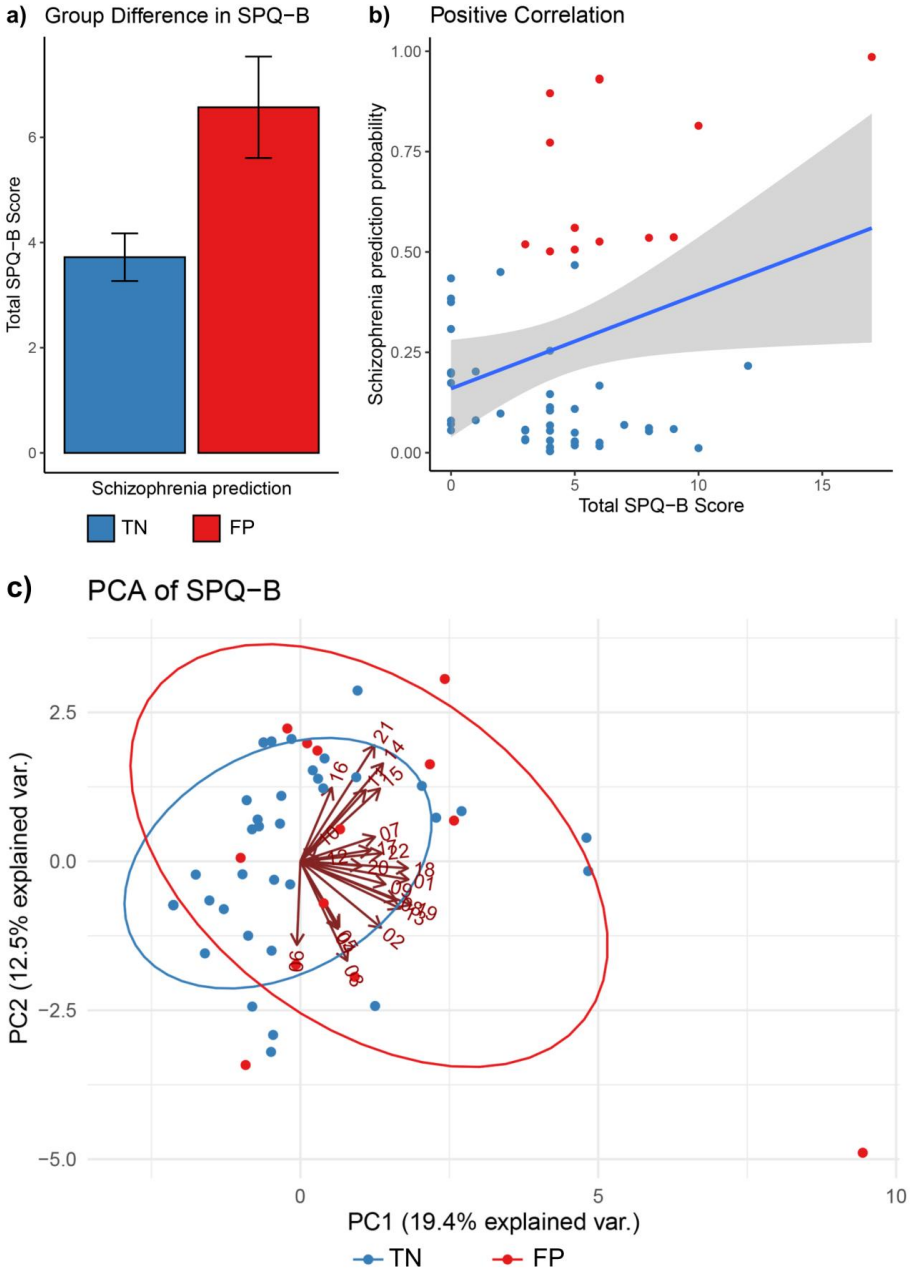


Figure 4.1: Relationship between schizophrenia prediction and SPQ-B score.

4.4 Results

This model classified 14 out of 57 FDR subjects as schizophrenia (FDR-SCZ), while the remaining 43 were classified as healthy controls (FDR-HC), based on the default threshold level of schizophrenia prediction probability > 0.5. We found that the FDR-SCZ group had a significantly higher total SPQ-B score" than that of FDR-HC (t =2.67, p = 0.01, Figure 4.1a); similarly, there was a significant positive correlation between probability of schizophrenia class and total SPQ-B score (Pearson's r = 0.28, p = 0.03, Figure 4.1b). FDR-SCZ and FDR-HC groups did not differ significantly on age (t = 1.02, p = 0.31) or sex distribution ($\chi^2 = 0.32$, p = 0.57).

To understand the effect of this machine classification further in relation to the latent structure of the SPQ-B questionnaire, we conducted a principal component analysis (PCA) of the 22 SPQ items. Figure 4.1c shows the biplot of two components of PCA (PC1 and PC2) along with the loadings of individual items as numbered arrows and the ellipses corresponding to FDR-SCZ or FDR-HC groups. In general, we observed that FDR-SCZ tended to show higher scores on PC1 comprising of items such as 'unable to get close to people' (item-18), 'people find me aloof and distant' (item-1), 'often pick up hidden threats' (item-9) and 'tend to keep my feelings to myself' (item-22).

4.5 Discussion

Schizotypy is considered as a marker of vulnerability for schizophrenia that runs within families [131]. Furthermore, it provides a useful framework to investigate the aetiological factors of schizophrenia spectrum disorders [10]. This study, for the first time, demonstrates a cross-application of a machine-learned schizophrenia diagnostic model in identifying subjects with high levels of schizotypy. However, whether similar prediction performance holds for a larger population without familial association remains to be explored. Further application of this approach holds significant promise for exploring related and comorbid symptom clusters in psychiatry.

Chapter 5

Schizophrenia Prognosis Prediction

5.1 Introduction

Schizophrenia is a devastating psychiatric disorder, including cognitive impairments, delusions, lack of motivation, hallucinations, and disabling [155]. The heterogeneous property of this disorder comes from the environmental factors and overlapping between the underlying biology of diseases. Early-onset Schizophrenia becomes a large burden due to its chronic behavior as well as the long-term poor treatment outcomes [155].

The primary treatment of this disease is antipsychotic drugs, and this kind of treatment contributes to the brain changes during the schizophrenia progression [38], [151]. However, the treatment is partially effective in many cases [42]. Some characteristics of Schizophrenia patients are resistant to antipsychotic medication treatment – in particular, hallucinations [88], [103], insight [15], cognitive deficits [94], [97] and others [3], [82]. But, recent studies reported that these antipsychotic treatment-resistant patients show significant improvement in their condition when treated with the transcranial Direct Current Stimulation (tDCS) treatment [14], [64], [70].

Treatment goals for schizophrenia include identifying the illness as early as possible, effectively treating core symptoms, and maintaining clinical improvement to prevent relapses [100]. The heterogeneous properties among the patients make the goal more challenging. There is inter-patient variabil-

ity in responses to the same treatment, and this also applies to debilitating side-effects that often lead to poor treatment compliance [87]. It follows that there is a pressing need to develop a personalized medicine system in the psychotic research field. We believe this could be greatly beneficial, as it can help clinicians to make important clinical decisions correctly – such as 'what drug/medication to prescribe'.

To advance the personalized treatment, researchers are working on building evidence-based approaches using various MRI modalities and clinical information. High-resolution structural imaging provides anatomical information pertaining to gray and white matter volumes, while resting state functional MRI provides indicators of spontaneous neuronal oscillations and diffusion tensor imaging provides fractional anisotropy that reflects on integrity of white matter tracts of the brain. Most researchers have tried to find a correlation between these neuroimaging biomarkers and the treatment response.

In a study on 107 first-episode schizophrenia patients and 20 healthy controls, Lieberman et al. [72] studied the relationship between ventricular volume and at least 12 months of clinical outcomes of antipsychotic treatment using structural MRI (sMRI) data. Despite any significant reduction in cortical and hippocampal volumes during the treatment, they found that patients with poor treatment outcomes showed ventricular volume enlargement. However, control subjects and patients with better treatment outcomes did not show this property. Ho et al. [51] also found similar findings using 73 schizophrenic patients and 23 controls. In addition to these, different studies showed the relationship between the treatment outcomes and the different areas of sMRI. For patients who responded better to treatment, frontal and limbic grey matter show higher densities [60], parahippocampal cortex has a larger volume, and there is a positive relationship between the changes of temporal grey matter volumes and the better outcomes [12], [13]. On the other hand, for patients with poorer treatment responses, grey volumes in the middle frontal gyrus and the temporal cortex are reduced [105], frontal cortical asymmetry shows abnormal patterns compared to healthy controls, and better responded patients [135].

Along with the use of sMRI, few research projects studied the changes in DTI measures concerning the treatment outcomes. Luck et al. [73] found the reduction in fractional anisotropy (FA) in the superior longitudinal and uncinate fasciculi for the poorer responded patients, and Zeng et al. [166] found that patients with better treatment outcomes show the FA increase in superior longitudinal fasciculus. Several studies also focused on using functional MRI to investigate brain activation changes during the treatment. Fahim et al. [36] found the significant activation increase in the dorsolateral prefrontal cortex, anterior cingulate cortex, and striatum, and similarly, Kumari et al. [69] found that high dorsolatral prefrontal activity for the responded patients using task-based fMRI data. In a study on 24 first-episode schizophrenia patients, Sarpal et al. [118] and Kragulijac et al. [66] found functional connectivity increase in the striatum, prefrontral regions and hippocampus, and Sambataro et al. [117] observed increase connectivity in Default mode network for the successful treatment.

While a greater number of research works have focused on doing association studies, there are a few numbers of studies that tried to build a machine-learned model that can predict the treatment response. Cao et al. [20] produced a treatment response method on 43 drug-naive first-episode schizophrenia patients, that achieved 82.5% accuracy in predicting the antipsychotic treatment response. They computed mutual information and functional connectivity correlation between bilateral superior temporal cortex and all of the cortical regions, and then passed these extracted features to support vector machine.

Inspired by all of these previous works, we tried to build a model to predict the treatment response for two types of treatments (tDCS treatment, and antipsychotic medication). In Section 5.2, we tried convolutional neural network (CNN) based different deep learning methods, and neuroanatomical and neurofunctional knowledge based different methods for the tDCS treatment response prediction. We found that prior neurobiological knowledge based method gives better performance compared to other methods. We achieved 77.5% accuracy from the prior neurobiological knowledge based method. In Section 5.3, for the antipsychotic treatment response prediction, we also used

similar CNN based deep learning methods. We have used 3 fMRI datasets (CGI, SAPS/SANS, and Hallucination datasets) along with the clinical information. We achieved around 64% accuracy on SANS dataset to predict the treatment response, but unfortunately, we were unable to achieve significant results for other datasets.

5.2 tDCS Schizophrenia Treatment Response Prediction

5.2.1 Dataset

We used a dataset from National Institute of Mental Health Neu-rosciences (NIMHANS), India, which includes fMRI and clinical data – including baseline and follow-up PSYRATS Auditory Hallucination (PsyAH) scores. We have a total 34 patients in this dataset with baseline (BL) and follow-up (FU) PsyAH scores, taken (respectively) at the time of tDCS administration, and 5 days afterwards. We used the percentage of improvement (PI) in terms of PsyAH scores for categorizing the least, versus the most, improvement patients using PI_PsyAH= $\frac{\text{BL\_PsyAH- FU\_PsyAH}}{\text{BL\_PsyAH}}$, where BL_PsyAH denotes baseline PsyAH scores and FU_PsyAH denotes follow-up PsyAH scores. After that, we defined tDCS responders as PI_PsyAH ≥ 0.25, otherwise non-responders. Our dataset had an even split of 17 responders and 17 non-responders by this criteria, thereby yielding a chance-level prediction performance of 50%. The basic demographic information of these subjects is provided in Table 5.1.

Characteristic	Responded	Not Responded	Stat	p
N	17	17		
Sex [M:F]	7:10	12:5	1.90^b	0.16
Age	30.06 ± 7.89	32.23 ± 7.47	0.80^a	0.42

a Independent Sample Test [t]
b Chi-Square test [χ^2]

Table 5.1: Demographic profile of tDCS-SCZ subjects.

5.2.2 Data Preprocessing and Feature Extraction

Data Preprocessing

Clinical Features We have 95 clinical features, some (e.g., features like weight, height, and weight circumference) of which were missing completely at random manner because the missing values have no relationship with either observed or missing values. That is why we used Multivariate Imputation by Chained Equation (MICE) that assumes data are missing at random [19] to impute the missing values.

fMRI data Preprocessed 4D fMRI data were provided by NIMHANS research group, India. They followed the fMRI preprocessing pipeline, which is described in Section 2.2.1.

Feature Extraction: Seed-based Functional Connectivity Features

Features were extracted using seed-based functional connectivity with two seed regions: left temporoparietal junction (LTPJ) and right dorsolateral prefrontal cortex in the fronto-parietal network (RDLPFC). We used a 15 mm radius sphere around MNI coordinates (-50, -34, 20) and (43, 21, 38) for LTPJ and RDLPFC, respectively [8], [144]. Then, we computed the Pearson Correlation (PC) between the mean time series of seed points with individual time series from all other voxels in the brain using $PC(v_i, v_j) = \frac{\sum(v_i - \overline{v_i})(v_j - \overline{v_j})}{\sqrt{\sum(v_i - \overline{v_i})^2 \sum(v_j - \overline{v_j})^2}}$, where v_i denotes the mean time series of i-th seed point, v_j denotes the time series of j-th voxel in the brain, and $\overline{v_k}$ denotes the mean of all the time points values of k-th voxel in the brain. The correlation values were then normalized using Fisher Z-transformation formula [90]. This procedure yielded a scalar feature value per voxel of the brain, generating a 3D feature matrix of size 91 x 109 x 91 for each seed-point.

5.2.3 Methods

Selected Regions of Brain based Method (SRM)

We extracted raw feature values from voxels (not average of all the voxels in a region) belonging to the apriori selected regions of the brain based on their neurobiological basis in pathogenesis of auditory hallucinations [62], [67], [113], [169], using Harvard cortical and subcortical atlases [1]. These regions are listed in Appendix B - Table B.2 along with references. We concatenated all the extracted raw features to yield a single vector for each subject. These vectors were passed to the learner model – L1-regularized logistic regression model (with fixed regularized value, C = 1).

Deep Learning Methods

As all the extracted features from the 4D fMRI data are three dimensional, we used a 3D convolutional deep neural network. The general framework of this neural network is depicted in Figure 5.1, where Panel A shows the main classification model (we took the idea of this 3D CNN classification model from the VGG-16 model [126], which was proposed for 2D data), and Panel B shows the architectures of how we added the clinical features to the main classification model.

We also tried transfer learning with convolutional neural network (CNN). The details of the deep learning models with and without transfer learning are described below:

Without Transfer Learning Our extracted LTPJ and RDLPFC seed-based Pearson Correlation (PC) features are two 3 dimensional matrices. We considered these two feature types as 2 input channels of the CNN model. We used a 3D CNN model architecture, which is shown in Figure 5.1 (Panel A), for categorizing the level of treatment response. For this model, it takes a fixed input size of $2 \times 91 \times 109 \times 91$, here the first number denotes the number of input channels (= 2 channels), and other ones denote the size of

[1] `http://www.cma.mgh.harvard.edu/fsl_atlas.html`

the 3-dimensional feature matrix ($= 91 \times 109 \times 91$).

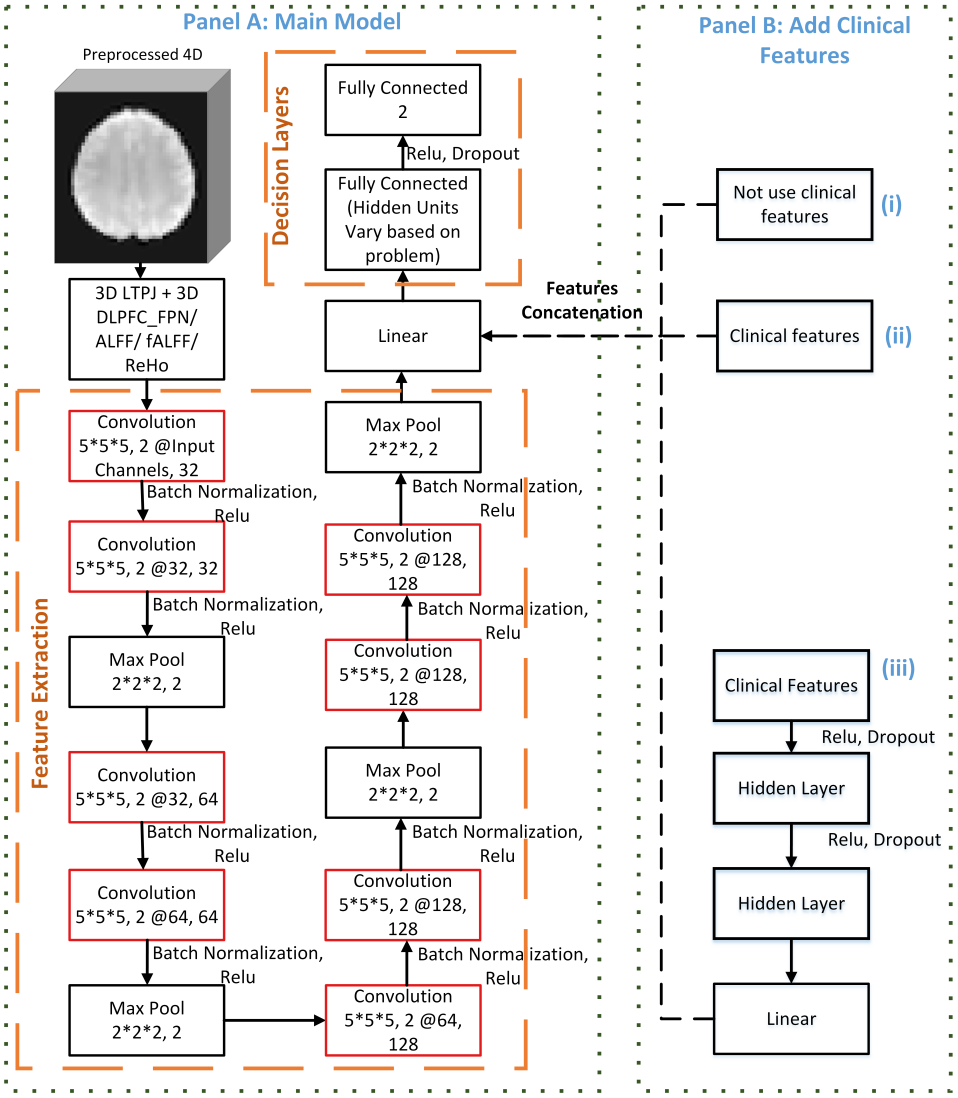


Figure 5.1: Neural network models for prognosis prediction: each convolution layer with 4 parameters depicted in order – kernel size, padding, input channel numbers and the number of filters. This is followed by max-pooling layer with 2 parameters – kernel size and stride. After the feature extraction step, each fully connected layer with 1 parameter – the number of units.

