Schizophrenia Research 214 (2019) 51–59

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Identifying schizophrenia subgroups using clustering and supervised learning

Alexandra Talpalaru [a,b,*], Nikhil Bhagwat [b,c], Gabriel A. Devenyi [b,d], Martin Lepage [b,d], M. Mallar Chakravarty [a,b,d,*]

[a] Biological & Biomedical Engineering, McGill University, 845 Sherbrooke Street West, Montreal, Quebec H3A 0G4, Canada
[b] Douglas Mental Health University Institute, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada
[c] Institute of Biomaterials and Biomedical Engineering, University of Toronto, 27 King's College Cir, Toronto, ON M5S 3H7, Canada
[d] Department of Psychiatry, McGill University, 845 Sherbrooke Street West, Montreal, Quebec H3A 0G4, Canada

ARTICLE INFO

Article history:
Received 2 July 2018
Received in revised form 28 May 2019
Accepted 30 May 2019
Available online 24 August 2019

Keywords:
Schizophrenia
Heterogeneity
Single-subject prediction
MRI
Clustering
Machine learning

ABSTRACT

Schizophrenia has a 1% incidence rate world-wide and those diagnosed present with positive (e.g. hallucinations, delusions), negative (e.g. apathy, asociality), and cognitive symptoms. However, both symptom burden and associated brain alterations are highly heterogeneous and intimately linked to prognosis. In this study, we present a method to predict individual symptom profiles by first deriving clinical subgroups and then using machine learning methods to perform subject-level classification based on magnetic resonance imaging (MRI) derived neuroanatomical measures. Symptomatic and MRI data of 167 subjects were used. Subgroups were defined using hierarchical clustering of clinical data resulting in 3 stable clusters: 1) high symptom burden, 2) predominantly positive symptom burden, and 3) mild symptom burden. Cortical thickness estimates were obtained in 78 regions of interest and were input, along with demographic data, into three machine learning models (logistic regression, support vector machine, and random forest) to predict subgroups. Random forest performance metrics for predicting the group membership of the high and mild symptom burden groups exceeded those of the baseline comparison of the entire schizophrenia population versus normal controls (AUC: 0.81 and 0.78 vs. 0.75). Additionally, an analysis of the most important features in the random forest classification demonstrated consistencies with previous findings of regional impairments and symptoms of schizophrenia.

1. Introduction

The clinical presentation of schizophrenia contains numerous symptom dimensions that are differentially expressed in individuals who suffer from this disorder (Tandon et al., 2009). The interaction between various risk factors (genetic, environmental, social, and lifestyle) have been implicated in the onset of schizophrenia, potentially impacting the development of a diverse number of neural circuits (Selemon and Zecevic, 2015; Patel et al., 2014). Differential exposure and dosing of these risk factors (during critical periods), and individual-specific susceptibility factors, may result in the heterogeneity commonly observed in responsivity to treatment, frequencies and lengths of active and remissive periods, different levels of functional impairment, and diverging functional outcomes (Selemon and Zecevic, 2015; Tandon et al., 2009). Several anatomical, volumetric, and morphological brain abnormalities have previously been observed in schizophrenia including overall reductions in brain volume and increases in ventricular sizes noted in postmortem analyses (Brown et al., 1986), and decreases in the volume and cortical thickness of frontal and temporal regions measured from magnetic resonance imaging (MRI) data (Van Erp et al., 2018; Kuperberg et al., 2003; Honea et al., 2005; Glahn et al., 2008). Correlations between regional cortical thinning and schizophrenia symptomatology has been demonstrated in large scale studies (Walton et al., 2018, 2017). However, similar to clinical presentation, structural alterations of schizophrenia patients are also immensely heterogeneous. This variability is thought to reflect multiple etiologies of this spectrum disorder, which can be identified in data-driven ways by grouping patients with similar clinical or anatomical presentation.

Significant evidence exists for the presence of clinical subtypes of schizophrenia (Crow et al., 1986; Liddle, 1987; Dollfus et al., 1996; Carpenter Jr et al., 1976) which in turn have been shown to have distinct patterns of structural alterations (Nenadic et al., 2015; Nenadic et al., 2010; Koutsouleris et al., 2008). Conversely, data-driven methods have also been implemented to subtype patients based on structural

* Corresponding authors at: Computational Brain Anatomy Laboratory, Brain Imaging Center, Douglas Mental Health University Institute, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada.
E-mail addresses: alexandra.talpalaru@mail.mcgill.ca (A. Talpalaru), mallar@cobralab.ca (M.M. Chakravarty).

https://doi.org/10.1016/j.schres.2019.05.044
0920-9964/

and functional biomarkers (Yang et al., 2014; Clementz et al., 2016; Dwyer et al., 2018; Brodersen et al., 2014). Thus, understanding the link between brain structure and symptom heterogeneity may be a critical first step to improving outcomes for patients. For example, previous studies have linked the persistence of negative symptoms to greater brain abnormalities to explain poorer functional outcomes (Li et al., 2018; Makowski et al., 2017)

Machine learning methods are powerful in identifying patterns of structural and functional impairment in widely-distributed brain regions for single-subject prediction tasks in schizophrenia (Zarogianni et al., 2013; Arbabshirani et al., 2017; Davatzikos et al., 2005). Current classifiers leveraging neuroimaging-based biomarkers are able to differentiate patients suffering with schizophrenia from normal controls with a sensitivity and specificity of about 80% (Kambeitz et al., 2015). Higher accuracies have been reported when patients are grouped based on symptom presentation (Nenadic et al., 2010). Improvements in predictive power based on clinical variables has also been demonstrated when patients have first been stratified into biologically homogeneous subgroups (Dwyer et al., 2018). Thus, defining clinically meaningful subgroups and using neuroanatomy to predict them may be an important step forward to understanding the neural basis for heterogeneity observed in schizophrenia.