Transfer Learning Convolutional neural networks (CNNs) have become a state-of-the-art method for solving various prediction tasks in computer vision. However, CNNs are not efficient in performance in terms of accuracy when trained on smaller datasets. These kinds of situations are prevalent in clinical problems; for example, in our case, we have data from 34 tDCS pa-

tients for training and evaluating the model. So, we made use of the transfer learning method. In our case, we pretrained our CNNs with a different dataset, and then subsequently continued to train on our target dataset of 34 tDCS subjects. We anticipate this would done in order to get better representation of fMRI features in lower layers of CNN, which then would yield better initialization of model parameters for training on the target task.

For pretraining, we used different datasets of resting state fMRI images from several cohorts, including – healthy controls, obsessive compulsive disorder (OCD) and schizophrenia (SCZ), unaffected first degree relatives of schizophrenia patients (FDR-SCZ). We used the Clinical global impression score (CGI) of subjects taken at the time of fMRI acquisition as the label for the pretraining task. 186 healthy controls and 62 FDR-SCZ with CGI score $= 0$ were assigned the label of 0, whereas 44 SCZ and 149 OCD patients who were at least moderately ill (CGI score > 3) were assigned the label of 1. This yielded a total of 442 instances: 248 and 193 for classes 0 and 1, respectively (shown in Table 5.2). We oversampled instances manually in the minority class based on sex, age to balance the classes (here, two classes are not significantly different in terms of sex and age), and finally, this process gave 496 samples for pretraining. We chose CGI as a label for the pretraining task in order to capture the global changes in fMRI signals that are associated with psychiatric illnesses in general. Imaging and clinical assessments for these cohorts as well as the study subjects were conducted in the same medical center (NIMHANS). Note that this pretraining process did not use any subjects from the target dataset.

Our CNN architecture is depicted in Figure 5.1 (Panel A). After pretraining with 496 instances, we froze the first four layers of the feature extraction, replaced the decision layers, and then continued to train on the target dataset and target label. We report results for our CNN models - both with and without pretraining.

All the models were implemented in PyTorch (v1.0.1) [99], and trained on a computer with Intel(R) Xeon(R) Platinum 8168 CPU, 32GB RAM and a 32GB Tesla V100-SXM2 GPU.

Pretraining (N)	Label 0: CGI = 0	Label 1: CGI >3
Healthy Controls (186)	186	0
FDR - Schizophrenia (62)	62	0
OCD (149)	0	149
SCZ (44)	0	44
Total	248	193

Table 5.2: Sample distribution in transfer learning model for predicting tDCS treatment response.

5.2.4 Results

We evaluated the learned models in 5 shuffled iterations of a 10-fold balanced (with respect to class label) cross-validation approach (90% training set, 10% test set; for a total of 50 train-test splits), estimating the model's generalization performance on the held-out fold, in terms of accuracy. We also computed precision, sensitivity, and specificity. For each variant, we report the mean and standard errors for these metrics over all 50 train-test splits. Also, we report the mean and standard errors for elements of confusion matrices for the 5 iterations.

Table 5.3 presents the 5×10–fold cross-validation prediction performance of the different models. We achieved highest performance – 77.5% accuracy – using selected regions based method. But, without (respectively, with) transfer learning based methods provide 57.2% (respectively, 66.2%) accuracies, and the results show that transfer learning helps to improve the deep learning model performance significantly.

We also added the clinical features with each of the models, but unfortunately, we were unable to get any extra benefit from the additional features.

Finally, for our best-performed model (selected regions based method), we used SHAP values [74] to estimate the relative importance of features contributed by individual brain regions. Figure 5.2 lists important regions identified using top-1000 SHAP values. For finding these regions, we initially selected the 1000 voxels corresponding to top SHAP values. We then computed the percentage contribution of each region – how many voxels from the selected

Table 5.3: Comparison of performance across different methods for tDCS-SCZ treatment response prediction.

Model Type	Accuracy	Precision	Sensitivity	Specificity	True positive	True negative	False positive	False negative
Selected Regions based	**77.5 (3.1)**	77.6 (3.5)	87.0 (3.7)	68.0 (5.1)	14.6 (0.5)	11.6 (0.2)	5.4 (0.2)	2.4 (0.5)
Without Transfer Learning	57.2 (3.1)	62.5 (5.4)	48.0 (6.3)	64.0 (6.1)	8.0 (0.8)	11.4 (1.3)	5.6 (1.3)	9.0 (0.9)
With Transfer Learning	66.2 (3.0)	70.83 (4.0)	71.0 (5.1)	62.0 (5.8)	12.0 (0.4)	10.6 (0.8)	6.4 (0.8)	5.0 (0.4)

1000 voxels belong to each region.

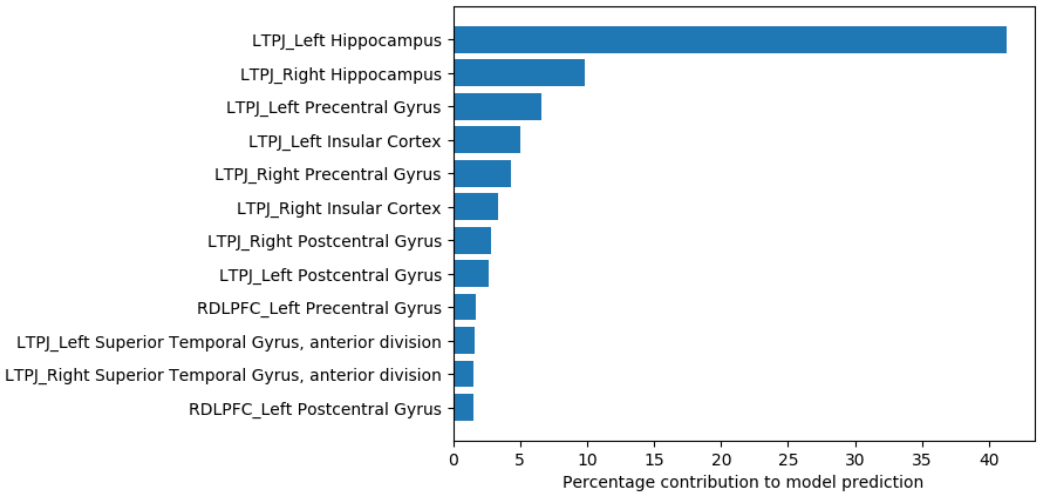


Figure 5.2: Percentage contributions of brain regions based on voxels with top 1000 SHAP values.

5.2.5 Other Experiments

To learn a classifier that could produce the prognosis prediction, we also tried different methods based on average and individual voxels features of each predefined region (here, average voxels features mean averaging all the voxels per region, whereas individual voxels features mean considering all the raw voxels features of each region). We used different ensembling methods (like meta-classifier using logistic regression or majority voting). Moreover, we implemented adaptive thresholding and feature concatenation methods. The details of these methods are given in Appendix B.1. In place of the Pearson correlation in SRM, we also used different measures: partial correlation, skewness, variance, and Kurtosis. From all of these experiments, we were unable to achieve significant improvement compared to SRM.

5.2.6 Discussion

In this part, our goal is to learn a classification framework that can accurately distinguish tDCS-SCZ patients with high treatment response from low treatment response. Along with the deep learning methods, we tried some prior neurobiological knowledge based methods. The empirical results show

that adding prior knowledge helped to improve the model performance for predicting the tDCS treatment response.

5.3 Antipsychotic Schizophrenia Treatment Response Prediction

5.3.1 Dataset

We have collected a dataset from National Institute of Mental Health & Neurosciences (NIMHANS), India, that includes fMRI and clinical data, including some baseline and some follow-up severity scores: CGI, SAPS/SANS, Hallucination. Based on the availability of these scores, we have divided the dataset into three sub-groups.

CGI Dataset

We have 99 patients in this dataset with baseline (BL) and follow-up (FU) CGI scores. These follow-up scores are collected between 2 to 4 months from the first visit. We have used two types of values – the percentage of improvement (PI_CGI) and the follow-score – for categorizing the least, versus the most improvement patients using the following ways.

1. Percentage of improvement: We computed the percentage of improvement by using PI_CGI= $\frac{\text{BL\_CGI-FU\_CGI}}{\text{BL\_CGI}}$, where BL_CGI denotes baseline CGI score and FU_CGI denotes the follow-up CGI score. After that, we used 40% as a threshold for categorization.

2. Follow-up CGI scores: We used FU_CGI = 3 value as a threshold value for sub-grouping the patients. If patients have FU_CGI scores less than 3, then they will be in "class 0", otherwise in "class 1".

SAPS/SANS Dataset

For 53 patients, we have the baseline and the follow-up SAPS/SANS scores. Total scores for the SANS range from 0 to 125 and 0 to 170 for the SAPS.

Here, we followed the below procedure for categorizing the least, versus the most, improved patients.

1. Percentage of improvement: We computed the percentage of improvement by using PI = $\frac{\text{BL-FU}}{\text{BL}}$, where BL denotes baseline SAPS/SANS score and FU denotes follow-up SAPS/SANS score. After that, we used 40% and 90% as thresholds for SAPS, SANS, respectively for categorization.

2. Follow-up SAPS/SANS scores:

   (a) Based on SAPS Scores: We used SAPS follow-up (FU) scores 0 as a threshold for sub-grouping the patients. If patients have FU scores equal to 0 (full improvement), then they will be in "class 0", otherwise in "class 1".

   (b) Based on SANS Scores: We used SANS follow-up (FU) scores 18 – which is a median SANS follow-up score– as a threshold. If patients have FU scores less than this threshold, then they will be in "class 0", otherwise in "class 1".

Hallucination Dataset

This dataset contains baseline and follow-up hallucination scores for 47 patients. We computed the percentage of improvement using these scores. We considered 100% improvement – which means the percentage of improvement (PI) score is equal to 1 – as "class 1", and other ones are "class 0".

5.3.2 Data Preprocessing and Feature Extraction

Data Preprocessing

Clinical Features We have 495 clinical features, some of which were missing. To handle missing values for learning algorithms like logistic regression, which cannot handle missing variables inherently, we used Multivariate Imputation by Chained Equation (MICE) that assumes data are missing at random [19] to impute the missing values.

fMRI Data We preprocesssed around 40% of the fMRI data using DPARSFA (Data Processing Assistant for Resting-State fMRI — Advanced Version) [159], which is an SPM-based Matlab toolbox. We collected the rest of the preprocessed fMRI data from the National Institute of Mental Health Neurosciences (NIMHANS), India. Initially, we discarded the first ten volumes of each functional time-series before reaching steady magnetization and to allow the participants to adapt to the scanning noise. Finally, we followed the fMRI preprocessing pipeline, which is described in Section 2.2.1.

Feature Extraction

Regional based fMRI features We followed the fMRI preprocessing pipeline, which is described in Section 2.2.1, for extracting the ALFF, fALFF, Reho features. Here, each type of feature is a 3-dimensional matrix with a size of $61 \times 73 \times 61$.

Seed-based Functional Connectivity features We used the seed-based feature extraction procedure (described in Section 5.2.2) for computing the functional connectivity matrix from the 4D fMRI data. Here, we also used LTPJ and RDLPFC as two seed points. From the feature extraction process, we got a 3-dimensional matrix with a size of $61 \times 73 \times 61$ for each seed-point.

5.3.3 Methods

Without Transfer learning

We have five extracted different feature types (ALFF, fALFF, ReHo, and LTPJ and RDLPFC seed based). As each of the extracted features are 3-dimensional, we chose to use the similar 3D convolutional neural network (CNN) – that is used for tDCS SCZ prognosis prediction (shown in Figure 5.1 (Panel A)) – to predict the least, versus the most improvement patients for each feature type and each of the treatment response condition, described in Section 5.3.1.

a **Seed based model:** We considered the two seed points (LTPJ and RDLPFC) based feature matrices as two input channels of the CNN

model (shown in Figure 5.1 - Panel A). So, the model takes a fixed input size of $2 \times 61 \times 73 \times 61$, here the first number denotes the number of input channels (= 2 channels), and other ones denote the size of the 3-dimensional feature matrix (= $61 \times 73 \times 61$).

b **Other Feature types:** For each of the other features (ALFF, fALFF, and ReHo), we built different CNN models separately (shown in Figure 5.1 - Panel A). Each model takes a fixed input size of $1 \times 61 \times 73 \times 61$, here the number of channels = 1, and the size of feature matrix = $61 \times 73 \times 61$.

With Transfer Learning

For building a prognosis model for antipsychotic SCZ patients, we also used the similar transfer learning technique that is similar to the one described in Section 5.2.3. In addition to our current experimental datasets (described in Section 5.3.1), we have additional 473 fMRI samples that include healthy controls, first degree relative of Schizophrenia, OCD, tDCS SCZ (these SCZ subjects received transcranial direct current stimulation (tDCS)) and tDCS OCD (these OCD subjects received tDCS) patients with baseline CGI scores. Subject with CGI score = 0 were assigned the label of 0, whereas subject with CGI score > 3 were assigned the label of 1. The sample distributions of these analyses are provided in Table 5.4. We did the oversampling manually on 'Label 1' based on sex, age to balance the classes, and finally, this process gave 496 samples (in which the number of distinct instances are 473) for our pretraining purposes.

In the first step, we trained a CNN architecture (shown in Figure 5.1 Panel A) on 496 samples, and we used this for transfer learning. Then, we froze the first five layers of the feature extraction part, removed the decision layers, and added new decision layers. Finally, the weights of the unfrozen layers are fine-tuned/trained with each of our experimental datasets (described in Section 5.3.1).

Pretraining (N)	Label 0: CGI = 0	Label 1: CGI >3
Healthy Controls (186)	186	0
FDR - Schizophrenia (62)	62	0
OCD (149)	0	149
tDCS- SCZ (33)	0	33
tDCS - OCD (20)	0	20
Total	248	202

Table 5.4: Sample distribution in transfer learning model for predicting the antipsychotic treatment response.

Stacking CNN Models

We used a stacking CNNs based model that includes a separate CNN architecture for each feature type to do the feature extraction. The feature extraction used in this model is similar to the feature extractor of Figure 5.1. The extracted features from each CNN model are concatenated, and then passed to hidden layers. For training the model, we used Relu as an activation function for each layer, cross-entropy as a loss function, and 50% dropout in the hidden layers. The general architecture of the stacking CNN model is depicted in Figure 5.3.

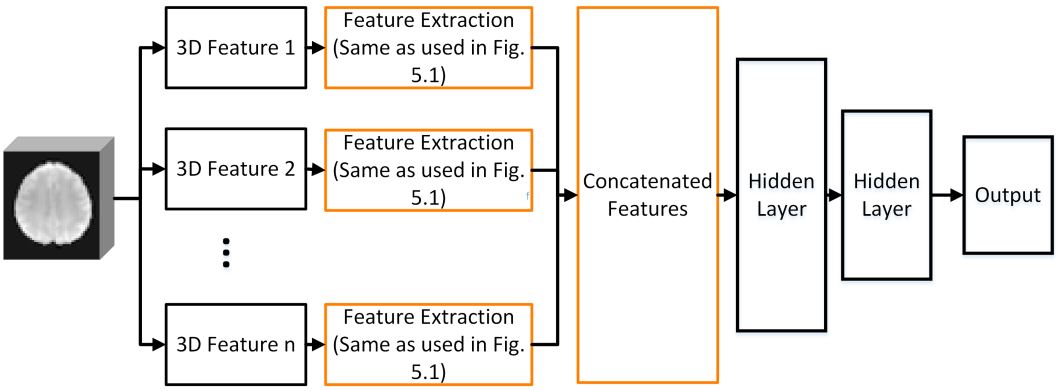


Figure 5.3: Stacked neural network models for prognosis prediction.

Densely Connected Convolutional Networks (Densenet)

We used one of the most popular deep learning methods, Densenet, which was originally proposed for 2D data [55]. For our experiment, we used 3D

convolutions in place of 2D convolutions in Densenet to adapt the model for 3D fMRI data. This model takes an input with a fixed size of $2 \times 61 \times 73 \times 61$, where first channel corresponds to 3D LTPJ features, and the second channel corresponds to 3D RDLPFC features.

Combining Clinical Features with fMRI Data

We combined clinical features with the fMRI data to get a better and stable neural network model. To do this, we considered two different ways to combine the data modalities.

1. Option 1: We concatenated clinical features with extracted features of CNN in the linear layer (shown in Figure 5.1 Panel B (ii)). Here, we used two sets of clinical features for building two different models. The first set of features are 'Accumulative_drug_record', 'Binned_time_leave_medication', 'Final_days_record', and 'Time_leave_medication', which may be highly predictive factors for classification according to our prior knowledge. The second set of features includes all other 491 features.

2. Option 2: We used a separate fully connected neural network (FNN) (shown in Figure 5.1 Panel B (iii)) that takes all 495 clinical features as input. Then, we combined both linear layers of CNN model and FNN model, and finally, we passed concatenated features into another fully connected neural network.

5.3.4 Results

We used the classification frameworks, which are described in Section 5.3.3, to classify patients into low improvement versus high improvement based on different clinical baseline and follow-up scores. The 5×10-fold cross-validation classification results for SANS dataset are shown in Tables 5.5, which shows that the performance of all of the methods, except the transfer learning, was around the baseline for the SANS scores based treatment response prediction.

Model Name	Class Labels based on	
	Percentage of Improvement (Baseline: 50.94%)	Follow-up SANs Score (Baseline: 50.94%)
3D CNN Model using LTPJ and RDLPFC	53.96 ± 1.37	54.33 ± 2.58
Densenet Model using LTPJ and RDLPFC	58.49 ± 1.41	56.60 ± 1.41
3D CNN Model using ALFF	59.25 ± 1.47	59.24 ± 1.14
3D CNN Model using fALFF	52.84 ± 1.84	51.32 ± 2.09
3D CNN Model using ReHo	52.46 ± 1.45	50.94 ± 0.92
Stacked 3D CNN Model Using LTPJ, RDLPFC and ReHo	56.60 ± 1.68	56.98 ± 1.64
Transfer Learning using LTPJ and RDLPFC	$\mathbf{63.77 \pm 1.45}$	$\mathbf{59.57 \pm 0.98}$

Table 5.5: Treatment response prediction results based on baseline and follow-up SANS scores: average (standard errors)- 5×10-fold CV.

Here, transfer learning method gave 63.77% and 59.57% accuracies for the percentage of improvement based model and follow-up SANs score based model, respectively.

However, for the other two clinical scores (SAPS and CGI Severity), none of the methods were able to provide significant performance above from the baseline. Results for these two datasets are provided in Appendix B (Tables B.4, and B.5). We tried all of these methods on Hallucination dataset, whose baseline is 63.82%. None of the methods were able to reach significantly better performance.

With the hope of building a more accurate model, we added the clinical features to each model. However, these additional features did not improve model performance.

5.3.5 Discussion

In this study, we tried to build a classification framework that can accurately distinguish patients with high treatment response from low treatment response. Here, we built different CNN models using different combinations of the different feature types (LTPJ, RDLPFC, ALFF, fALFF, and ReHo). The transfer learning method helps to improve performance on SANS dataset, which justifies future research into the transfer learning method, where we have only access to the limited amount of target dataset. For SAPS and CGIs datasets, the transfer learning models were unable to provide significant improvement compared to other methods. Maybe the reasons behind these unsuccessful experiments are the smaller number of subjects compared to the high-dimensional fMRI features, and the lack of information of the patients' medication history – Did the patients actually follow the clinicians' prescribed medication or not– which might affect their improvement process.

We did not try to use standard classification models (e.g., logistic regression, linear SVM, xgboost), because these types of models were already explored in Ghoreishiamiri's thesis [43]. Unfortunately, Ghoreishiamiri did not able to get any significant performance from this dataset.

Chapter 6

Conclusion

Machine learning applications using neuroimaging are becoming increasingly common in psychiatric research, as it provides a multi-dimensional, data-driven approach that captures the level of complexity necessary for aiding diagnosis, and prognosis in an objective manner. But, as many psychiatric disorders usually manifest with a myriad of overlapping symptoms, reliable clinical diagnosis/ prognosis is a challenging task. The challenges are also amplified by the smaller number of neuroimaging samples with high dimensional features. So, we should be careful in terms of model architecture selection to avoid overfitting. Briefly, in this dissertation, we empirically show that (a) incorporating prior knowledge in the training process produces models that help to improve the performance in OCD diagnosis and SCZ prognosis prediction; (b) transfer learning across psychiatric diagnoses (SCZ to OCD) improved the OCD diagnosis prediction with fewer sets; (c) a cross-application of a machine-learned SCZ diagnostic model could help to identify FDR subjects with high levels of schizotypy.

In the first study, our objective was to build a classifier that could distinguish healthy control versus OCD patients. Here, we used both models that incorporate prior neurobiological knowledge (parcellations) and knowledge-agnostic (neural networks) models. The "prior neurobiological knowledge based" model gave 80.3% accuracy, whereas the best accuracy obtained from neural network based models is 76.7%. In this experiment, it shows that a simple linear model with neurobiology knowledge based features outperforms

complex neural network models, even though neural network models can do the feature extraction automatically. We also showed that it is possible to transfer learning across psychiatric diagnoses (SCZ to OCD). Despite these successful stories, we were not successful in identifying subtypes based on symptom dimensions or severity. From the experimental results, we anticipate that fMRI data might not have enough information to perform more complicated psychiatric decisions, at least within the scope of our training size and algorithms.

In our second study, we attempted to build a cross-application of a machine learned schizophrenia diagnostic model in identifying FDR subjects with high schizotypy scores. Results show that FDR subjects who were classified as schizophrenia patients by our model learned to distinguish schizophrenia from healthy control using schizophrenia dataset had significantly higher 'schizotypal personality scores' than those who were not classified. However, whether similar prediction performance holds for a larger population without familial association remains to be explored.

In the last study of this dissertation, we aimed to build two prognosis models for two types of treatment (antipsychotic medication, and tDCS treatment) of SCZ patients. For antipsychotic SCZ patients, we found mild success (63.77% accuracy) from the deep transfer learning model to predict the treatment response. However, we achieved 77.5% accuracy using the prior knowledge based model for tDCS SCZ patients. One of the interesting findings from the tDCS prognosis experiments is that: at least with this training set and learning algorithms, raw ROI features gave comparable performance compared to average ROI features, when we incorporated more specific prior neurological knowledge (regions – that are implicated during the disorder and it's treatment) in the learning models.

From our diagnosis and prognosis works, we empirically showed that any type of prior knowledge incorporation in the model helps to improve the model performance. Briefly, in the neural network, we have added prior knowledge through transfer learning ideas, whereas in other methods, we used the predefined brain atlases, and additionally, we handpicked few regions, which have

been identified to be relevant to the particular disorder, from the literature. We believe that our research works could be greatly beneficial, as the clinicians can use our diagnosis model to help doctors make important clinical decisions – such as 'what treatment to prescribe'. Also, this drives a practice of medicine that is truly 'evidence based', which is particularly crucial for psychiatry.

References

[1] A. Aboraya, E. Rankin, C. France, A. El-Missiry, and C. John, "The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis.," *Psychiatry (Edgmont (Pa. : Township))*, 2006, ISSN: 1550-5952. 47

[2] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in Neuroinformatics*, 2014, ISSN: 16625196. DOI: 10.3389/fninf.2014.00014. arXiv: 1412.3919. 28

[3] S. M. Agarwal, V. Shivakumar, A. Bose, A. Subramaniam, H. Nawani, H. Chhabra, S. V. Kalmady, J. C. Narayanaswamy, and G. Venkatasubramanian, *Transcranial direct current stimulation in schizophrenia*, 2013. DOI: 10.9758/cpn.2013.11.3.118. 55

[4] A. J. Allen, M. E. Griss, B. S. Folley, K. A. Hawkins, and G. D. Pearlson, *Endophenotypes in schizophrenia: A selective review*, 2009. DOI: 10.1016/j.schres.2009.01.016. 50

[5] American Psychiatric Association, *American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.)* 2013, ISBN: 9780890425541. DOI: 10.1176/appi.books.9780890425596.744053. 23

[6] A. Amerio, A. Odone, C. C. Liapis, and S. N. Ghaemi, *Diagnostic validity of comorbid bipolar disorder and obsessive-compulsive disorder: A systematic review*, 2014. DOI: 10.1111/acps.12250. 21

[7] *Amplitude of Low Frequency Fluctuations (ALFF) and fractional ALFF (f/ALFF) — C-PAC 1.6.2 Beta documentation*. [Online]. Available: https://fcp-indi.github.io/docs/user/alff.html (visited on 05/17/2020). 12

[8] X. M. Androulakis, K. Krebs, B. L. Peterlin, T. Zhang, N. Maleki, S. Sen, C. Rorden, and P. Herath, "Modulation of intrinsic resting-state fMRI networks in women with chronic migraine," *Neurology*, 2017, ISSN: 1526632X. DOI: 10.1212/WNL.0000000000004089. 59

[9] *Atlases - FslWiki*. [Online]. Available: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases (visited on 09/21/2019). 96

[10] N. Barrantes-Vidal, P. Grant, and T. R. Kwapil, "The role of schizotypy in the study of the etiology of schizophrenia spectrum disorders," in *Schizophrenia Bulletin*, 2015. DOI: 10.1093/schbul/sbu191. 2, 54

[11] P. Bellec, P. Rosa-Neto, O. C. Lyttelton, H. Benali, and A. C. Evans, "Multi-level bootstrap analysis of stable clusters in resting-state fMRI," *NeuroImage*, 2010, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.02.082. 28, 112

[12] M. Bodnar, P. O. Harvey, A. K. Malla, R. Joober, and M. Lepage, "The parahippocampal gyrus as a neural marker of early remission in first-episode psychosis: A voxel-based morphometry study," *Clinical Schizophrenia and Related Psychoses*, 2011, ISSN: 19351232. DOI: 10.3371/CSRP.4.4.2. 56

[13] M. Bodnar, A. K. Malla, R. Joober, C. Lord, E. Smith, J. Pruessner, and M. Lepage, "Neural markers of early remission in first-episode schizophrenia: A volumetric neuroimaging study of the parahippocampus," *Psychiatry Research - Neuroimaging*, 2012, ISSN: 09254927. DOI: 10.1016/j.pscychresns.2011.07.012. 56

[14] A. Bose, V. Shivakumar, S. M. Agarwal, S. V. Kalmady, S. Shenoy, V. S. Sreeraj, J. C. Narayanaswamy, and G. Venkatasubramanian, "Efficacy of fronto-temporal transcranial direct current stimulation for refractory auditory verbal hallucinations in schizophrenia: A randomized, double-blind, sham-controlled study," *Schizophrenia Research*, 2018, ISSN: 15732509. DOI: 10.1016/j.schres.2017.08.047. 55

[15] A. Bose, V. Shivakumar, J. C. Narayanaswamy, H. Nawani, A. Subramaniam, S. M. Agarwal, H. Chhabra, S. V. Kalmady, and G. Venkatasubramanian, "Insight facilitation with add-on tDCS in schizophrenia," *Schizophrenia Research*, 2014, ISSN: 15732509. DOI: 10.1016/j.schres.2014.03.029. 55

[16] L. Breiman, "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)," *Statistical Science*, 2001, ISSN: 0883-4237. DOI: 10.1214/ss/1009213726. 2

[17] X. Bu, X. Hu, L. Zhang, B. Li, M. Zhou, L. Lu, X. Hu, H. Li, Y. Yang, W. Tang, Q. Gong, and X. Huang, "Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder," *Translational Psychiatry*, 2019, ISSN: 21583188. DOI: 10.1038/s41398-018-0362-9. 2, 21, 22, 48

[18] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. Thomas Yeo, "The organization of the human cerebellum estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, 2011, ISSN: 00223077. DOI: 10.1152/jn.00339.2011. 27

[19] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of Statistical Software*, 2011, ISSN: 15487660. DOI: 10.18637/jss.v045.i03. 59, 67

[20] B. Cao, R. Y. Cho, D. Chen, M. Xiu, L. Wang, J. C. Soares, and X. Y. Zhang, "Treatment response prediction and individualized identification of first-episode drug-naïve schizophrenia using brain functional connectivity," *Molecular Psychiatry*, 2018, ISSN: 14765578. DOI: 10.1038/s41380-018-0106-5. 57

[21] R. Caruana, "Multitask Learning," in *Learning to Learn*, Springer US, 1998, pp. 95–133. DOI: 10.1007/978-1-4615-5529-2_5. 23

[22] X. Chang, Y. B. Xi, L. B. Cui, H. N. Wang, J. B. Sun, Y. Q. Zhu, P. Huang, G. Collin, K. Liu, M. Xi, S. Qi, Q. R. Tan, D. M. Miao, and H. Yin, "Distinct inter-hemispheric dysconnectivity in schizophrenia patients with and without auditory verbal hallucinations," *Scientific Reports*, 2015, ISSN: 20452322. DOI: 10.1038/srep11218. 27

[23] D. Chen, S. Liu, P. Kingsbury, S. Sohn, C. B. Storlie, E. B. Habermann, J. M. Naessens, D. W. Larson, and H. Liu, "Deep learning and alternative learning strategies for retrospective real-world clinical data," *npj Digital Medicine*, 2019, ISSN: 2398-6352. DOI: 10.1038/s41746-019-0122-0. 22

[24] B. N. Cuthbert and T. R. Insel, "Toward the future of psychiatric diagnosis: The seven pillars of RDoC," *BMC Medicine*, 2013, ISSN: 17417015. DOI: 10.1186/1741-7015-11-126. 23, 48

[25] R. De Filippis, E. A. Carbone, R. Gaetano, A. Bruni, V. Pugliese, C. Segura-Garcia, and P. De Fazio, *Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: A systematic review*, 2019. DOI: 10.2147/NDT.S202418. 2, 21

[26] S. J. De Wit, P. Alonso, L. Schweren, D. Mataix-Cols, C. Lochner, J. M. Menchón, D. J. Stein, J. P. Fouche, C. Soriano-Mas, J. R. Sato, M. Q. Hoexter, D. Denys, T. Nakamae, S. Nishida, J. S. Kwon, J. H. Jang, G. F. Busatto, N. Cardoner, D. C. Cath, K. Fukui, W. H. Jung, S. N. Kim, E. C. Miguel, J. Narumoto, M. L. Phillips, J. Pujol, P. L. Remijnse, Y. Sakai, N. Y. Shin, K. Yamada, D. J. Veltman, and O. A. Van Den Heuvel, "Multicenter voxel-based morphometry mega-analysis of structural brain scans in obsessive-compulsive disorder," *American Journal of Psychiatry*, 2014, ISSN: 15357228. DOI: 10.1176/appi.ajp.2013.13040574. 42, 49

[27] M. Debbané and N. Barrantes-Vidal, "Schizotypy from a developmental perspective," in *Schizophrenia Bulletin*, 2015. DOI: 10.1093/schbul/sbu175. 51

[28] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "A sulcal depth-based anatomical parcellation of the cerebral cortex.," *NeuroImage*, 2009, ISSN: 10538119. DOI: 10.1016/s1053-8119(09)71561-7. 27, 28, 112

[29] *Diffusion Tensor Imaging 101*. [Online]. Available: http://www.diffusion-imaging.com/2009/05/diffusion-tensor-imaging-101.html (visited on 08/21/2019). 8

[30] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. Pruett, D. M. Barch, S. E. Petersen, and B. L. Schlaggar, "Prediction of individual brain maturity using fMRI," *Science*, 2010, ISSN: 00368075. DOI: 10.1126/science.1194144. 28, 112

[31] *ECR 2018 / C-2545 / Simultaneous Multi-Slice DW Imaging: Speed and Quality in different body regions - EPOS^{TM}*. [Online]. Available: https://epos.myesr.org/esr/viewing/index.php?module=viewing%7B%5C_%7Dposter%7B%5C&%7Dtask=viewsection%7B%5C&%7Dpi=141702%7B%5C&%7Dti=495443%7B%5C&%7Dsi=1718 (visited on 05/23/2020). 17

[32] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates," *Proceedings of the National Academy of Sciences of the United States of America*, 2016, ISSN: 10916490. DOI: 10.1073/pnas.1602413113. 2

[33] *ENIGMA-DTI diffusivity protocol « ENIGMA*. [Online]. Available: http://enigma.ini.usc.edu/protocols/dti-protocols/enigma-dti-diffusivity-protocol/ (visited on 03/04/2019). 14

[34] *ENIGMA-DTI diffusivity protocol « ENIGMA*. [Online]. Available: http://enigma.ini.usc.edu/protocols/dti-protocols/enigma-dti-diffusivity-protocol/ (visited on 03/04/2019). 16

[35] F. Esposito, S. Ferilli, T. M. A. Basile, and N. Di Mauro, *Machine Learning in Document Analysis and Recognition*, S. Marinai and H. Fujisawa, Eds., ser. Studies in Computational Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 90, pp. 105–138, ISBN: 978-3-540-76279-9. DOI: 10.1007/978-3-540-76280-5. [Online]. Available: http://dblp.uni-trier.de/db/series/sci/sci90.html%7B%5C#%7DEspositoFBM08. 49

[36] C. Fahim, E. Stip, A. Mancini-Marïe, A. Gendron, B. Mensour, and M. Beauregard, "Differential hemodynamic brain activity in schizophrenia patients with blunted affect during quetiapine treatment," *Journal of Clinical Psychopharmacology*, 2005, ISSN: 02710749. DOI: 10.1097/01.jcp.0000168880.10793.ed. 57

[37] K. J. Friston, A. P. Holmes, K. J. Worsley, J. .-. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, 1994, ISSN: 10970193. DOI: 10.1002/hbm.460020402. 11

[38] P. Fusar-Poli, R. Smieskova, M. J. Kempton, B. C. Ho, N. C. Andreasen, and S. Borgwardt, *Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies*, 2013. DOI: 10.1016/j.neubiorev.2013.06.001. 55

[39] P. Fusar-Poli, J. Perez, M. Broome, S. Borgwardt, A. Placentino, E. Caverzasi, M. Cortesi, P. Veggiotti, P. Politi, F. Barale, and P. McGuire, *Neurofunctional correlates of vulnerability to psychosis: A systematic review and meta-analysis*, 2007. DOI: 10.1016/j.neubiorev.2006.11.006. 1

[40] S. Gao, V. D. Calhoun, and J. Sui, *Machine learning in major depression: From classification to treatment outcome prediction*, 2018. DOI: 10.1111/cns.13048. 2, 21

[41] Y. Gao, Y. Zhang, H. Wang, X. Guo, and J. Zhang, "Decoding Behavior Tasks from Brain Activity Using Deep Transfer Learning," *IEEE Access*, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2907040. 30

[42] D. L. Garver, J. A. Holcomb, and J. D. Christensen, "Heterogeneity of response to antipsychotics from multiple disorders in the schizophrenia spectrum," *Journal of Clinical Psychiatry*, 2000, ISSN: 01606689. DOI: 10.4088/JCP.v61n1213. 55

[43] S. Ghoreishiamiri, "Addressing the Challenges of Applying Machine Learning for Predicting Mental Disorders and Their Prognosis Using Two Case Studies," PhD thesis, 2018. 16, 73

[44] M. Goodkind, S. B. Eickhoff, D. J. Oathes, Y. Jiang, A. Chang, L. B. Jones-Hagata, B. N. Ortega, Y. V. Zaiko, E. L. Roach, M. S. Korgaonkar, S. M. Grieve, I. Galatzer-Levy, P. T. Fox, and A. Etkin, "Identification of a common neurobiological substrate for mental Illness," *JAMA Psychiatry*, 2015, ISSN: 2168622X. DOI: 10.1001/jamapsychiatry.2014.2206. 23, 49

[45] W. K. Goodman, L. H. Price, S. A. Rasmussen, C. Mazure, R. L. Fleischmann, C. L. Hill, G. R. Heninger, and D. S. Charney, "The Yale-Brown Obsessive Compulsive Scale: I. Development, Use, and Reliability," *Archives of General Psychiatry*, 1989, ISSN: 15383636. DOI: 10.1001/archpsyc.1989.01810110048007. 24

[46] I. I. Gottesman and T. D. Gould, *The endophenotype concept in psychiatry: Etymology and strategic intentions*, 2003. DOI: 10.1176/appi.ajp.160.4.636. 3, 50

[47] T. A. Greenwood, A. Shutes-David, and D. W. Tsuang, "Endophenotypes in Schizophrenia: Digging Deeper to Identify Genetic Mechanisms," *Journal of Psychiatry and Brain Science*, vol. 4, no. 2, 2019. DOI: 10.20900/jpbs.20190005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/31245629%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6594566. 50