In this manuscript, we use a data-driven approach to characterize the clinical heterogeneity in a schizophrenia sample. Further, we assess the feasibility of using neuroanatomical variables as predictors of individual clinical profiles. This combines two methodologies: 1) clinical variables were used to produce subgroups in a data-driven manner, then 2) cortical thickness features were used to predict symptom profiles.

Data from SchizConnect (http://schizconnect.org) of 104 patients and 63 normal controls was used. Hierarchical clustering was performed on the patient symptom severity data, resulting in three stable clusters and representing patients with high symptom burden, predominantly positive symptom burden, and low symptom burden. Demographic variables and the average cortical thickness in 78 brain regions defined by the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002) parcellations were used as input features into three machine learning algorithms (logistic regression, support vector machine, and random forest), and the subgroups as class labels. Random forest performance metrics for predicting the group membership of the high symptom burden and the mild symptom burden groups exceeded those of the baseline comparison of all patients versus normal controls. Further, the cortical regions that were the most informative predictors in each random forest classification task were different for each subgroup, indicating distinct neuroanatomical impairments in each subtype. Additionally, important features in the subgroup classification task were shown to be consistent with previous findings of regional impairments (Glahn et al., 2008; Bora et al., 2011; Honea et al., 2005) and symptom associations in schizophrenia (Allen et al., 2008; Sumich et al., 2005; Walton et al., 2017; Wylie and Tregellas, 2010) (e.g.: right superior temporal gyrus (Shenton et al., 2001; Sun et al., 2009), left Heschl's gyrus (Hirayasu et al., 2000) and the right insula (Nesvåg et al., 2008) for the high symptom burden group, the bilateral insula (Bora et al., 2011) for the predominantly positive symptom burden group, and the right anterior cingulate and paracingulate gyri and left insula (Kuperberg et al., 2003) for the mild symptom burden group). Once validated, integration of this type of technique in the clinic may potentially improve patient specific diagnosis and personalized treatment options.

2. Materials and methods

2.1. Dataset

The T1-weighted magnetic resonance imaging (MRI) data and the scales for the assessment of positive and negative symptoms (SAPS/SANS) clinical measures of the Northwestern University Schizophrenia Data and Software Tool (NUSDAST) dataset (Wang et al., 2013) were downloaded from the SchizConnect website (http://schizconnect.org/). Demographics of the included subjects and the total scores of all SAPS and SANS items are shown in Table 1.

Table 1
Demographic characteristics of the included subjects of the NUSDAST dataset.

Demographic	Schizophrenia patients Mean (standard deviation)	Normal controls Mean (standard deviation)
Sex	63 M/41 F	35 M/28 F
Age	33.1 (12.4)	26.1 (10.5)
SAPS	24.2 (18.0)	–
SANS	31.8 (19.1)	–

2.2. MRI pre-processing and cortical thickness estimation

MRI pre-processing was performed using the automated minc-bpipe-library pipeline (https://github.com/CobraLab/minc-bpipe-library). Briefly, bpipe performs bias field correction using a variant of N4ITK (Tustison et al., 2010), registration to a common space using a 12 parameter affine transformation (Collins et al., 1994), removal of non-head regions (such as the neck), and brain extraction using BEaST (Eskildsen et al., 2012). Then, cortical thickness was estimated at 81,924 vertices using the fully automated CIVET pipeline (http://www.bic.mni.mcgill.ca/ServicesSoftware/CIVET, version: 1.1.12) (Ad-Dab'bagh et al., 2006; Zijdenbos et al., 2002). The typical image processing steps of CIVET are: linear registration to the MNI ICBM 152 average (Collins et al., 1994), non-uniformity correction (Sled et al., 1998), brain extraction (Smith, 2002), tissue classification into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) using priors derived from nonlinear registration and accounting for partial volume effects (Tohka et al., 2004), GM-WM and GM-CSF surface extraction and registration (MacDonald et al., 2000; Kim et al., 2005; Kabani et al., 2001), and estimation of cortical thickness at 81,924 vertices (Lerch and Evans, 2005). All the steps of CIVET were performed except for the non-uniformity correction, which was performed using the improved N4ITK as a pre-processing step (Tustison et al., 2010). The cortical thickness values measured by CIVET were reduced to 78 mean cortical thickness values using parcellations defined by the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

2.3. Machine learning analyses

The clinical heterogeneity in a schizophrenia dataset was characterized by demarcating subtypes using clustering of SAPS/SANS clinical variables. Further, cortical thickness features were used as predictors of clinical profiles (i.e. cluster memberships) at the single-subject level.

2.3.1. Cluster definition and stability

Agglomerative hierarchical clustering was implemented on z-scored clinical data (59 SAPS/SANS items) of the study participants with a schizophrenia diagnosis ($N = 104$), to identify clinical subgroups of disease presentation. Hierarchical clustering was performed by building a dendrogram, where subject similarity was Euclidean distance, and group linkage was based on Ward's criterion (Ward, 1963). The number of clusters present in the dataset was determined based on a cluster stability analysis introduced by Ben-Hur et al. (Ben-Hur et al., 2002). Briefly, two subsamples were randomly selected, each consisting of 80% of the original number of study participants and then independently clustered, into 2-, 3-, and 4- clusters. In each of the two subsamples, a comparison matrix is generated where an entry of 1 is given if two study participants are clustered together. The similarity between the comparison matrices of the two subsamples is assessed by computing the fraction of entries on which the two agree, in the set of subjects that they have in common, resulting in a matching fraction between 0

and 1 (Ben-Hur et al., 2002). To probe stability, this process was repeated 1000 times, for each of the 2-, 3-, and 4- number of clusters, and plotted as individual histograms. Permutation analysis was performed to ensure that the clusters are representative of distinct groups irrespective of the input dataset. Differences in the distributions of stability fractions between each of the histograms were assessed using Kolmogorov-Smirnov tests (Kolmogorov, 1992; Smirnov, 1939). The optimal number of clusters was selected based on the histogram distribution where the majority matching fractions in the permutations were greater than random, and which significantly differed from other cluster solutions. The 5-cluster solution was not significantly different from the 4-cluster solution ($p = 0.08$) based on a two-sample K-S test, therefore it was not included in the analysis. The 4-cluster solution shows very poor stability, being left-skewed, with the center on 0.5. The 2-cluster solution showed the greatest stability, with most matching fractions exceeding 0.6. However, an analysis of the heatmap, Fig. 1A, information indicates that the second cluster contains a subgroup of patients with higher SAPS values, evident in the warmer colors grouped together on the right side and near the center of the heatmap; thus, partitioning the patients into two clusters does not created clinically homogeneous