[48] P. Gruner, A. Vo, M. Argyelan, T. Ikuta, A. J. Degnan, M. John, B. D. Peters, A. K. Malhotra, A. M. Uluğ, and P. R. Szeszko, "Independent component analysis of resting state activity in pediatric obsessive-compulsive disorder," *Human Brain Mapping*, 2014, ISSN: 10970193. DOI: 10.1002/hbm.22551. 21, 22, 48

[49] G. Gundersen, B. Dumitrascu, J. T. Ash, and B. E. Engelhardt, "End-to-end training of deep probabilistic CCA on paired biomedical observations," in *35th Conference on Uncertainty in Artificial Intelligence, UAI 2019*, 2019. 101

[50] G. Haddock, J. McCarron, N. Tarrier, and E. B. Faragher, "Scales to measure dimensions of hallucinations and delusions: The psychotic symptom rating scales (PSYRATS)," *Psychological Medicine*, 1999, ISSN: 00332917. DOI: 10.1017/S0033291799008661. 20

[51] B. C. Ho, N. C. Andreasen, P. Nopoulos, S. Arndt, V. Magnotta, and M. Flaum, "Progressive structural brain abnormalities and their relationship to clinical outcome: A longitudinal magnetic resonance imaging study early in schizophrenia," *Archives of General Psychiatry*, 2003, ISSN: 0003990X. DOI: 10.1001/archpsyc.60.6.585. 56

[52] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, *Causability and explainability of artificial intelligence in medicine*, 2019. DOI: 10.1002/widm.1312. 23

[53] D. Hoyer, E. P. Zorrilla, P. Cottone, S. Parylak, M. Morelli, N. Simola, N. Simola, M. Morelli, J. D. Lane, M. M. Morgan, M. J. Christie, C. J. Hillard, A. J. Budney, C. J. Hillard, A. J. Budney, R. G. Vandrey, T. Robbins, P. Nencini, M. S. Milella, N. (Mirza, D. Hoyer, V. L. Harvey, T. Dickenson, S. Miyamoto, O. Almeida, N. Sousa, S. Yu, A. Absalom, D. Menon, and A. e. a. Absalom, "Clinical Global Impression Scales," in *Encyclopedia of Psychopharmacology*, 2010. DOI: 10.1007/978-3-540-68706-1_1422. 19

[54] X. Hu, L. Zhang, X. Bu, H. Li, B. Li, W. Tang, L. Lu, X. Hu, S. Tang, Y. Gao, Y. Yang, N. Roberts, Q. Gong, and X. Huang, "Localized connectivity in obsessive-compulsive disorder: An investigation combining univariate and multivariate pattern analyses," *Frontiers in Behavioral Neuroscience*, 2019, ISSN: 16625153. DOI: 10.3389/fnbeh.2019.00122. 2, 21, 22, 48

[55] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, ISBN: 9781538604571. DOI: 10.1109/CVPR.2017.243. arXiv: 1608.06993. 70

[56] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, S. Zhao, L. Guo, and T. Liu, "Modeling Task fMRI Data Via Deep Convolutional Autoencoder," *IEEE Transactions on Medical Imaging*, 2018, ISSN: 1558254X. DOI: 10.1109/TMI.2017.2715285. 30

[57] a. S. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging, Second Edition*, 2009. 8

[58] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, *Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review*, 2019. DOI: 10.1007/s40489-019-00158-x. 2, 21

[59] S. E. Hyman, *Can neuroscience be integrated into the DSM-V?* 2007. DOI: 10.1038/nrn2218. 23, 48

[60] E. Jääskeläinen, P. Juola, J. Kurtti, M. Haapea, M. Kyllönen, J. Miettunen, P. Tanskanen, G. K. Murray, S. Huhtaniska, A. Barnes, J. Veijola, and M. Isohanni, "Associations between brain morphology and outcome in schizophrenia in a general population sample," *European Psychiatry*, 2014, ISSN: 17783585. DOI: 10.1016/j.eurpsy.2013.10.006. 56

[61] Y. C. Janardhan Reddy, A. Sundar, J. Narayanaswamy, and S. Math, *Clinical practice guidelines for Obsessive-Compulsive Disorder*, 2017. DOI: 10.4103/0019-5545.196976. 21

[62] R. Jardri, A. Pouchet, D. Pins, and P. Thomas, "Cortical activations during auditory verbal hallucinations in schizophrenia: A coordinate-based meta-analysis," *American Journal of Psychiatry*, 2011, ISSN: 0002953X. DOI: 10.1176/appi.ajp.2010.09101522. 60, 114, 116

[63] S. V. Kalmady, R. Greiner, R. Agrawal, V. Shivakumar, J. C. Narayanaswamy, M. R. Brown, A. J. Greenshaw, S. M. Dursun, and G. Venkatasubramanian, "Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning," *npj Schizophrenia*, 2019, ISSN: 2334265X. DOI: 10.1038/s41537-018-0070-8. 22, 24, 26, 28, 29, 33, 43

[64] M. Kekic, E. Boysen, I. C. Campbell, and U. Schmidt, *A systematic review of the clinical efficacy of transcranial direct current stimulation (tDCS) in psychiatric disorders*, 2016. DOI: 10.1016/j.jpsychires.2015.12.018. 55

[65] M. G. Kendall, "Rank Correlation Methods.," *Biometrika*, 1957, ISSN: 00063444. DOI: 10.2307/2333282. 12

[66] N. V. Kraguljac, D. M. White, N. Hadley, J. A. Hadley, L. V. Hoef, E. Davis, and A. C. Lahti, "Aberrant hippocampal connectivity in unmedicated patients with schizophrenia and effects of antipsychotic medication: A longitudinal resting state functional mri study," *Schizophrenia Bulletin*, 2016, ISSN: 17451701. DOI: 10.1093/schbul/sbv228. 57

[67] S. Kühn and J. Gallinat, "Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia," *Schizophrenia Bulletin*, 2012, ISSN: 05867614. DOI: 10.1093/schbul/sbq152. 60, 114, 116

[68] S. Kumari, M. MPH, M. Malik, M. C. Florival, M. P. Manalai, MD, and S. S. MD, "An Assessment of Five (PANSS, SAPS, SANS, NSA-16, CGI-SCH) commonly used Symptoms Rating Scales in Schizophrenia and Comparison to Newer Scales (CAINS, BNSS)," *Journal of Addiction Research & Therapy*, 2017. DOI: 10.4172/2155-6105.1000324. 19

[69] V. Kumari, E. R. Peters, D. Fannon, E. Antonova, P. Premkumar, A. P. Anilkumar, S. C. Williams, and E. Kuipers, "Dorsolateral Prefrontal Cortex Activity Predicts Responsiveness to Cognitive-Behavioral Therapy in Schizophrenia," *Biological Psychiatry*, 2009, ISSN: 00063223. DOI: 10.1016/j.biopsych.2009.04.036. 57

[70] M. F. Kuo, P. S. Chen, and M. A. Nitsche, *The application of tDCS for the treatment of psychiatric diseases*, 2017. DOI: 10.1080/09540261.2017.1286299. 55

[71] X. Li, J. Hect, M. Thomason, and D. Zhu, "Interpreting Age Effects of Human Fetal Brain from Spontaneous fMRI using Deep 3D Convolutional Neural Networks," Jun. 2019. arXiv: 1906.03691. [Online]. Available: http://arxiv.org/abs/1906.03691. 30

[72] J. Lieberman, M. Chakos, H. Wu, J. Alvir, E. Hoffman, D. Robinson, and R. Bilder, "Longitudinal study of brain morphology in first episode schizophrenia," *Biological Psychiatry*, 2001, ISSN: 00063223. DOI: 10.1016/S0006-3223(01)01067-8. 56

[73] D. Luck, L. Buchy, Y. Czechowska, M. Bodnar, G. B. Pike, J. S. Campbell, A. Achim, A. Malla, R. Joober, and M. Lepage, "Fronto-temporal disconnectivity and clinical short-term outcome in first episode psychosis: A DTI-tractography study," *Journal of Psychiatric Research*, 2011, ISSN: 00223956. DOI: 10.1016/j.jpsychires.2010.07.007. 57

[74] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017. arXiv: 1705.07874. 63

[75] K. Macy, W. Staal, C. Kraper, A. Steiner, T. D. Spencer, L. Kruse, M. Azimova, M. J. Weiss, T. Zane, M. J. Weiss, F. Orlich, C. Ray-Subramanian, M. D. Powers, S. Butler, C. Lord, J. Rohrer, S. Butler, C. Lord, S. Butler, C. Lord, R. Munday, B. Rogé, M. D. Powers, M. D. Powers, S. Bendiske, M. H. Charlop, C. A. Miltenberger, A. Snow, P. Cavanagh, S. Egan, M. Azimova, J. Molteni, S. A. Mason, M. South, J. Palilla, F. R. Volkmar, J. W. Thomas, F. R. Volkmar, M. D. Kaiser, J. R. Gaag, T. R. Welch, T. Soto, C. Kraper, K. A. Pelphrey, M. J. Weiss, G. Vivanti, J. McCullagh, K. O'Hearn, K. O'Hearn, J. Parr, A. S. Le Couteur, E. R. Eernisse, M. South, J. Palilla, A. Stevens, R. Bernier, R. Gilroy, Y.-S. Kim, S. Hwang, B. Leventhal, L. D. Scahill, and L. D. Scahill, "Blood-Oxygen-Level-Dependent (BOLD) Signal," in *Encyclopedia of Autism Spectrum Disorders*, Springer New York, 2013, pp. 465–466. DOI: 10.1007/978-1-4419-1698-3_550. 7

[76] K. H. Madsen, L. G. Krohne, X. L. Cai, Y. Wang, and R. C. Chan, *Perspectives on Machine Learning for Classification of Schizotypy Using fMRI Data*, 2018. DOI: 10.1093/schbul/sby026. 51

[77] D. Maust, M. Cristancho, L. Gray, S. Rushing, C. Tjoa, and M. E. Thase, "Psychiatric rating scales," in *Handbook of Clinical Neurology*, 2012. DOI: 10.1016/B978-0-444-52002-9.00013-9. 17

[78] J. M. McClellan, E. Susser, and M. C. King, *Schizophrenia: A common disease caused by multiple rare alleles*, 2007. DOI: 10.1192/bjp.bp.106.025585. 2, 50

[79] T. H. Mcglashan, B. C. Walsh, D. Ph, S. W. Woods, J. Addington, K. Cadenhead, T. Cannon, B. Cornblatt, L. Davidson, R. Heinssen, R. Hoffman, T. K. Larsen, J. Rosen, M. Tsuang, and E. Walker, "STRUCTURED INTERVIEW FOR PSYCHOSIS-RISK SYNDROMES," Tech. Rep., 2014. 51

[80] M. McGue, I. I. Gottesman, and D. C. Rao, "The transmission of schizophrenia under a multifactorial threshold model," *American Journal of Human Genetics*, 1983, ISSN: 00029297. 50

[81] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, 1998, ISSN: 10659471. DOI: 10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBM5>3.0.CO;2-1. 28

[82] J. E. Mervis, R. J. Capizzi, E. Boroda, and A. W. Macdonald, *Transcranial direct current stimulation over the dorsolateral prefrontal cortex in schizophrenia: A quantitative review of cognitive outcomes*, 2017. DOI: 10.3389/fnhum.2017.00044. 55

[83] R. J. Meszlényi, K. Buza, and Z. Vidnyánszky, "Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture," *Frontiers in Neuroinformatics*, 2017, ISSN: 16625196. DOI: 10.3389/fninf.2017.00061. arXiv: 1707.06682. 30

[84] M. R. Milad and S. L. Rauch, *Obsessive-compulsive disorder: Beyond segregated cortico-striatal pathways*, 2012. DOI: 10.1016/j.tics.2011.11.003. 42, 49

[85] *MNI Space*. [Online]. Available: http://brainmap.org/training/BrettTransform.html (visited on 04/30/2020). 11

[86] *Modulation at Structural Brain Mapping Group*. [Online]. Available: http://www.neuro.uni-jena.de/vbm/segmentation/modulation/ (visited on 03/10/2019). 13

[87] J. Moncrieff, "Antipsychotic Maintenance Treatment: Time to Rethink?" *PLoS Medicine*, 2015, ISSN: 15491676. DOI: 10.1371/journal.pmed.1001861. 56

[88] M. Mondino, F. Haesebaert, E. Poulet, M. F. Suaud-Chagny, and J. Brunelin, *Fronto-temporal transcranial Direct Current Stimulation (tDCS) reduces source-monitoring deficits and auditory hallucinations in patients with schizophrenia*, 2015. DOI: 10.1016/j.schres.2014.10.054. 55

[89] S. Mori and J. D. Tournier, *Introduction to diffusion tensor imaging: And higher order models: Second edition*. Elsevier Inc., 2013, pp. 1–126, ISBN: 9780123983985. DOI: 10.1016/C2011-0-07607-X. 9

[90] G. S. Mudholkar, "Fisher's ¡i¿Z¡/i¿ -Transformation," in *Encyclopedia of Statistical Sciences*, Hoboken, NJ, USA: John Wiley & Sons, Inc., Aug. 2006. DOI: 10.1002/0471667196.ess0796.pub2. [Online]. Available: http://doi.wiley.com/10.1002/0471667196.ess0796.pub2. 59

[91] B. Nelson, P. D. McGorry, M. Wichers, J. T. Wigman, and J. A. Hartmann, *Moving from static to dynamic models of the onset of mental disorder a review*, 2017. DOI: 10.1001/jamapsychiatry.2017.0001. 23

[92] *Neuroimaging in Python - Pipelines and Interfaces — nipy pipeline and interfaces package*. [Online]. Available: https://nipype.readthedocs.io/en/latest/interfaces/generated/interfaces.fsl/preprocess.html (visited on 09/14/2019). 14

[93] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, ISBN: 9783319467221. DOI: 10.1007/978-3-319-46723-8_25. 30

[94] T. M. Nienow, A. W. MacDonald, and K. O. Lim, *TDCS produces incremental gain when combined with working memory training in patients with schizophrenia: A proof of concept pilot study*, 2016. DOI: 10.1016/j.schres.2016.01.053. 55

[95] *Nipype Beginner's Guide — All you need to know to become an expert in Nipype*. [Online]. Available: https://miykael.github.io/nipype-beginner-s-guide/neuroimaging.html (visited on 03/23/2020). 6

[96] L. J. O'Donnell and C. F. Westin, *An introduction to diffusion tensor image analysis*, 2011. DOI: 10.1016/j.nec.2010.12.004. 9

[97] N. D. Orlov, D. K. Tracy, D. Joyce, S. Patel, J. Rodzinka-Pasko, H. Dolan, J. Hodsoll, T. Collier, J. Rothwell, and S. S. Shergill, "Stimulating cognition in schizophrenia: A controlled pilot study of the effects of prefrontal transcranial direct current stimulation upon memory and learning," *Brain Stimulation*, 2017, ISSN: 18764754. DOI: 10.1016/j.brs.2016.12.013. 55

[98] E. Parrado-Hernández, V. Gómez-Verdejo, M. Martínez-Ramón, J. Shawe-Taylor, P. Alonso, J. Pujol, J. M. Menchón, N. Cardoner, and C. Soriano-Mas, "Discovering brain regions relevant to obsessive-compulsive disorder identification through bagging and transduction," *Medical Image Analysis*, 2014, ISSN: 13618415. DOI: 10.1016/j.media.2014.01.006. 48

[99] A. Paszke, G. Chanan, Z. Lin, S. Gross, E. Yang, L. Antiga, and Z. Devito, "Automatic differentiation in PyTorch," in *31st Conference on Neural Information Processing Systems*, 2017, ISBN: 9788578110796. DOI: 10.1017/CBO9781107707221.009. arXiv: arXiv:1011.1669v3. 33, 62

[100] K. R. Patel, J. Cherian, K. Gohil, and D. Atkinson, *Schizophrenia: Overview and treatment options*, Sep. 2014. 55

[101] W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. 2007, ISBN: 9780123725608. DOI: 10.1016/B978-0-12-372560-8.X5000-1. 13

[102] J. B. Poline and M. Brett, *The general linear model and fMRI: Does love last forever?* 2012. DOI: 10.1016/j.neuroimage.2012.01.133. 1

[103] P. H. Pondé, E. P. De Sena, J. A. Camprodon, A. N. De Araújo, M. F. Neto, M. DiBiasi, A. F. Baptista, L. M. Moura, and C. Cosmo, *Use of transcranial direct current stimulation for the treatment of auditory hallucinations of schizophrenia - a systematic review*, 2017. DOI: 10.2147/NDT.S122016. 55

[104] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen, "Functional Network Organization of the Human Brain," *Neuron*, 2011, ISSN: 08966273. DOI: 10.1016/j.neuron.2011.09.006. 27, 28, 112

[105] M. Quarantelli, O. Palladino, A. Prinster, V. Schiavone, B. Carotenuto, A. Brunetti, A. Marsili, M. Casiello, G. Muscettola, M. Salvatore, and A. De Bartolomeis, "Patients with poor response to antipsychotics have a more severe pattern of frontal atrophy: A voxel-based morphometry study of treatment resistance in Schizophrenia," *BioMed Research International*, 2014, ISSN: 23146141. DOI: 10.1155/2014/325052. 56

[106] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016. arXiv: 1511.06434. 101

[107] J. Radua, O. A. Van Den Heuvel, S. Surguladze, and D. Mataix-Cols, "Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs other anxiety disorders," *Archives of General Psychiatry*, 2010, ISSN: 0003990X. DOI: 10.1001/archgenpsychiatry.2010.70. 42, 49

[108] A. Raine and D. Benishay, "The SPQ-B: A brief screening instrument for schizotypal personality disorder," *Journal of Personality Disorders*, 1995, ISSN: 0885579X. DOI: 10.1521/pedi.1995.9.4.346. 52, 53

[109] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral mr images," *IEEE Transactions on Medical Imaging*, 1997, ISSN: 02780062. DOI: 10.1109/42.563663. 13

[110] G. M. Reed, P. Sharan, T. J. Rebello, J. W. Keeley, M. Elena Medina-Mora, O. Gureje, J. Luis Ayuso-Mateos, S. Kanba, B. Khoury, C. S. Kogan, V. N. Krasnov, M. Maj, J. de Jesus Mari, D. J. Stein, M. Zhao, T. Akiyama, H. F. Andrews, E. Asevedo, M. Cheour, T. Domínguez-Martínez, J. El-Khoury, A. Fiorillo, J. Grenier, N. Gupta, L. Kola, M. Kulygina, I. Leal-Leturia, M. Luciano, B. Lusu, J. Nicolas, I. Martínez-López, C. Matsumoto, L. Umukoro Onofa, S. Paterniti, S. Purnima, R. Robles, M. K. Sahu, G. Sibeko, N. Zhong, M. B. First, W. Gaebel, A. M. Lovell, T. Maruta, M. C. Roberts, and K. M. Pike, "The ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders: results among adult patients in mental health settings of 13 countries," *World Psychiatry*, 2018, ISSN: 20515545. DOI: 10.1002/wps.20524. 48

[111] D. A. Regier, W. E. Narrow, D. E. Clarke, H. C. Kraemer, S. J. Kuramoto, E. A. Kuhl, and D. J. Kupfer, "DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses," *American Journal of Psychiatry*, 2013, ISSN: 15357228. DOI: 10.1176/appi.ajp.2012.12070999. 48

[112] *Regional Homogeneity — C-PAC 1.6.2 Beta documentation*. [Online]. Available: https://fcp-indi.github.io/docs/user/reho (visited on 04/30/2020). 12

[113] C. P. Rollins, J. R. Garrison, J. S. Simons, J. B. Rowe, C. O'Callaghan, G. K. Murray, and J. Suckling, "Meta-analytic Evidence for the Plurality of Mechanisms in Transdiagnostic Structural MRI Studies of Hallucination Status," *EClinicalMedicine*, 2019, ISSN: 25895370. DOI: 10.1016/j.eclinm.2019.01.012. 60, 114, 116

[114] M. C. do Rosário, M. Batistutto, and Y. Ferrao, *Symptom Heterogeneity in OCD*, C. Pittenger, Ed. Oxford University Press, Oct. 2017, vol. 1. DOI: 10.1093/med/9780190228163.003.0008. [Online]. Available: http://www.oxfordmedicine.com/view/10.1093/med/9780190228163.001.0001/med-9780190228163-chapter-8. 1, 21

[115] A. M. Ruscio, D. J. Stein, W. T. Chiu, and R. C. Kessler, "The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication," *Molecular Psychiatry*, 2010, ISSN: 13594184. DOI: 10.1038/mp.2008.94. 21

[116] R. B. Rutledge, A. M. Chekroud, and Q. J. Huys, *Machine learning and big data in psychiatry: toward clinical applications*, 2019. DOI: 10.1016/j.conb.2019.02.006. 2, 21

[117] F. Sambataro, G. Blasi, L. Fazio, G. Caforio, P. Taurisano, R. Romano, A. Di Giorgio, B. Gelao, L. Lo Bianco, A. Papazacharias, T. Popolizio, M. Nardini, and A. Bertolino, "Treatment with olanzapine is associated with modulation of the default mode network in patients with schizophrenia," *Neuropsychopharmacology*, 2010, ISSN: 0893133X. DOI: 10.1038/npp.2009.192. 57

[118] D. K. Sarpal, D. G. Robinson, T. Lencz, M. Argyelan, T. Ikuta, K. Karlsgodt, J. A. Gallego, J. M. Kane, P. R. Szeszko, and A. K. Malhotra, "Antipsychotic treatment and functional connectivity of the striatum in first-episode schizophrenia," *JAMA Psychiatry*, 2015, ISSN: 2168622X. DOI: 10.1001/jamapsychiatry.2014.1734. 57

[119] S. Sarraf, D. D. DeSouza, J. Anderson, G. Tofighi, and f. t. A. D. N. Initiativ, "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI," *bioRxiv*, 2017. DOI: 10.1101/070441. 30

[120] F. Schirmbeck and M. Zink, *Comorbid obsessive-compulsive symptoms in schizophrenia: Contributions of pharmacological and genetic factors*, 2013. DOI: 10.3389/fphar.2013.00099. 21

[121] H. G. Schnack and R. S. Kahn, "Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters," *Frontiers in Psychiatry*, 2016, ISSN: 16640640. DOI: 10.3389/fpsyt.2016.00050. 22

[122] B. Sen, G. A. Bernstein, T. Xu, B. A. Mueller, M. W. Schreiner, K. R. Cullen, and K. K. Parhi, "Classification of obsessive-compulsive disorder from resting-state fMRI," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016, ISBN: 9781457702204. DOI: 10.1109/EMBC.2016.7591508. 21, 22, 48

[123] N. Shahid, T. Rappon, and W. Berta, *Applications of artificial neural networks in health care organizational decision-making: A scoping review*, 2019. DOI: 10.1371/journal.pone.0212356. 22

[124] D. V. Sheehan, Y. Lecrubier, K. H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, and G. C. Dunbar, "The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10," in *Journal of Clinical Psychiatry*, vol. 59, 1998, pp. 22–33. 24, 51

[125] S. K. Shenas, U. Halici, and M. Cicek, "Detection of obsessive compulsive disorder using resting-state functional connectivity data," in *Proceedings of the 2013 6th International Conference on Biomedical Engineering and Informatics, BMEI 2013*, IEEE Computer Society, 2013, pp. 132–136. DOI: 10.1109/BMEI.2013.6746921. 21, 22, 48

[126] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015. arXiv: 1409.1556. 33, 60

[127] K. Siu, D. M. Stuart, M. Mahmoud, and A. Moshovos, "Memory Requirements for Convolutional Neural Network Hardware Accelerators," in *2018 IEEE International Symposium on Workload Characterization, IISWC 2018*, 2018, ISBN: 9781538667804. DOI: 10.1109/IISWC.2018.8573527. 33

[128] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences of the United States of America*, 2009, ISSN: 00278424. DOI: 10.1073/pnas.0905267106. 28

[129] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. Behrens, "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data," *NeuroImage*, 2006, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2006.02.024. 16

[130] J. M. Soares, P. Marques, V. Alves, and N. Sousa, "A hitchhiker's guide to diffusion tensor imaging," *Frontiers in Neuroscience*, 2013, ISSN: 16624548. DOI: 10.3389/fnins.2013.00031. 8

[131] J. Soler, P. Ferentinos, C. Prats, S. Miret, M. Giralt, V. Peralta, L. Fañanás, and M. Fatjó-Vilas, "Familial aggregation of schizotypy in schizophrenia-spectrum disorders and its relation to clinical and neurodevelopmental characteristics," *Journal of Psychiatric Research*, 2017, ISSN: 18791379. DOI: 10.1016/j.jpsychires.2016.09.026. 51, 54

[132] C. Soriano-Mas, J. Pujol, P. Alonso, N. Cardoner, J. M. Menchón, B. J. Harrison, J. Deus, J. Vallejo, and C. Gaser, "Identifying patients with obsessive-compulsive disorder using whole-brain anatomy," *NeuroImage*, 2007, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2007.01.011. 2, 48, 49

[133] G. S. Steketee, J. B. Grayson, and E. B. Foa, "Obsessive-compulsive disorder: Differences between washers and checkers," *Behaviour Research and Therapy*, 1985, ISSN: 00057967. DOI: 10.1016/0005-7967(85)90028-2. 99

[134] P. F. Sullivan, K. S. Kendler, and M. C. Neale, "Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies," *Archives of General Psychiatry*, 2003, ISSN: 0003990X. DOI: 10.1001/archpsyc.60.12.1187. 2, 50

[135] P. R. Szeszko, K. L. Narr, O. R. Phillips, J. McCormack, S. Sevy, H. Gunduz-Bruce, J. M. Kane, R. M. Bilder, and D. G. Robinson, "Magnetic resonance imaging predictors of treatment response in first-episode schizophrenia," *Schizophrenia Bulletin*, 2012, ISSN: 05867614. DOI: 10.1093/schbul/sbq126. 56

[136] A. Tahmassebi, M. H. Schulte, A. H. Gandomi, A. E. Goudriaan, I. McCann, and A. Meyer-Baese, "Deep learning in medical imaging: FMRI big data analysis via convolutional neural networks," in *ACM International Conference Proceeding Series*, 2018, ISBN: 9781450364461. DOI: 10.1145/3219104.3229250. 30

[137] Y. Takagi, Y. Sakai, G. Lisi, N. Yahata, Y. Abe, S. Nishida, T. Nakamae, J. Morimoto, M. Kawato, J. Narumoto, and S. C. Tanaka, "A neural marker of obsessive-compulsive disorder from whole-brain functional connectivity," *Scientific Reports*, 2017, ISSN: 20452322. DOI: 10.1038/s41598-017-07792-7. 2, 21, 22, 48

[138] J. Talairach and P. Tournoux, *Co-Planar Stereotaxis Atlas of the Human Brain: : an approach to cerebral imaging.* 1988, ISBN: 0865772932. 27

[139] X. Tan, "Artifacts Caused by Eddy Current in Diffusion MRI," Tech. Rep., 2010. 15

[140] *TBSS/UserGuide - FslWiki*. [Online]. Available: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide%7B%5C#%7DRunning%7B%5C_%7DTBSS (visited on 03/04/2019). 17

[141] *The diffusion tensor, and its relation to FA, MD, AD and RD*, 2015. [Online]. Available: http://www.diffusion-imaging.com/2015/10/what-is-diffusion-tensor.html. 16

[142] B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fisch, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, 2011, ISSN: 00223077. DOI: 10.1152/jn.00338.2011. 28, 112

[143] A. W. Thomas, K. R. Müller, and W. Samek, "Deep Transfer Learning for Whole-Brain FMRI Analyses," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, ISBN: 9783030326944. DOI: 10.1007/978-3-030-32695-1_7. arXiv: 1907.01953. 30

[144] F. Thomas, N. Bouaziz, C. Gallea, P. Schenin-King Andrianisaina, F. Durand, O. Bolloré, R. Benadhira, C. Isaac, S. Braha-Zeitoun, V. Moulier, A. Valero-Cabré, and D. Januel, "Structural and functional brain biomarkers of clinical response to rTMS of medication-resistant auditory hallucinations in schizophrenia patients: Study protocol for a randomized sham-controlled double-blind clinical trial," *Trials*, 2019, ISSN: 17456215. DOI: 10.1186/s13063-019-3311-x. 30, 59

[145] L. R. Trambaiolli, C. E. Biazoli, J. B. Balardin, M. Q. Hoexter, and J. R. Sato, "The relevance of feature selection methods to the classification of obsessive-compulsive disorder based on volumetric measures," *Journal of Affective Disorders*, 2017, ISSN: 15732517. DOI: 10.1016/j.jad.2017.06.061. 2

[146] D. Tromp, *DTI Processing - Distortion Correction*, 2012. [Online]. Available: http://www.diffusion-imaging.com/2012/03/dti-preprocessing-distortion-correction.html. 15

[147] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, 2002, ISSN: 10538119. DOI: 10.1006/nimg.2001.0978. 27, 28, 112

[148] P. Vakli, R. J. Deák-Meszlényi, P. Hermann, and Z. Vidnyánszky, "Transfer learning improves resting-state functional connectivity pattern analysis using convolutional neural networks," *GigaScience*, 2018, ISSN: 2047217X. DOI: 10.1093/gigascience/giy130. 30

[149] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, ISBN: 9783642220913. DOI: 10.1007/978-3-642-22092-0_46. 27, 28

[150] B. Viswanath, J. C. Narayanaswamy, R. P. Rajkumar, A. V. Cherian, T. Kandavel, S. B. Math, and Y. C. Reddy, "Impact of depressive and anxiety disorder comorbidity on the clinical expression of obsessive-compulsive disorder," *Comprehensive Psychiatry*, 2012, ISSN: 0010440X. DOI: 10.1016/j.comppsych.2011.10.008. 1, 21

[151] A. Vita, L. De Peri, G. Deste, and E. Sacchetti, *Progressive loss of cortical gray matter in schizophrenia: A meta-analysis and meta-regression of longitudinal MRI studies*, 2012. DOI: 10.1038/tp.2012.116. 55

[152] N. D. Volkow, G. J. Wang, J. S. Fowler, D. Tomasi, and R. Baler, "Neuroimaging of Addiction," in *Imaging of the Human Brain in Health and Disease*, Elsevier Inc., Jan. 2014, pp. 1–26, ISBN: 9780124186774. DOI: 10.1016/B978-0-12-418677-4.00001-4. 5

[153] *What Is FMRI? - Center for Functional MRI - UC San Diego*. [Online]. Available: http://fmri.ucsd.edu/Research/whatisfmri.html (visited on 05/01/2020). 7

[154] *White matter fibres, DTI scan - Stock Image - C042/6476 - Science Photo Library*. [Online]. Available: https://www.sciencephoto.com/media/986713/view/white-matter-fibres-dti-scan (visited on 04/29/2020). 8

[155] H. A. Whiteford, L. Degenhardt, J. Rehm, A. J. Baxter, A. J. Ferrari, H. E. Erskine, F. J. Charlson, R. E. Norman, A. D. Flaxman, N. Johns, R. Burstein, C. J. Murray, and T. Vos, "Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010," *The Lancet*, 2013, ISSN: 1474547X. DOI: 10.1016/S0140-6736(13)61611-6. 55

[156] C. W. Woo, A. Krishnan, and T. D. Wager, "Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations," *NeuroImage*, 2014, ISSN: 10959572. DOI: 10.1016/j.neuroimage.2013.12.058. 1

[157] T. S. Woodward, K. Jung, H. Hwang, J. Yin, L. Taylor, M. Menon, E. Peters, E. Kuipers, F. Waters, T. Lecomte, I. E. Sommer, K. Daalman, R. Van Lutterveld, D. Hubl, J. Kindler, P. Homan, J. C. Badcock, S. Chhabra, M. Cella, S. Keedy, P. Allen, A. Mechelli, A. Preti, S. Siddi, and D. Erickson, "Symptom dimensions of the psychotic symptom rating scales in psychosis: A multisite study," *Schizophrenia Bulletin*, 2014, ISSN: 17451701. DOI: 10.1093/schbul/sbu014. 19

[158] *YALE-BROWN OBSESSIVE COMPULSIVE SCALE (Y-BOCS)*. [Online]. Available: https://iocdf.org/wp-content/uploads/2016/04/04-Y-BOCS-w-Checklist.pdf (visited on 05/30/2020). 17

[159] Yan Chao-Gan and Zang Yu-Feng, "DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI," *Frontiers in System Neuroscience*, 2010, ISSN: 16625137. DOI: 10.3389/fnsys.2010.00013. 9, 11, 27, 68

[160] W. Yan, V. Calhoun, M. Song, Y. Cui, H. Yan, S. Liu, L. Fan, N. Zuo, Z. Yang, K. Xu, J. Yan, L. Lv, J. Chen, Y. Chen, H. Guo, P. Li, L. Lu, P. Wan, H. Wang, H. Wang, Y. Yang, H. Zhang, D. Zhang, T. Jiang, and J. Sui, "Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data," *EBioMedicine*, 2019, ISSN: 23523964. DOI: 10.1016/j.ebiom.2019.08.023. 30

[161] H. Yang, J. Zhang, Q. Liu, and Y. Wang, "Multimodal MRI-based classification of migraine: using deep learning convolutional neural network," *BioMedical Engineering OnLine*, 2018. DOI: 10.1186/s12938-018-0587-0. 30

[162] X. Yang, X. Hu, W. Tang, B. Li, Y. Yang, Q. Gong, and X. Huang, "Multivariate classification of drug-naive obsessive-compulsive disorder patients and healthy controls by applying an SVM to resting-state functional MRI data," *BMC Psychiatry*, 2019, ISSN: 1471244X. DOI: 10.1186/s12888-019-2184-6. 2, 21, 22, 48

[163] R. Zafar, A. S. Malik, A. N. Shuaibu, M. Javvad Ur Rehman, and S. C. Dass, "Classification of fMRI data using support vector machine and convolutional neural network," in *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017*, 2017, ISBN: 9781509055593. DOI: 10.1109/ICSIPA.2017.8120630. 30

[164] Y. F. Zang, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," *Brain and Development*, 2007, ISSN: 03877604. DOI: 10.1016/j.braindev.2006.07.002. 12

[165] Y. Zang, T. Jiang, Y. Lu, Y. He, and L. Tian, "Regional homogeneity approach to fMRI data analysis," *NeuroImage*, 2004, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2003.12.030. 12

[166] B. Zeng, B. A. Ardekani, Y. Tang, T. Zhang, S. Zhao, H. Cui, X. Fan, K. Zhuo, C. Li, Y. Xu, D. C. Goff, and J. Wang, "Abnormal white matter microstructure in drug-naive first episode schizophrenia patients before and after eight weeks of antipsychotic treatment," *Schizophrenia Research*, 2016, ISSN: 15732509. DOI: 10.1016/j.schres.2016.01.051. 57

[167] Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, L. Guo, X. Hu, J. Han, and T. Liu, "Automatic Recognition of fMRI-Derived Functional Networks Using 3-D Convolutional Neural Networks," *IEEE Transactions on Biomedical Engineering*, 2018, ISSN: 15582531. DOI: 10.1109/TBME.2017.2715281. 30

[168] K. Zilles and K. Amunts, *Receptor mapping: Architecture of the human cerebral cortex*, 2009. DOI: 10.1097/WCO.0b013e32832d95db. 27

[169] L. Zmigrod, J. R. Garrison, J. Carr, and J. S. Simons, *The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies*, 2016. DOI: 10.1016/j.neubiorev.2016.05.037. 60, 114, 116

[170] L. Zou, J. Zheng, C. Miao, M. J. McKeown, and Z. J. Wang, "3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI," *IEEE Access*, 2017, ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2762703. 30

[171] Q. H. Zou, C. Z. Zhu, Y. Yang, X. N. Zuo, X. Y. Long, Q. J. Cao, Y. F. Wang, and Y. F. Zang, "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF," *Journal of Neuroscience Methods*, 2008, ISSN: 01650270. DOI: 10.1016/j.jneumeth.2008.04.012. 12

Appendix A

Obsessive-Compulsive Disorder

A.1 OCD Diagnosis Prediction

A.1.1 Adding sMRI and DTI features in EMPaSchiz framework

Initially, we did sMRI and DTI preprocessing by using the preprocessing pipelines described in Sections 2.2.2 and 2.2.3. Then, we considered the above extracted sMRI and DTI data as additional features of the EMPaSchiz framework. sMRI features were parcellated using the same 14 parcellations, which were used in the EMPaSchiz framework [63]. For extracting DTI features, we took two procedures: (1) we used JHU atlas (ICBM-DTI-81 white-matter labels atlas) [9] for extracting the mean ROI values, according to Engima's protocol; (2) we used Neuromorphometics, LPBA40, and Hammers, cobra atlases with a computational anatomy toolbox, CAT12[1] for extracting the ROI (region-of-interest) based features.