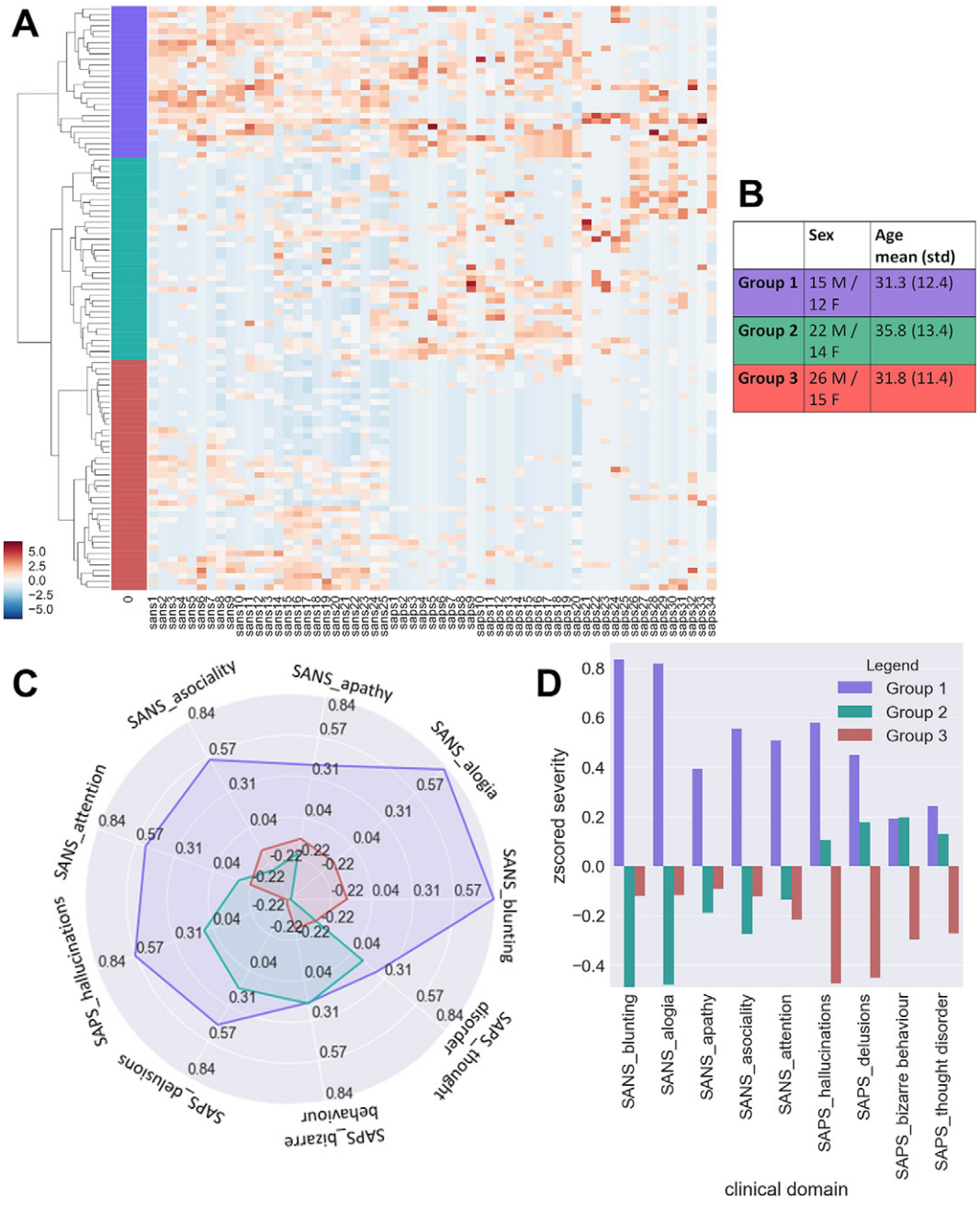

Fig. 1. A) Heatmap and left-adjacent dendrogram (tree-diagram) representing agglomerative hierarchical clustering of the 104 NUSDAST study participants (rows of the heatmap) according to their 59 z-scored SAPS/SANS clinical features (columns of the heatmap). The blue, green, and red delineations illustrate the 3-cluster solution, separating patients into those having high SAPS/SANS, high SAPS, and low SAPS/SANS respectively, relative to the average. B) Group-wise demographic information of the three clusters. C) Radar plot and D) bar graph of the average z-scored SAPS/SANS items in the high symptom burden (Group 1, blue), predominantly positive symptom burden (Group 2, green), and mild symptom burden (Group 3, red) groups averaged across the major categories of the SAPS (hallucinations, delusions, bizarre behaviour, and positive formal thought disorder) and SANS (blunting, alogia, apathy, asociality, attention) scales.

groups. The 3-cluster solution was relatively stable, with a higher proportion of permutations having an acceptable value ≥0.6, as compared to those falling on or below 0.5. Additionally, the 3- and 4-cluster solutions had significantly different distributions ($p = 2.4 \times 10^{-18}$, K-S test), and the 4-cluster solution had only 8 patients in one of the clusters; thus, the 3-cluster solution was considered to be more reliably stable for subsequent analyses. Based on the results of these histograms, the 3-cluster solution (3 subgroups) was selected for further analysis.