Unfortunately, we could not get any significant improvement in predicting OCD versus HC after combining these two types of features to other 6 types of features (ALFF, fALFF, ReHo, FC-Precision, FC-correlation, and FC-partial correlation) in EMPaSchiz framework. We got 80.1% (respectively, 79.5% and 79.0%) accuracy for the combined features of fMRI and JHU-based DTI (respectively, the combined features of fMRI and CAT12-based DTI, and the combined features of sMRI and fMRI) features.

[1] `http://www.neuro.uni-jena.de/cat/`

A.1.2 Implement EMPaSchiz Framework using Neural Network Structure

EMPaSchiz uses two levels of learning to produce the final classification model. At the first label, it trains 84 different logistic regression classifiers with L2-regularization. In the second level, it trains a single logistic regression classifier with L2-regularization from the previously trained 84 classifiers to predict OCD versus HC. So, EMPaSchiz's 2nd level model is trained on the output of the first level 84 models, and it, unlike neural networks, does not able to back propagate the 2nd level model's error to the first level 84 models. That is why initial 84 models are not able to adjust themselves based on the second level outputs.

As a result, initially, we assumed that if we were able to learn the final model in a single level using neural network structure, we could be able to back propagate the loss/error to the initial input. So, we might be able to produce a better classification model. However, the accuracy of this model was 79.2%, whereas EMPaSchiz gave 80.3% accuracy in terms of distinguishing between OCD and HC. So, we were not able to get any advantage of using a single level learning method over the two levels learning structure. The architecture of this stacked neural network (= single level learning method) is shown in Figure A.1.

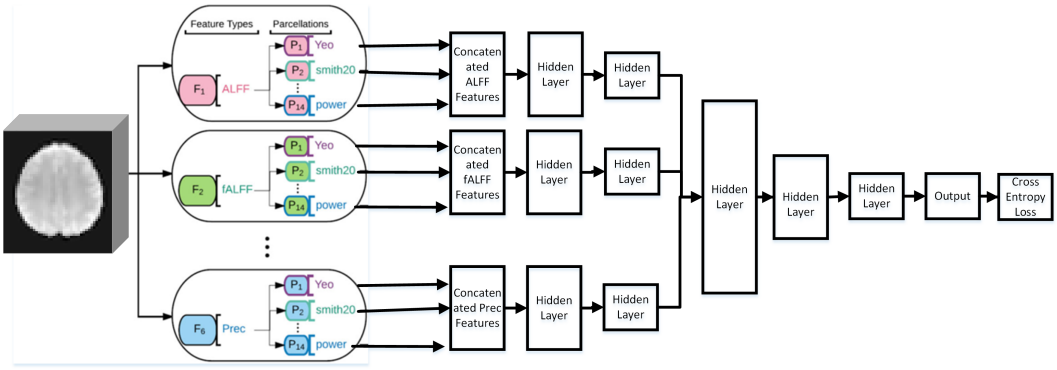


Figure A.1: EMPaSchiz framework implemented in a structure of neural network.

A.2 Multi-Class Classification Model

We have a dataset of first degree relatives (FDRs) of schizophrenia patients along with the schizophrenia (SCZ), and healthy control (HC) samples. The details of the FDR subjects are given in Section 4.2. We tried to build a 3-way classification among SCZ, FDR, and HC. As the number of subjects in HCs is more than double compared to SCZ and FDR subjects, we used the undersampling method inside the EMPaSchiz framework. A multinomial logistic regression classifier was used in EMPaSchiz framework to handle the multi-class classification problem. We used 5 shuffled iterations of 10-fold cross-validation (balanced based on class labels) for evaluating the model, and the model was evaluated in terms of the following metrics: Accuracy, Precision, and Sensitivity. For each metric, we have reported the mean and standard errors overall 50 train-test splits. The classification results are shown in Table A.1. Here, the overall model accuracy is around 69%, but as it is observable in the results, model performance on predicting FDR is not significant – precision and sensitivity for FDR prediction are less than 50%. Furthermore, under-sampling, over-sampling, and assigning class weights did not help to improve model performance.

	Prediction Class				
		HC	SCZ	FDR	Sensitivity
True Class	HC (191)	140.4 (0.82)	18.4 (0.82)	32.2 (0.52)	73.54 (0.44)
	SCZ (81)	6.6 (0.77)	64.0 (0.63)	10.4 (0.45)	78.86 (0.78)
	FDR (72)	21.4 (0.66)	16.6 (0.35)	34.0 (0.69)	47.04 (0.91)
	Precision	84.12 (0.41)	65.87 (0.15)	46.33 (1.12)	Overall Accuracy 69.28 (0.14)

Table A.1: EMPaSchiz performance on HC versus SCZ versus FDR.

Similarly, we attempted to learn another 3-way classification model for classifying among OCD, SCZ, and HC. The results are given in Table A.2. The overall model accuracy was around 68%, but the sensitivity score for SCZ class is 24.21% (too low). Here, most of the SCZ patients were classified as OCD patients.

	Prediction Class				
True Class		HC	OCD	SCZ	Sensitivity
	HC (191)	146.4 (0.92)	44.0 (0.63)	0.6 (0.35)	76.63 (1.37)
	OCD (184)	35.6 (0.66)	144.6 (0.82)	3.8 (0.52)	78.56 (1.51)
	SCZ (81)	11.8 (1.24)	49.6 (1.21)	19.6 (0.87)	24.21 (1.85)
	Precision	75.76 (0.76)	60.78 (0.71)	84.86 (3.21)	Overall Accuracy 68.11 (0.13)

Table A.2: EMPaSchiz performance on HC versus OCD versus SCZ.

A.3 Symptom Severity Prediction

A.3.1 DY-BOCS based Severity Prediction

The researcher used Dimensional Yale–Brown Obsessive–Compulsive Scale (DY-BOCS) for accessing the presence and severity of OCD. We have 35 DY-BOCS sub-items in our OCD sample. We then applied the PCA to get the 35 PCA components, and we used the first 5 PCA components to compute the class labels. We used the first and last quartile of each component to categorize the least versus the most severely, symptomatic patients. So, finally, we had five sets of class levels. For each set of class levels, we trained 84 single source models (SSMs) (described in Section 3.4.1) and EMPaSchiz model (ensemble of 84 single source models).

We found that for the 5-th PCA component based class levels, some of the single source models gave accuaries in the range 60% to 68%, but EMPaSchiz gave 55% accuracy where the baseline accuracy is 50%. For other PCA components based class levels, SSMs performance is around 60% accuracy, and we got around 55% accuracy from EMPaSchiz. We found that identifying severe cases was not an easy task – we got mild success here.

A.3.2 Distinguish Washers versus Checkers

The most common behaviours in OCD patients are washing and checking [133]. Sometimes one patient may have both of these. But, in most cases [133], one type of behaviour predominates, and it helps to identify that OCD subject as a washer or checker. That is why, we tried EMPaSchiz model to classify washer

versus checker. Unfortunately, the model was not able to perform significantly better than the baseline F1-score.

A.3.3 Other Severity Measurements based Symptoms' Severity Prediction

Along with the obsessions and compulsion severity measurements of the OCD discussed in Section 3.7.4, there are few other scales for measuring OCD severity. Here is the list of severity measurements which ones we analyzed: Age at onset; Ill duration; Global Assessment of Functioning; Clinical Global Impression - Severity (CGI_S); current and lifetime obsession on contamination, aggressive, sexual, blasphemous, hoarding, pathological doubt, symmetrical, and miscellaneous; current and lifetime compulsions on washing, checking, repeating, arranging, hoarding, cognitive, miscellaneous. For each, we used the first and last quartile of these scales to categorize the least versus the most severely, symptomatic patients. Using EMPaSchiz with leave-one-out cross-validation gave accuracies around the baseline–not significantly above the baseline– for each case.

We also did the sub-group symptoms severity analysis. In our OCD sample, there are three sub-groups: Drug Naive (never taken a particular drug before), Drug-Free (not taken particular drugs for a certain period), On Treatment (currently taking the treatment). We used the above-mentioned severity measurements for each sub-group to predict the severity, but we were not able to reach a significant performance in terms of accuracy.

A.3.4 Multi-task Symptom Severity Prediction

One of the clinically important questions is to predict the current severity scores of (1) 2 subcategories of YBOCS (obsession and compulsion scores), (2) 8 subcategories of DYBOCS (Total score Contamination, Total score Hoarding and Collecting, Total score Symmetry, Ordering, Counting, and Arranging, Total score Aggression and related compulsion, Total score Sexual and Religious, Total score Miscellaneous, Total Score Pathological Doubt, Total score All Obsessions and Compulsions) from the current fMRI data (before

any treatment). For this experiment, we used two multitask neural network models (shown in Figures A.2, A.3). Architecture 1 (shown in Figure A.2) has initial few layers for reducing the dimension of each feature type, whereas architecture 2 (shown in Figure A.3) combines all the feature types into a single concatenated features vector. After that, for each model, there is a hidden layer that is shared among all the tasks, which helps the training process by simultaneously optimizing each task. As all the tasks are related to each other, learned information from one task can improve the results of other tasks. In this experiment, we used a subset of the original OCD dataset for which we had the YBOCS sub-items' scores and the DYBOCS sub-items' scores (N=136). We tried a different set of parameters for these two architectures, but we failed to get significant performance for each task in terms of F1-scores and mean squared errors.

A.3.5 Symptom Severity Prediction using Deep Probabilistic Canonical Correlation Analysis

Gundersen et al. [49] proposed DPCCA (Deep Probabilistic Canonical Correlation Analysis) to learn the shared latent structure of joint medical image and the gene expression data, and they showed that this learned model helps to obtain image features which are the highly predictive signature of gene expression data. This inspired us to use DPCCA to learn joint fMRI data and clinical features. The detailed description of the DPCCA system is provided in the original paper [49]. Figure A.4 shows the general architecture of DPCCA.

In our case, DPCCA consists of PCCA (Probabilistic Canonical Correlation Analysis), fMRI encoder and decoder, and clinical data encoder and decoder. To learn the shared model, we used 3D convolution-based DCGAN [106] neural network for fMRI data and a fully connected neural network for clinical data in DPCCA framework. In the DPCCA framework, PCCA forces the learning system to learn a model that maximally explains variation in the clinical data. We have the following features for doing the experiments:

1. fMRI based features: We used 3 types of fMRI features (ALFF, fALFF,

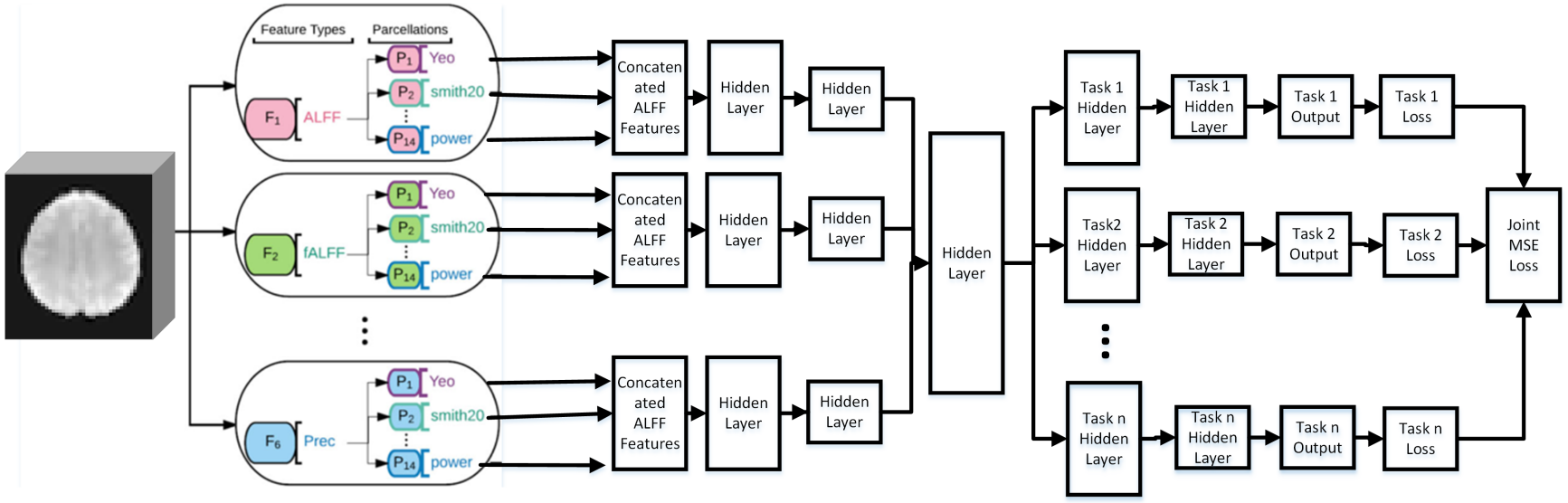


Figure A.2: Multitask neural network architecture 1.

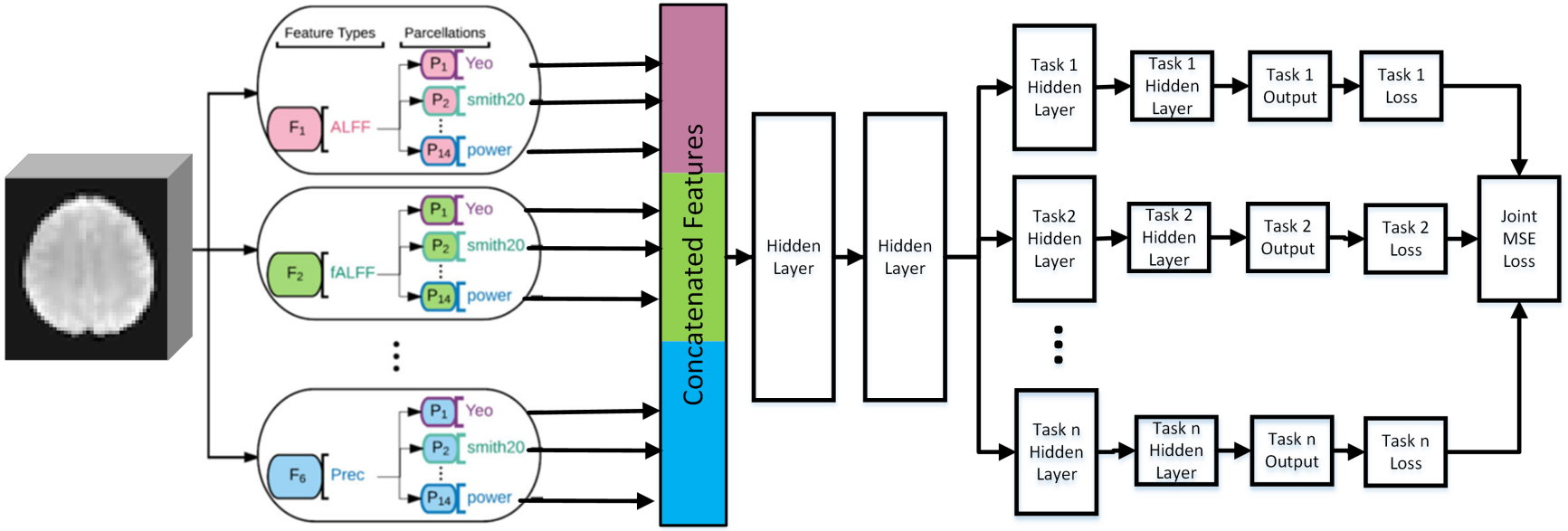


Figure A.3: Multitask neural network architecture 2.

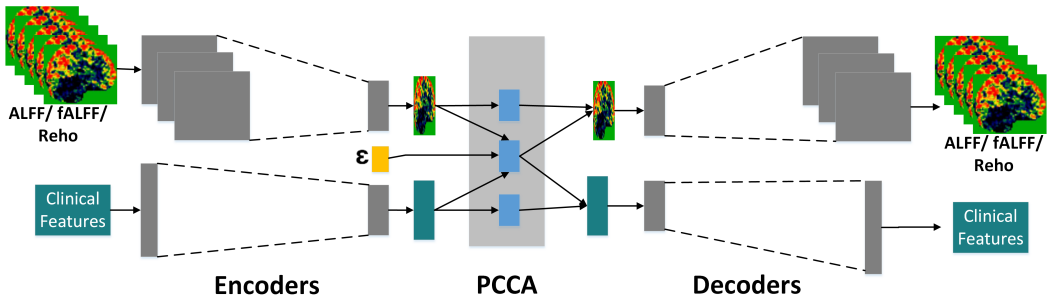


Figure A.4: General framework of DPCCA.

Reho), individually.

2. Clinical Features: All of the clinical features are on DYBOCS scale.

   (a) 35 sub-measurements from DYBOCS scale (provided in Appendix A, Table A.7)

   (b) Derived 35 PCA components from all the measures of DYBOCS scale (follows the same procedure which is used in Section A.3)

   (c) 7 sub-total measurements from DYBOCS scale: Those were Total score Contamination, Total score Hoarding and Collecting, Total score Symmetry, Ordering, Counting, and Arranging, Total score Aggression and related compulsion, Total score Sexual and Religious, Total score Miscellaneous, Total Score Pathological Doubt.

   (d) 6 Obsession and compulsion related sub-measurements: Those were Time score, Distress score, Interface score, Impairment score, Resistance score, Control score.

   We used these types of clinical features because those were related to our main performance task (output label).

We combined each type of fMRI feature with each type of clinical feature to classify the symptom severity (using two Y-BOCS measurements: obsessions or compulsions score). So, we trained 12 (3 types of fMRI features $\times$ 4 types of clinical features) different models for each measurement. The procedure of computing the class labels for these tasks is given in Section 3.7.4. After doing

all of these steps, we were not able to reach any significant results compared to the performance of EMPaSchiz (EMPaSchiz results on these tasks are given in Table 3.10).

A.4 Transfer learning using stochastic training process

For Schizophrenia dataset as inputs, EMPaSchiz's L1-regularization selected a different set of SSMs during 5 iterations. Figure A.5 (a) shows the number of times SSMs selected out of 5 times, for example, FC_prec with basc_multiscale_444 is selected for 4 times. Then, we built five different SCZ_to_OCD transfer models using a different set of SSMs. A different set of SSMs were obtained by considering how many times a particular SSM were selected during the 5 times runs. For example, FC_Corr with basc_multiscale_197, FC_prec with power, and FC_prec with dosenbach were selected more than 4 times. So, we used these three SSMs for building EMPaSchiz-3 model, where 3 denotes the number of SSMs.

For the first row in Table A.3, "Number of selected SSMs" column shows that 3 SSMs were selected more than 4 times (number of times selected is given in "Number of times SSM selected" column), and EMPaSchiz-t denotes EMPaSchiz framework used only t number of SSMs for building the final ensemble prediction model.