2.3.2. Single-subject prediction using supervised learning

Logistic regression (LR), support vector machine (SVM), and random forest (RF), were used to perform single-subject prediction into their respective class (schizophrenia vs. normal control, clinical subgroup vs. normal control, subgroup vs. subgroup). These classifiers have been previously successfully applied in schizophrenia classification studies (Arbabshirani et al., 2017; Zarogianni et al., 2013), and are relatively simple, easily interpretable, and robust. LR was chosen because it has few hyperparameters and is robust to overfitting with regularization. SVM and RF were chosen because they are both discriminative, capable of handling large amounts of data, and can capture non-linear relationships across input features. Three classification methods were investigated to provide a comparison between methods commonly used in literature and prevent making a priori assumptions on the results. Additionally, given the small sample size, achieving similar results with different models increases the confidence of our conclusions. Further, we have commonly employed this strategy in the examination of subject-level classification using neuroanatomical measures (Winterburn et al., 2017; Bhagwat et al., 2018).

All study participants in the NUSDAST cohort ($N = 167$) were used, with features of each subject being the mean cortical thickness measurements in the 78 AAL atlas regions and age and sex demographic information (80 input features). Each task was set up as a binary classification problem, where the prediction target is either a 0 or 1. First, the classification task of classifying all schizophrenia patients versus all normal controls was performed as the case-control study. This case-control task gives a baseline for comparison for the clinical subgroup classifications. Next, three classification tasks consisting of each of the three clinical subgroups versus normal controls was performed, to assess the ability of differentially diagnosing across subgroups. Finally, each clinical subgroup was classified against every other clinical subgroup. Demographic information is included as inputs to the models to account for the known effects of age (Tamnes and Østby, 2018), and sex differences (Ruigrok et al., 2014) on brain structure in both normal development and psychiatric disorders. Studies of normal brain development consistently observe reductions in cortical thickness beginning in childhood, continuing through the adolescent and adulthood periods (Lemaitre et al., 2012). Furthermore, sex differences in the trajectories of normal cortical maturation (Mutlu et al., 2013), and in schizophrenia (Narr et al., 2005) have been reported. Sex differences in prevalence (McGrath et al., 2008), age of onset (Abel et al., 2010), clinical presentation (Abel et al., 2010), and functional outcome (Grossman et al., 2008) are also observed. Also, from a feature selection perspective, inclusion of demographic information improves feature learning and predictive performance of classifiers (Struyf et al., 2008).

Stratified nested 5-fold cross-validation was performed to optimize the hyperparameters and assess the performance of each of the models (LR, SVM, RF). This procedure consists of 5-fold cross-validation in the outer loop, and 3-fold cross-validation in the nested loop as shown in Supplementary Fig. 1. In the inner loop, gridsearch was implemented to exhaustively search the space of all combinations from reasonable hyperparameter ranges (see Supplementary material). From the outer loop, several performance metrics can be calculated, including accuracy, sensitivity and specificity (see Supplementary Fig. 3). For our binary classification task, the receiver operating characteristic (ROC) curve was generated by plotting (1-specificity) versus sensitivity at various thresholds. The area under the ROC curve (AUC) is reported as the performance metric for each the test sets. Thus, the overall performance of the model is reported as the average performance over the five test sets. All machine learning analysis was performed in the Python programming language using the Scikit-learn library (http://scikit-learn.org/stable/, version: 0.19.1) (Garreta and Moncecchi, 2013).

2.3.3. Testing the generalizability of the proposed methods

There is evidence to suggest that attempting to cluster subjects into groups and then use machine learning techniques to predict this class membership may result in overfitting in some cases (Dinga et al., 2018). To examine if this was the case in the proposed framework, a proof-of-concept analysis was performed where within each of the 5 folds of the cross-validation task, clusters were generated, and the entire classification task was re-ran with these new labels in each of the 5 folds. AUC was used for pairwise comparison of the performance in each of the folds (Supplementary Table 1).

3. Results

3.1. Schizophrenia subgroup determination using hierarchical clustering

Hierarchical clustering results are shown in Fig. 1. The 2-, 3-, and 4-cluster stability analyses are shown in Supplementary Fig. 2A), B), and C), respectively, indicating the frequency of each matching fraction.

Dividing the dendrogram into three clusters (blue, green, and red in Fig. 1A) partitions the patients into those with (1) high loads of both negative and positive symptoms (blue), (2) predominantly positive symptoms (green), and (3) mild symptom burden (red), as compared to the average population (Fig. 1C and D show the average z-scored SAPS and SANS items across the subscales). The demographics of these groups are given in Fig. 1B, showing that the groups contain a similar number of study participants (Group 1: 27, Group 2: 36, Group 3: 41), have similar average ages (Group 1: 31.3, Group 2: 35.8, Group 3: 31.8), but have a higher proportion of males in each cluster (Group 1: 15 M/12 F, Group 2: 22 M/14 F, Group 3: 26 M/15 F).

3.2. Single-subject prediction of individual symptom profiles in schizophrenia

3.2.1. Case-control classification

Schizophrenia population vs. normal controls results are summarized in the first row of Fig. 2A as ROC curves, and in Fig. 2B summarizing AUC values. These figures show that the case-control classifiers perform well above chance and have similar performance (LR: 0.69 ± 0.13, RF: 0.75 ± 0.14, SVM: 0.71 ± 0.08), with RF performing the best.

3.3. Subgroups versus normal controls classification

Fig. 2B summarizes the AUC values of the classification results of groups 1 (high symptom burden), 2 (predominantly positive symptoms) and 3 (mild symptom burden) versus normal controls for each of the models. The AUC of the group 1 classification is consistently better than the case-control across all of the classifiers (LR: 0.74 ± 0.13, RF: 0.81 ± 0.11, SVM: 0.80 ± 0.14). However, the AUC of the group 2 classification is poorer than the case-control study for all classifiers (LR: 0.61 ± 0.11, RF: 0.61 ± 0.13, SVM: 0.61 ± 0.04). Additionally, for the RF model, the group 3 classification shows slight improvement in AUC over the baseline comparison (LR: 0.65 ± 0.11, RF: 0.78 ± 0.08, SVM: 0.63 ± 0.09). To assess the significance of the subgroup classifications performing better than the case-control study, the nonparametric Mann-Whitney-*U* test (Mann and Whitney, 1947) was performed on the AUC results following 10 repetitions of the 5-fold cross-validation tasks. Group 1 (high symptom burden; RF: $p = 0.013$) classification was significantly better than the case-control classification only for the RF classifier (LR: $p = 0.14$, SVM: $p = 0.35$). Additionally, the RF group

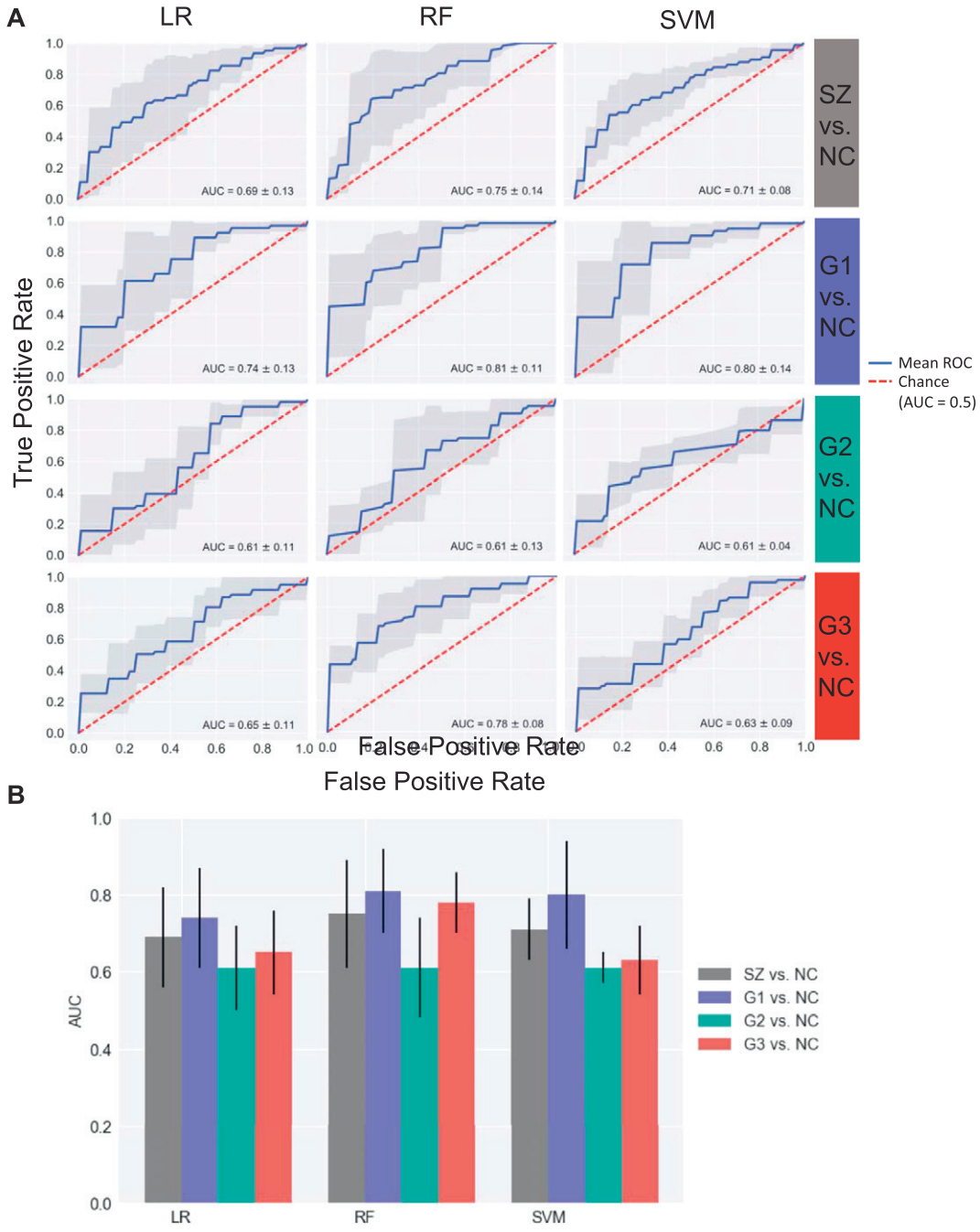

Fig. 2. A) LR, SVM, and RF receiver operator characteristic (ROC) curves for the four comparisons: all schizophrenia patients versus normal controls (SZ vs NC), group 1 (high symptom burden) (G1 vs NC), group 2 (predominantly positive symptom burden) (G2 vs NC), and group 3 (mild symptom burden) (G3 vs NC) versus normal controls. B) Summary of classification results of LR, RF, and SVM on patients in Group 1 (G1 vs. NC; blue), 2 (G2 vs. NC; green), 3 (G3 vs. NC; red), and all schizophrenia patients (SZ vs. NC; gray) versus normal controls. Legend: LR = logistic regression, RF = random forest, SVM = support vector machine, ROC = receiver operator characteristic, AUC = area under ROC curve, SZ = schizophrenia, NC = normal control, G1 = high symptom burden group, G2 = predominantly positive symptom burden group, G3 = mild symptom burden group.

3 (mild symptom burden) classification was significantly better than the case-control study ($p = 0.022$).

3.3.1. Subgroup by subgroup classification

The AUC values of the classification results showing the group 1 (G1; high symptom burden) versus group 2 (G2; predominantly positive symptoms), group 1 versus group 3 (G3; mild symptom burden) and group 2 versus group 3 are shown in Supplementary Fig. 4. From these results we can see that the AUCs for all the comparisons, across all the classifiers, was around chance (AUC: 0.50).