We also did the same things for SCZ_to_OCD _CommonHC, OCD_to_SCZ, and OCD_to_SCZ _CommonHC. The number of times SSMs selected for SCZ_to_OCD _CommonHC, OCD_to_SCZ, and OCD_to_SCZ _CommonHC are given in Figure A.5 (c, b, d), respectively. The results for these experiments are given in Tables A.4, A.5, A.6.

A.5 35 Measures under DYBOCS scale for OCD Subjects

The list of 35 measures under DYBOCS scale are given in Table A.7

Table A.3: EMPaSchiz results for OCD diagnosis prediction with feature selection based on Schizophrenia model (OCD (175), HC (88)).

SCZ_to_OCD models	NTSS	NSS	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
EMPaSchiz-3	>4	3	92.6 (0.8)	94.9 (0.8)	94.2 (0.7)	89.5 (1.7)	164.8 (1.0)	78.8 (0.8)	9.2 (0.8)	10.2 (1.0)
EMPaSchiz-5	>3	5	96.7 (0.4)	96.8 (0.5)	98.3 (0.4)	93.4 (1.1)	172.0 (0.6)	82.2 (0.3)	5.8 (0.3)	3.0 (0.6)
EMPaSchiz-13	>2	13	95.5 (0.4)	95.9 (0.5)	97.5 (0.4)	91.6 (1.1)	170.6 (0.8)	80.6 (0.6)	7.4 (0.6)	4.4 (0.8)
EMPaSchiz-15	>1	15	94.7 (0.5)	96.2 (0.5)	95.9 (0.7)	92.3 (1.1)	167.8 (0.9)	81.2 (0.5)	6.8 (0.5)	7.2 (0.9)
EMPaSchiz-18	>0	18	96.4 (0.4)	96.8 (0.4)	97.9 (0.3)	93.4 (1.0)	171.4 (0.4)	82.2 (0.6)	5.8 (0.6)	3.6 (0.4)
Original EMPaSchiz	All	84	93.6 (0.7)	94.4 (0.7)	96.3 (1.1)	88.1 (1.7)	168.6 (1.2)	77.6 (1.0)	10.4 (1.0)	6.4 (1.2)

N: OCD 175, HC 88; Baseline: 66.53
NTSS= Number of times SSM selected, NSS= Number of Selected SSMs

Table A.4: EMPaSchiz results for OCD diagnosis prediction with feature selection based on Schizophrenia model (OCD (175), HC (181)).

SCZ_to_OCD _CommonHC models	NTSS	NSS	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
EMPaSchiz-0	>4	0	-	-	-	-	-	-	-	-
EMPaSchiz-0	>3	0	-	-	-	-	-	-	-	-
EMPaSchiz-6	>2	6	73.9 (0.8)	72.4 (0.9)	76.5 (1.1)	71.5 (1.3)	133.8 (1.7)	129.4 (1.2)	51.6 (1.2)	41.2 (1.7)
EMPaSchiz-7	>1	7	73.5 (0.9)	71.5 (1.0)	77.3 (1.3)	69.8 (1.4)	135.2 (0.6)	126.4 (1.3)	54.6 (1.3)	39.8 (0.6)
EMPaSchiz-14	>0	14	75.2 (0.8)	73.2 (1.0)	79.0 (1.3)	71.5 (1.4)	138.2 (0.5)	129.4 (1.7)	51.6 (1.7)	36.8 (0.5)
Original EMPaSchiz	All	84	78.8 (1.0)	78.0 (1.3)	80.2 (1.3)	77.3 (1.8)	140.4 (1.8)	140.0 (1.5)	41.0 (1.5)	34.6 (1.8)

N: OCD 175, HC 181; Baseline: 0.5084
NTSS= Number of times SSM selected, NSS= Number of Selected SSMs

Table A.5: EMPaSchiz results for SCZ diagnosis prediction with feature selection based on OCD model (SCZ (81), HC (93)).

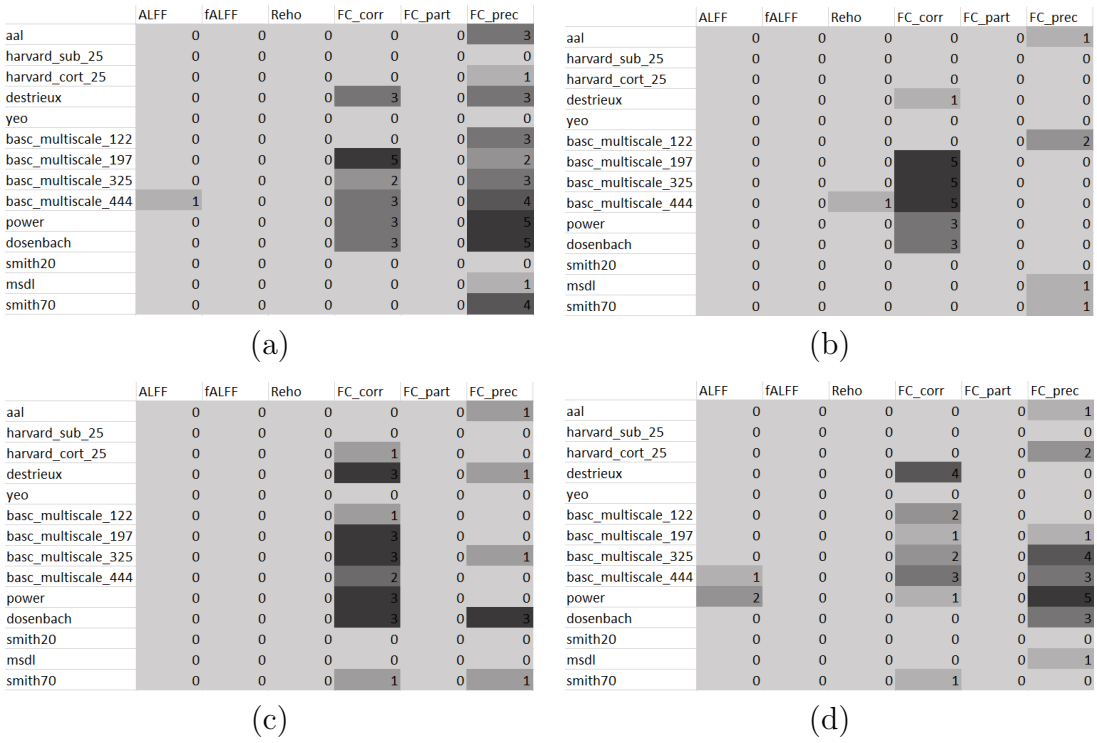
OCD_to_SCZ models	NTSS	NSS	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
EMPaSchiz-0	>4	0	-	-	-	-	-	-	-	-
EMPaSchiz-1	>3	1	74.5 (1.2)	72.7 (1.7)	74.8 (2.1)	74.3 (2.2)	60.6 (0.7)	69.0 (0.6)	24.0 (0.6)	20.4 (0.7)
EMPaSchiz-2	>2	2	75.5 (1.2)	73.3 (1.3)	75.1 (2.6)	75.9 (1.6)	60.8 (1.4)	70.6 (1.5)	22.4 (1.5)	20.2 (1.4)
EMPaSchiz-8	>1	8	81.7 (1.3)	84.9 (2.0)	74.6 (1.8)	87.8 (1.7)	60.4 (1.2)	81.6 (1.3)	11.4 (1.3)	20.6 (1.2)
EMPaSchiz-16	>0	16	82.8 (1.3)	84.3 (1.8)	78.5 (1.9)	86.5 (1.8)	63.6 (1.3)	80.4 (1.0)	12.6 (1.0)	17.4 (1.3)
Original EMPaSchiz	All	84	85.7 (1.0)	90.9 (1.4)	77.8 (2.0)	92.7 (1.2)	63.0 (0.4)	86.2 (0.5)	6.8 (0.5)	18.0 (0.4)

N: SCZ 81, HC 93; Baseline: 53.44
NTSS= Number of times SSM selected, NSS= Number of Selected SSMs

Table A.6: EMPaSchiz results for SCZ diagnosis prediction with feature selection based on OCD model (SCZ (81), HC (181)).

OCD_to_SCZ _CommonHC models	NTSS	NSS	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	True positive	True negative	False positive	False negative
EMPaSchiz-1	>4	1	83.6 (0.8)	83.9 (2.3)	59.8 (2.4)	94.2 (0.9)	48.4 (2.2)	170.6(1.0)	10.4 (1.0)	32.6 (2.2)
EMPaSchiz-3	>3	3	84.9 (1.0)	77.9 (1.9)	72.9 (2.1)	90.3 (1.1)	59.0 (0.8)	163.4 (1.3)	17.6 (1.3)	22.0 (0.8)
EMPaSchiz-6	>2	6	86.9 (1.0)	80.5 (2.0)	77.5 (2.4)	91.0 (1.1)	62.8 (0.5)	164.8 (1.9)	16.2 (1.9)	18.2 (0.5)
EMPaSchiz-10	>1	10	87.1 (0.8)	80.2 (1.5)	78.9 (2.2)	90.8 (1.0)	63.8 (0.7)	164.4 (0.8)	16.6 (0.8)	17.2 (0.7)
EMPaSchiz-17	>0	17	87.6 (0.9)	79.7 (1.6)	81.1 (1.9)	90.5 (0.9)	65.6 (0.5)	163.8 (0.5)	17.2 (0.5)	15.4 (0.5)
Original EMPaSchiz	All	84	83.2 (0.7)	84.8 (2.0)	56.8 (1.9)	95.0 (0.8)	46.0 (1.1)	172.0 (0.7)	9.0 (0.7)	35.0 (1.1)

N: SCZ 81, HC 181; Baseline: 69.08
NTSS= Number of times SSM selected, NSS= Number of Selected SSMs



	ALFF	fALFF	Reho	FC_corr	FC_part	FC_prec
aal	0	0	0	0	0	3
harvard_sub_25	0	0	0	0	0	0
harvard_cort_25	0	0	0	0	0	1
destrieux	0	0	0	3	0	3
yeo	0	0	0	0	0	0
basc_multiscale_122	0	0	0	0	0	3
basc_multiscale_197	0	0	0	5	0	2
basc_multiscale_325	0	0	0	2	0	3
basc_multiscale_444	1	0	0	3	0	4
power	0	0	0	3	0	5
dosenbach	0	0	0	3	0	5
smith20	0	0	0	0	0	0
msdl	0	0	0	0	0	1
smith70	0	0	0	0	0	4

(a)

	ALFF	fALFF	Reho	FC_corr	FC_part	FC_prec
aal	0	0	0	0	0	1
harvard_sub_25	0	0	0	0	0	0
harvard_cort_25	0	0	0	0	0	0
destrieux	0	0	0	1	0	0
yeo	0	0	0	0	0	0
basc_multiscale_122	0	0	0	0	0	2
basc_multiscale_197	0	0	0	5	0	0
basc_multiscale_325	0	0	0	5	0	0
basc_multiscale_444	0	0	1	5	0	0
power	0	0	0	3	0	0
dosenbach	0	0	0	3	0	0
smith20	0	0	0	0	0	0
msdl	0	0	0	0	0	1
smith70	0	0	0	0	0	1

(b)

	ALFF	fALFF	Reho	FC_corr	FC_part	FC_prec
aal	0	0	0	0	0	1
harvard_sub_25	0	0	0	0	0	0
harvard_cort_25	0	0	0	1	0	0
destrieux	0	0	0	3	0	1
yeo	0	0	0	0	0	0
basc_multiscale_122	0	0	0	1	0	0
basc_multiscale_197	0	0	0	3	0	0
basc_multiscale_325	0	0	0	3	0	1
basc_multiscale_444	0	0	0	2	0	0
power	0	0	0	3	0	0
dosenbach	0	0	0	3	0	3
smith20	0	0	0	0	0	0
msdl	0	0	0	0	0	0
smith70	0	0	0	1	0	1

(c)

	ALFF	fALFF	Reho	FC_corr	FC_part	FC_prec
aal	0	0	0	0	0	1
harvard_sub_25	0	0	0	0	0	0
harvard_cort_25	0	0	0	0	0	2
destrieux	0	0	0	4	0	0
yeo	0	0	0	0	0	0
basc_multiscale_122	0	0	0	2	0	0
basc_multiscale_197	0	0	0	1	0	1
basc_multiscale_325	0	0	0	2	0	4
basc_multiscale_444	1	0	0	3	0	3
power	2	0	0	1	0	5
dosenbach	0	0	0	0	0	3
smith20	0	0	0	0	0	0
msdl	0	0	0	0	0	1
smith70	0	0	0	1	0	0

(d)

Figure A.5: Number of times SSMs selected for a) SCZ_to_OCD, b) OCD_to_SCZ, c) SCZ_to_OCD_CommonHC, and d) OCD_to_SCZ_CommonHC transfer learning.

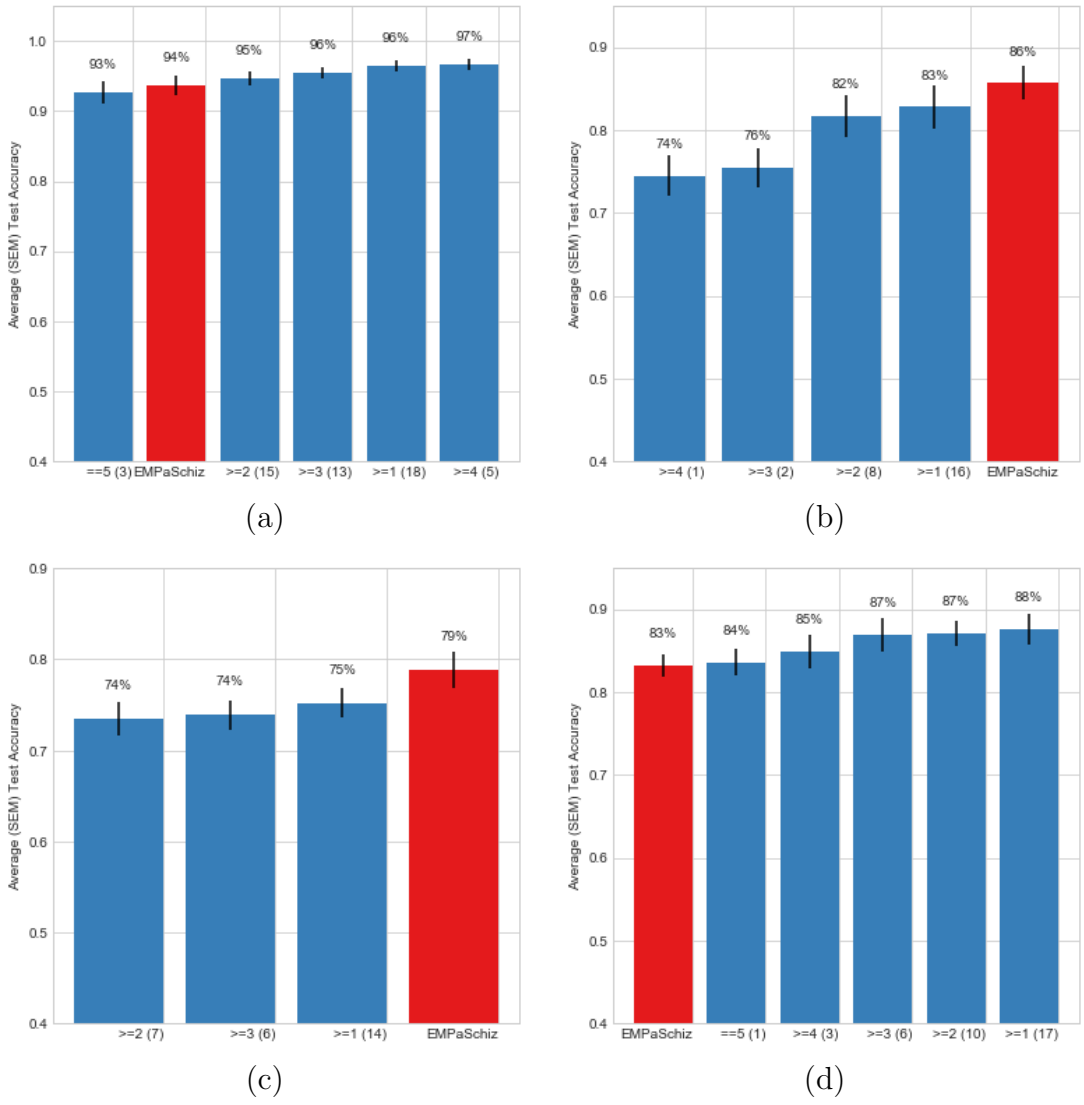


Figure A.6: Comparison across EMPaSchiz and Feature selection based EMPaSchiz using 5 fold cross-validation - 5 shuffled iterations : a) SCZ_to_OCD, b) OCD_to_SCZ, c) SCZ_to_OCD_CommonHC, and d) OCD_to_SCZ_CommonHC transfer learning model (X axis denotes the number of selected models out of 5 iterations).

Feature Category	Features Name
Contamination	Time score Distress score Interface score Resictance score Control score
Hoarding and Collecting	Time score Distress score Interface score Resistance score Control score
Symmetry,Ordering,Counting, and Arranging	Time score Distress score Interface score Resistance score Control score
Aggression & related compulsion	Time score Distress score Interface score Resistance score Control score
Sexual and Religious	Time score Distress score Interface score Resistance score Control score
Miscellaneous	Time score Distress score Interface score Resistance score Control score
Pathological Doubt DYBOC	Time Score Distress Score Interface score Resistance Score Pathological Control Score

Table A.7: 35 Measures under DYBOCS scale.

Appendix B

Schizophrenia Prognosis Prediction

B.1 Treatment response prediction on tDCS Schizophrenia Patients

B.1.1 Average Voxels Features based Methods

Initially, we projected the extracted LTPJ and RDLPFC based pearson correlation features onto 14 different parcellations schemes (mentioned in Section 3.3.2), and extracted average voxels features per region. It yielded 28 (2 feature types $\times$ 14 parcellations) feature sets. In Figure B.1, $F1$ and $F2$ denote the LTPJ, and RDLPFC seed-based pearson correlation features, respectively. And, P_j denotes the j-th parcellation, where j = 1, 2, ..., 14. Feature sets are shown as F_iP_j (i.e. F_1P_1 to F_2P_{14} each correspond to specific feature set), where i denotes the feature type, and j denotes the parcellation.

We applied PCA on each feature set separately to reduce the feature dimensionality. Here, we used all the PCA components –i.e., the number of PCA components is equal to the smaller values between the number of samples and the number of original features. After that, for each reduced feature set, we trained a different L2-logistic regression to predict the treatment response; we consider these 28 classifiers as 28 single source models (SSMs). We tried the below three different procedures for building an ensemble model from these single source models.