3.4. Analysis of random forest classification results

To assess which cortical regions were driving the classification in the three subgroups, the top ten most relevant features in the RF algorithm, were extracted (Fig. 3, full list of regions provided in Supplementary Table 2), where individual colors are not meaningful. The left anterior

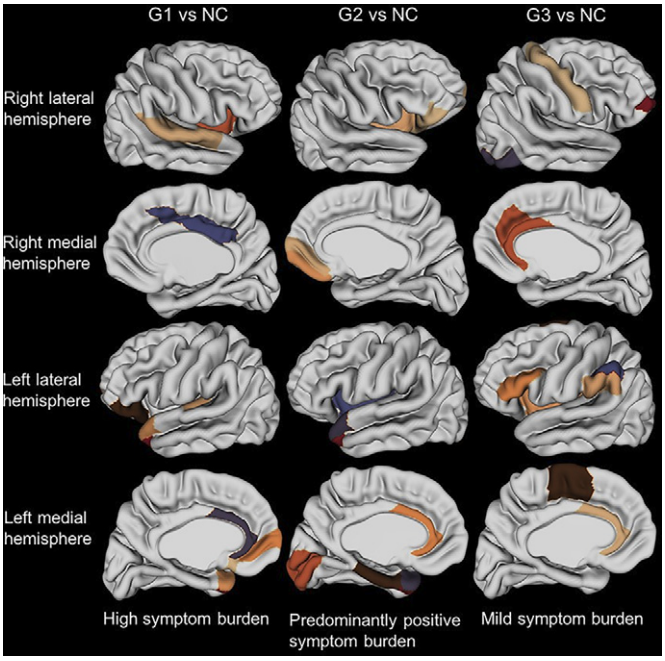

Fig. 3. The 10 most important features in performing the RF classification of the high symptom burden (G1 vs NC), predominantly positive symptom burden (G2 vs NC), and mild symptom burden (G3 vs NC) groups versus normal controls. Legend: NC = normal control, G1 = high symptom burden group, G2 = predominantly positive symptom burden group, G3 = mild symptom burden group.

cingulate and paracingulate gyri were of high importance for discriminating all groups against normal controls. The G1 and G2 classifications were additionally driven by the right insula, and the left temporal poles of the superior temporal and middle temporal gyri. Finally, G2 and G3 classification was commonly driven by the left insula. However, the regions unique to each subgroup classification provide better insight into their differentiability. In particular, for the high symptom burden group (G1 vs. NC) classification, the right superior temporal gyrus and the left Heschl's gyrus were important in classification. For the predominantly positive symptom burden group (G2 vs. NC), the orbital part of the right inferior frontal gyrus and the medial orbital part of the superior frontal gyrus were uniquely important regions for this classification. Finally, two of the important predictors of the mild symptom burden group (G3 vs. NC), were the right anterior cingulate and paracingulate gyri.

3.5. Generalizability of the proposed methods

Overall, only two of the three machine learning methods were able to predict the new group memberships properly after integrating the clustering in each of the 5 folds (Supplementary Table 1). RF performed extremely well across each of the folds and provided AUC values ranging from 0.67 to 0.94, in line with the results previously presented. Performance dips slightly for LR, where most folds demonstrate AUC between 0.65 and 0.91; however, in the classification of one of the clusters in two of the folds a drop in performance is observed (0.45 and 0.55). SVM, however demonstrates limited performance for this task, with AUC values ranging from 0.42 to 0.77 (although 0.94 was achieved in the classification of one cluster in a single fold). Taken together, these findings suggest that the methodology here may be generalizable with the correct combination of clustering technique and machine learning methodology.

4. Discussion

The goal of this manuscript was to investigate classification of data-driven clinical subgroups from neuroanatomical features. Using hierarchical clustering of SAPS/SANS individual items, three subgroups of clinical presentation were defined (high symptom burden, predominantly positive symptom burden, and low symptom burden). Cortical thickness measurements were used as inputs into three classifiers (LR, SVM, RF) to predict these clinical profiles. All three classifiers predicted the high symptom burden group with an AUC higher than the case-control study. Additionally, the RF classifier also outperformed the case-control study in predicting the mild symptom burden group. Our results suggest that these data-derived subgroups can be attributed to distinct structural alterations. Extracting the most predictive features in the RF classification suggested that different regions were more informative for the classification of each subgroup vs controls.

Data-driven methods for characterizing schizophrenia heterogeneity have previously been employed and are based on grouping either clinical or neuroanatomical attributes. Both Dollfus et al. and Carpenter et al. used hierarchical clustering to obtain schizophrenia subtypes (Dollfus et al., 1996). The present work most closely emulates the work by Dollfus et al. who also implemented hierarchical clustering using Ward's method on SAPS/SANS measures and identified four subtypes of presentation: mild symptom burden, high symptom burden (mixed), predominantly positive symptom burden, and predominantly negative symptom burden; three of which are similar to the subgroups presented here (Dollfus et al., 1996). However, a limitation of the study by Dollfus et al. is that they did not perform stability analysis, which would improve the confidence of the cluster solutions. More recently, Clementz et al. took a different approach and subgrouped patients based on an extensive biomarker panel, introducing psychosis biotypes (biologically distinctive phenotypes), which has since led many research groups to investigate the feasibility of clustering study participants based on structural and functional biomarkers for the goal of providing a neurobiological-based subtyping (Yang et al., 2014; Clementz et al., 2016; Dwyer et al., 2018; Brodersen et al., 2014). The biotypes concept has been invaluable to deepening our understanding of schizophrenia. However, a limitation of biotypes from the clinical perspective may be the lack of accepted biomarker panel validated for use in schizophrenia research. Further, the types of biomarker data acquired by the Bipolar and Schizophrenia Network for Intermediate Phenotypes (BSNIP) may be lengthy and costly to acquire (Tamminga et al., 2014). Also, treatment options are still typically linked to clinical performance and treatment tolerance. As such, clinical assessments remain the easiest and most cost-effective ways of providing a diagnosis and studying the clinical heterogeneity of schizophrenia.