(a) **Using Meta-Classifier:**

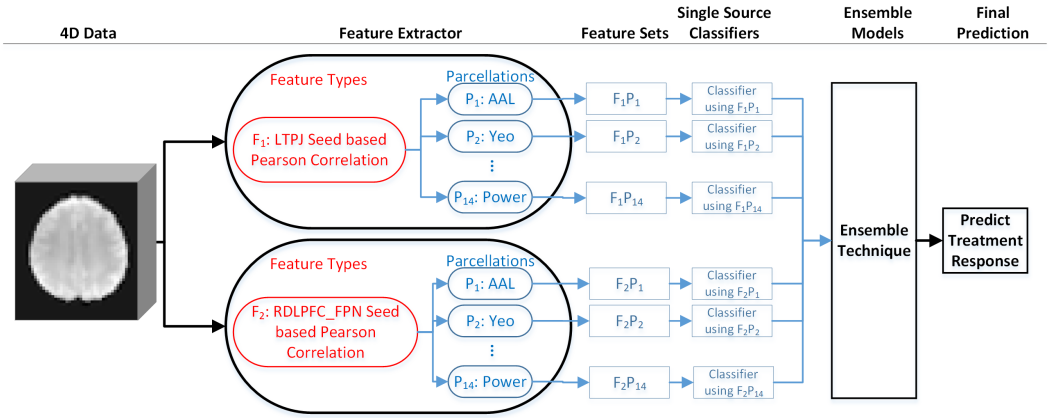


Figure B.1: Average voxels features based ensemble model.

(i) **Logistic Regression:** We took outputs (predicted probabilities) of 28 SSMs' predictions on the training set. After that, we passed these classes probabilities to a classifier, L2-logistic regression, to train the final ensemble model.

(ii) **Using Majority Voting:** Instead of taking the classes probabilities from 28 single source models' prediction on each training/ testing sample, we considered the predicted class labels (here we used 0.5 as threshold computing the class labels), and then applied the majority voting to determine the final class label.

(b) **Adaptive Thresholding:** At first, we computed the average of 28 SSMs' predicted probabilities on the training set, and then the method selected a threshold value by optimizing the objective function. Here, the objective function is to maximize the training accuracy by trying different threshold values on the above computed average probabilities. During the performance time on the test set, the model used the selected threshold on the mean outputs (or probabilities) of 28 single source models' predictions to compute the class label.

B.1.2 Individual Voxels Features based Methods

Similar to previous described "average voxels features" based method, we firstly projected extracted LTPJ and RDLPFC based pearson correlation fea-

tures onto 11 different parcellations schemes (power [104], dosenbach [30], yeo [142], aal [147], basc_multiscale_122, basc_multiscale_197, basc _multiscale_325, basc _multiscale_444 [11], destrieux [28], harvard_cort_25, and harvard_sub_25[1]). Instead of extracting average voxels signals, we extracted raw signal from each voxel per region – regions are predefined by each parcellation – for each parcellation (the number of regions of each parcellation is shown in Table B.1) and each feature type.

For each region, we have an individual feature set that contains all the voxels features of that particular region. It yielded 232 (2 feature types $\times$ 116 regions) feature sets for aal, and similarly for other parcellations. So, in total, we have 3822 ($2 \times 116 + 2 \times 264 + 2 \times 122 + 2 \times 197 + 2 \times 325 + 2 \times 444 + 2 \times 17 + 2 \times 22 + 2 \times 96 + 2 \times 148 + 2 \times 160$) feature sets. As we have two types of features for each region, we concatenated these two feature types for each region. As a result, there were 1911 (3822 feature sets / 2 feature types) feature sets for building an ensemble classification model.

Parcellation	Number of Regions
aal	116
power	264
basc_multiscale_122	122
basc_multiscale_197	197
basc_multiscale_325	325
basc_multiscale_444	444
yeo	17
harvard_sub_25	22
harvard_cort_25	96
destrieux	148
dosenbach	160

Table B.1: Number of regions in each parcellation scheme.

From these 1911 feature sets, we trained 1911 different logistic regression classifiers (we considered each classifier as a single source model) to predict the treatment response. We tried ensemble procedures (meta-classifiers, and adaptive thresholding methods), which also are used for "average voxels fea-

[1] `http://www.cma.mgh.harvard.edu/fsl_atlas.html`

tures" based methods, for building an ensemble model from these single source models. The model structure is shown in Figure B.2.

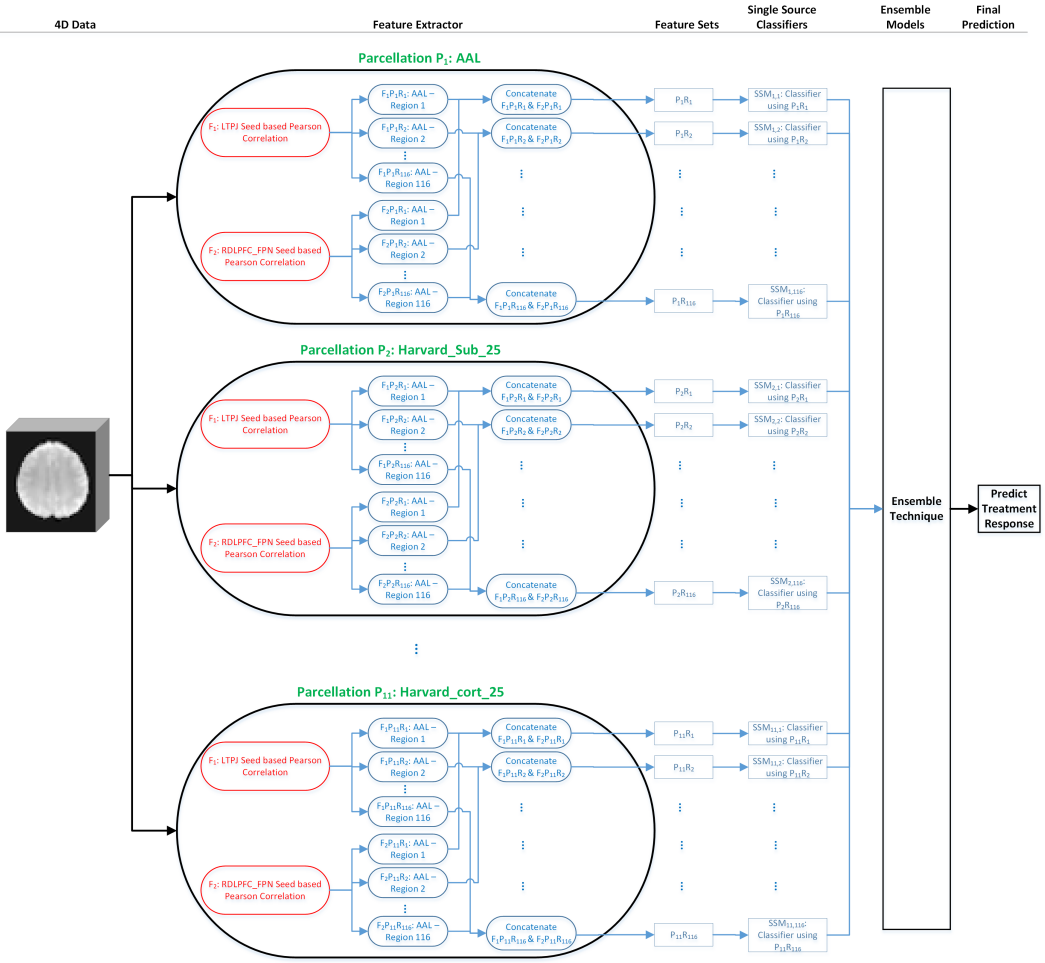


Figure B.2: Individual voxels features based ensemble model.

B.1.3 Feature Concatenation

a) **Multiple Parcellations based Methods (MPM):**

i) **Average Voxels Features based Method** Feature extractor of the "average voxels features" based method (shown in Figure B.1) provides 28 feature sets. Then, we simply concatenated all of these 28 feature sets and built a L2-logistic regression model for predicting the treatment response.

ii) **Individual Voxels Features based Method** Similarly, we got 3822 feature sets from the feature extractor of the "individual voxels features" based method (shown in Figure B.2). Due to the huge amount of features, we built a classification model using Logistic regression with L1-regularization, where L1-regularization helps to achieve sparsity.

b) **Selected Regions based Method (SRM):** Instead of considering all of the brain regions based on different parcellations, we considered few regions, those are implicated during Schizophrenia Auditory hallucination and it's treatment according to the previous works [62], [67], [113], [169]. We used 2 atlases (harvard_sub_25, harvard_cort_25) for extracting these regional features. List of selected regions is provided in Table B.2.

i) **Average Voxels Features based Method** We extracted the average voxels features (described in "Average Voxels Features" based Methods section) from each of the selected regions. After that, we concatenated all the features, and learnt a L2-regularized logistic regression.

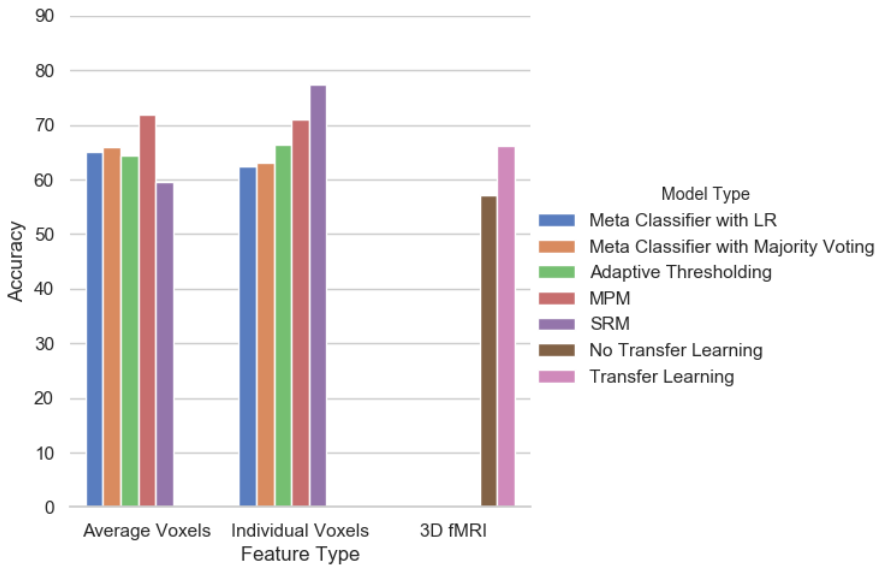


Figure B.3: Comparison of performance across different methods for tDCS-SCZ treatment response.

B.1.4 Discussion

From these analysis, we found that we were able to achieve 72% and 71% accuracies using the predefined set of atlases as prior knowledge for the average voxels and individual voxels features based methods, respectively. These results inspired us to add more knowledge in our learning model in terms of selecting a set of regions (refer to Table B.2) based on apriori knowledge.

Selected regions based methods (SRMs) gave 71.0% and 77.5% accuracies using average voxels and individual voxels features, respectively (refer to Figure B.3). When we used the average voxels features of those selected regions, we were unable to achieve significant performance. It indicates that we might lose some information through the process of averaging the voxel of those regions.

B.2 Treatment response prediction on Antipsychotic Schizophrenia Patients

For two datasets (SAPS and CGI dataset), none of deep learning methods were able to provide significant better performance above from the baseline. Results for these two datasets are provided in Tables B.4, and B.5.

SL No.	Brain Area	Reference
1.	Left Superior Temporal Gyrus, anterior division	Zmigrod et al. 2016 [169], Kühn and Gallinat 2012 [67], Jardri et al. 2011 [62]
2.	Left Superior Temporal Gyrus, posterior division	
3.	Right Superior Temporal Gyrus, anterior division	Zmigrod et al. 2016 [169]
4.	Right Superior Temporal Gyrus, posterior division	
5.	Left Insula Cortex	Zmigrod et al. 2016 [169], Rollins et al. 2019 [113], Jardri et al. 2011 [62]
6.	Right Insula Cortex	Zmigrod et al. 2016 [169]
7.	Right Inferior Frontal Gyrus, pars triangularis	Zmigrod et al. 2016 [169], Rollins et al. 2019 [113]
8.	Right Inferior Frontal Gyrus, pars opercularis	
9.	Left Precentral Gyrus	Jardri et al. 2011 [62]
10.	Right Precentral Gyrus	Zmigrod et al. 2016 [169]
11.	Left Postcentral Gyrus	Kühn and Gallinat 2012 [67]
12.	Right Postcentral Gyrus	
13.	Left Cingulate Gyrus, anterior division	Kühn and Gallinat 2012, Rollins et al. 2019 [113], Jardri et al. 2011 [62]
14.	Left Middle Temporal Gyrus, anterior division	Kühn and Gallinat 2012 [67], Rollins et al. 2019 [113], Zmigrod et al. 2016 [169]
15.	Left Middle Temporal Gyrus, posterior division	
16.	Left Middle Temporal Gyrus, temporooccipital part	
17.	'Left Inferior Frontal Gyrus, pars triangularis'	Kühn and Gallinat 2012 [67]
18.	Left Inferior Frontal Gyrus, pars opercularis	
19.	Left Parietal Operculum Cortex	
20.	Right Supramarginal Gyrus, anterior division	Rollins et al. 2019 [113]
21.	Right Supramarginal Gyrus, posterior division	
22.	Left Supramarginal Gyrus, anterior division	Jardri et al. 2011 [62]
23.	Left Supramarginal Gyrus, posterior division	
24.	Left Hippocampus	
25.	Right Hippocampus	
26.	Left Parahippocampal Gyrus, anterior division	
27.	Right Parahippocampal Gyrus, anterior division	
28.	Left Parahippocampal Gyrus, posterior division	
29.	Right Parahippocampal Gyrus, posterior division	

Table B.2: Regions selected based on prior research works.

Table B.3: Comparison of performance across different methods for tDCS-SCZ treatment response prediction.

Feature Type	Model Type		Accuracy	Precision	Sensitivity	Specificity	True positive	True negative	False positive	False negative
Avearge Voxels	Meta-Classifier	Logistic Regression	65.0 (3.5)	69.2 (4.1)	65.0 (4.7)	65.0 (5.1)	11.0 (0.3)	10.8 (0.7)	6.2 (0.7)	6.0 (0.3)
		Majority Voting	66.0 (4.2)	72.5 (4.6)	64.0 (4.9)	68.0 (5.8)	10.6 (0.2)	11.2 (0.2)	5.8 (0.2)	6.4 (0.2)
	Adaptive Thresholding		64.5 (3.6)	68.1 (4.3)	64.0 (4.9)	65.0 (5.1)	10.8 (0.3)	10.8 (0.7)	6.2 (0.7)	6.2 (0.3)
	MPM		72.0 (3.3)	77.8 (3.8)	67.0 (4.8)	77.0 (4.3)	11.0 (0.4)	12.8 (0.2)	4.2 (0.2)	6.0 (0.4)
	SRM		59.5 (4.3)	61.2 (5.0)	60.0 (5.5)	59.0 (5.6)	10.2 (0.7)	9.8 (0.8)	7.2 (0.8)	6.8 (0.7)
Individual Voxels	Meta-Classifier	Logistic Regression	62.5 (3.8)	62.9 (4.1)	72.0 (4.7)	53.0 (5.4)	11.8 (0.3)	9.0 (0.6)	8.0 (0.6)	5.2 (0.3)
		Majority Voting	63.0 (3.6)	65.9 (4.5)	60.0 (5.8)	66.0 (5.4)	10.4 (0.2)	11.0 (0.0)	6.0 (0.0)	6.6 (0.2)
	Adaptive Thresholding		66.5 (3.9)	72.0 (4.4)	65.0 (5.3)	68.0 (5.6)	11.0 (0.0)	11.6 (0.2)	5.4 (0.2)	6.0 (0.0)
	MPM		71.0 (3.7)	69.7 (3.8)	83.0 (4.4)	59.0 (5.0)	14.4 (0.2)	10.0 (0.9)	7.0 (0.9)	2.6 (0.2)
	SRM		**77.5 (3.1)**	77.6 (3.5)	87.0 (3.7)	68.0 (5.1)	14.6 (0.5)	11.6 (0.2)	5.4 (0.2)	2.4 (0.5)
3D fMRI	3D CNN Model		57.2 (3.1)	62.5 (5.4)	48.0 (6.3)	64.0 (6.1)	8.0 (0.8)	11.4 (1.3)	5.6 (1.3)	9.0 (0.9)
	Transfer Learning		66.2 (3.0)	70.83 (4.0)	71.0 (5.1)	62.0 (5.8)	12.0 (0.4)	10.6 (0.8)	6.4 (0.8)	5.0 (0.4)

Model Name	Class Labels based on	
	Percentage of Improvement (Baseline: 50.94%)	Follow-up SAPs Score (Baseline: 54.71%)
3D CNN Model using LTPJ and RDLPFC	52.83 ± 2.44	52.45 ± 1.2
Densenet Model using LTPJ and RDLPFC	50.56 ± 1.47	50.19 ± 3.13
3D CNN Model using ALFF	51.33 ± 1.72	50.18 ± 0.86
3D CNN Model using fALFF	55.47 ± 2.48	50.94 ± 0.80
3D CNN Model using ReHo	52.83 ± 0.92	53.20 ± 0.98
Stacked 3D CNN Model Using LTPJ, RDLPFC and ReHo	54.34 ± 1.24	52.08 ± 2.75
Transfer Learning using LTPJ and RDLPFC	54.72 ± 1.84	55.47 ± 0.67

Table B.4: Treatment response prediction results based on baseline and follow-up SAPS scores: average (standard errors)- 5×10-fold CV.

Model Name	Class Labels based on	
	Percentage of Improvement (Baseline: 57.57%)	Follow-up CGIs Score (Baseline: 50.51%)
3D CNN Model using LTPJ and RDLPFC	55.75 ± 0.66	51.0 ± 1.24
Densenet Model using LTPJ and RDLPFC	54.31 ± 1.06	51.41 ± 0.58
3D CNN Model using ALFF	55.96 ± 1.19	51.77 ± 0.32
3D CNN Model using fALFF	53.53 ± 0.57	52.42 ± 0.38
3D CNN Model using ReHo	55.55 ± 0.49	50.41 ± 0.54
Stacked 3D CNN Model Using LTPJ, RDLPFC and ReHo	55.53 ± 1.08	51.18 ± 0.86
Transfer Learning using LTPJ and RDLPFC	56.16 ± 1.04	53.34 ± 1.49

Table B.5: Treatment response prediction results based on baseline and follow-up CGI scores: average (standard errors)- 5×10-fold CV.