The choice of the number of clusters included in a specific analysis using data driven techniques is often left to the discretion of the researcher and is often considered a design choice. From this perspective, the decision to use three clusters rather than 2 is an important design decision. Firstly, three subtypes have been defined in the past using other data driven techniques (Liddle, 1987; Malla et al., 1993; Nenadic et al., 2015). Further the symptom dimensions that we describe in the different clusters are indeed consistent with previous studies (Dollfus et al., 1996). From the perspective of design choices there are several considerations that one must give towards specific methodological choices made, particularly when using a data-driven approach. We have had success employing hierarchical clustering as a method for elucidating subtypes in the past (Bhagwat et al., 2018) using a method similar to the one used here (albeit using longitudinal data). Others have criticized this design decision due to concerns over the imposed tree structure, sensitivity to the similarity metric used to assess similarity, and subjective evaluation required to define clusters (Bouveyron and Brunet-Saumard, 2014). While we did employ hierarchical clustering, we also computed a measure of stability using permutations. However, when comparing the stability results associated with different permutation, we further incorporated the information associated with performing multiple random selections of subsets of the data. It should also be noted that hierarchical clustering provides an easily interpretable finding that consistently gets used in the literature (Ellegood

et al., 2015; Felice Reddy et al., 2014; Velthorst et al., 2019). Further examination of clustering methodology and its impact on subtype definition is required.

Zarogianni et al., 2013 and Arbabshirani et al., 2017 both reviewed the use of MRI-derived features in predicting a schizophrenia diagnosis, summarizing numerous studies in which measures of brain morphometry have been shown to be discriminative (Arbabshirani et al., 2017; Davatzikos et al., 2005). Of which, Davatzikos et al. was the first to show the capability of an SVM classifier in assigning a subject-specific diagnosis based on tissue density maps and achieving 81% accuracy (Davatzikos et al., 2005). Kambeitz et al. performed a meta-analytic study of 38 papers and showed that structural features were able to differentiate schizophrenia patients from normal controls with sensitivity and specificity of 76.4% and 79% respectively (Kambeitz et al., 2015). Unfortunately, these and other schizophrenia diagnosis classification results presented in the literature are not yet at a level that would be acceptable for clinical use; thus, Dwyer et al. investigated whether patient subtyping based on gray matter volumes, as derived from MRI, would increase the predictive power of classification against normal controls (Dwyer et al., 2018). They showed that classification accuracy using clinical variable inputs in predicting the two subgroups, was higher than predicting the entire schizophrenia patient population versus normal controls (accuracy: 73% and 78.8% versus 68.3%) (Dwyer et al., 2018). Based on a similar hypothesis, in the present study, cortical thickness features were used as predictors of our defined subgroups, to test whether clinical heterogeneity within the schizophrenia population is presenting as different structural alterations and hindering the classifier from learning the patterns of pathological presentation, thus saturating the case-control performance. Our results show that RF performed better at predicting the group membership of the high and mild symptom burden groups, as compared to the baseline classification task (AUC: 0.81 and 0.78 vs. 0.75).

Many of the features extracted by our RF-feature importance analysis were in line with previous findings of structural impairments and have been linked to the clinical presentation of schizophrenia. For each of the subgroup versus normal control classifications the insula was identified as an important feature in classification (high symptom burden: right insula, predominantly positive symptom burden: bilaterally, mild symptom burden: left insula). This region is implicated in emotional and sensory processing, and impairments have been hypothesized to be associated with negative symptoms and hallucinations (Wylie and Tregellas, 2010). Numerous studies have shown volume reduction in the insula (Glahn et al., 2008; Bora et al., 2011), in addition to cortical thickness differences in the left (Kuperberg et al., 2003), right (Nesvåg et al., 2008), and bilaterally (Bora et al., 2011; Allen et al., 2008) in schizophrenia patients. The results presented here indicate that the right, bilateral and left insula are important predictors of high symptom burden, predominantly positive symptom burden, and mild symptom burden subgroups respectively. This may reflect the importance of patient subtyping based on clinical measures, and the assessment of the insular effects of each subtype. The left anterior cingulate and paracingulate gyrus were important features in the classification of all the three subgroups. These regions are associated with executive function in healthy individuals (Carter et al., 1999), with leftward asymmetry linked specifically to executive tasks in males (Fornito et al., 2004). Volume reductions of the anterior cingulate have been noted in schizophrenia patients (Glahn et al., 2008; Bora et al., 2011). Since each subgroup has a higher proportion of males, it is expected that this region would appear as a discriminant feature and be lateralized to the left hemisphere. The right superior temporal gyrus and left Heschl's gyrus were two of the most important features in the RF classification of the high symptom burden group. Reductions of both the volume of the superior temporal gyrus (Shenton et al., 2001; Sun et al., 2009), and the volume of Heschl's gyrus (Hirayasu et al., 2000) have been noted in schizophrenia patients. These two regions contain the primary auditory cortex, and impairments have been associated with auditory hallucinations in schizophrenia (Bora et al., 2011; Allen et al., 2008; Honea et al., 2005). Walton et al. found a negative relationship between positive symptom severity and bilateral thickness in the superior temporal gyrus in the ENIGMA Schizophrenia Working Group consortium dataset containing nearly 2000 schizophrenia patients (Walton et al., 2017). Similarly, in a study by Sumich et al. (2005), reductions in the left Heschl's gyrus were associated with hallucinations and delusions.

A number of limitations should be noted in this study. The results presented in this manuscript were based on a relatively small sample set, consisting of a total of 167 study participants, patients were divided into three subgroups, and study participants were further split into 5 partitions for cross-validation. Recently numerous studies have criticized the use of small sample sizes for single-subject prediction of brain disorders tasks (Arbabshirani et al., 2017), however, our study is limited by data availability, the decision to prioritize data quality, and practicing proper machine learning protocols. To assess the generalizability of the method presented in this manuscript it is important to test it on an external validation dataset. Only one dataset contained in Schizconnect (http://schizconnect.org/) was used in this analysis; however, data from multiple other consortia are available, a subset of which provide SAPS/SANS clinical patient information. The functional biomedical information research network (fBIRN) (Keator et al., 2016) and the Neuromorphometry by Computer Algorithm Chicago (NMorphCH) (http://nunda.northwestern.edu/nunda/data/projects/NMorphCH) datasets contained in SchizConnect both have T1-weighted MRIs and SAPS/SANS data. However, each site has different scanning protocols and clinical measure raters. First attempts to add the study participants from the fBIRN and NMorph datasets to the clustering analysis presented in this manuscript also produced three clusters, however these separated patients into their respective datasets. Future work includes investigating methods for integrating multiple sites into this analysis to increase the sample size. However, we did attempt to address this issue by integrating the clustering within the 5-fold cross-validation performed in our study. As suggested by others, the nature in which the clustering and training and testing is performed may have a significant impact on the results obtained and may limit the generalizability of the final model. We observe that two of our machine learning models perform well in this configuration but with limited generalizability for SVM-based classifications. This recalls the no free lunch theorem which states that, on average, no machine learning algorithm will perform better than others over a broad class of problems (Davatzikos, 2018). However, in the present context, it is likely that the distributions being examined are most easily separated using simple linear models as we show here. In fact, experts in the field have called for examining the problem over a wide range of machine learning classes (as we have done here and previously, Bhagwat et al., 2018; Winterburn et al., 2017) rather than a single one (as done in Dinga et al., 2019) to determine which models may be best for a specific problem. Another limitation of this study is that binary classification of each subgroup was performed against normal controls. The results of this type of classification are highly dependent on the neuroanatomical data of the cohort of controls. Future analysis should include validating these results with a different cohort of study participants. Additionally, a multi-class classifier, capable of predicting whether an individual belongs to one of the subgroups or the class of normal controls, would have more clinical utility; however, previous studies have also employed the strategy used here since subgroups may have overlapping patterns of alterations (Gould et al., 2014). Poor performance was achieved when multi-class and subgroup versus subgroup classifiers were implemented. It is likely that larger sample sizes are required to properly perform classification at the single subject-level. Further, we had attempted to integrate MRI data from other datasets to determine if we could augment the sample size using multi-site data and attempted to generalize the findings across samples. However, we found that the clustering methodology for defining the initial subtypes was more sensitive to site groupings

rather than clinical groupings. While there are many initiatives underway to deal with site differences in univariate analyses related to schizophrenia (van Erp et al., 2016; Thompson et al., 2014), there are limited methods available at the moment to accommodate this in small-to-moderate samples sizes. This is likely due to differences in severity related to inclusion/exclusion criteria or the nature of the population under study as well (e.g.: community dwelling vs. institutionalized). As data-driven and machine learning approaches increase in popularity, these issues related to site differences will require resolution. From the perspective of experimental design, other measures outside of cortical thickness, such as surface area or subcortical volume, may have increased classifier accuracy or provided a more complete picture of the effects of the pathology. Finally, it is important to note that there is a demonstrated effect of antipsychotics on brain volumes and typical antipsychotics have been associated with greater cortical thinning than atypical antipsychotics (Ho et al., 2011; van Haren et al., 2011; Ansell et al., 2015). However, this was not studied here, and thereby no conclusions can be drawn concerning the effects on our results.

Overall, this study demonstrates that significant heterogeneity exists in the clinical presentation of the schizophrenia population, and techniques such as the one proposed are a step towards providing single-subject diagnosis based on objective measures of disease. A future direction would be to predict symptom severity scores directly from neuroanatomical variables for the future goal of detecting subjects that are at risk for developing the disorder.

Contributors

AT developed the study design, performed the data processing and analyses, and wrote the manuscript. NB assisted with the implementation and troubleshooting of the machine learning analyses and editing the manuscript. GAD helped with the image processing and providing feedback to the manuscript. ML aided in interpreting the results and writing the manuscript. MMC conceived the study, provided supervision and guidance, and edited the manuscript. All authors contributed to and have approved the final manuscript.

Role of funding source

MMC is funded by Canadian Institutes of Health Research, National Sciences and Engineering Research Council of Canada, and les Fonds de Recherches Sante Quebec. The funding sources did not play a role in the production of this manuscript.

Declaration of Competing Interest

None.

Acknowledgements

MMC is funded by Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Fonds de Recherche du Québec - Santé.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2019.05.044.

References

Abel, Kathryn M., Drake, Richard, Goldstein, Jill M., 2010. Sex differences in schizophrenia. Int. Rev. Psychiatry 22 (5), 417–428.

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J.-S., Mok, K., Ivanov, O., Vincent, R.D., et al., 2006. "A fully automated comprehensive pipeline for anatomical neuroimaging research." Presented at the 12th Annual Meeting of the Organization for Human Brain Mapping, Florence, Italy. http://www.bic.mni.mcgill.ca/users/yaddab/Yasser-HBM2006-Poster.pdf.

Allen, Paul, Larøi, Frank, McGuire, Philip K., Aleman, Andrè, 2008. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. Neurosci. Biobehav. Rev. 32 (1), 175–191.

Ansell, B.R.E., Dwyer, D.B., Wood, S.J., Bora, E., Brewer, W.J., Proffitt, T.M., Velakoulis, D., McGorry, P.D., Pantelis, C., 2015. Divergent effects of first-generation and second-generation antipsychotics on cortical thickness in first-episode psychosis. Psychol. Med. 45 (3), 515–527.

Arbabshirani, Mohammad R., Plis, Sergey, Sui, Jing, Calhoun, Vince D., 2017. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. NeuroImage 145 (Pt B), 137–165.

Ben-Hur, Asa, Elisseeff, Andre, Guyon, Isabelle, 2002. A stability based method for discovering structure in clustered data. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, pp. 6–17.

Bhagwat, Nikhil, Viviano, Joseph D., Voineskos, Aristotle N., Mallar Chakravarty, M., Alzheimer's Disease Neuroimaging Initiative, 2018. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. PLoS Comput. Biol. 14 (9), e1006376.

Bora, Emre, Fornito, Alex, Radua, Joaquim, Walterfang, Mark, Seal, Marc, Wood, Stephen J., Yücel, Murat, Velakoulis, Dennis, Pantelis, Christos, 2011. Neuroanatomical abnormalities in schizophrenia: a multimodal voxelwise meta-analysis and meta-regression analysis. Schizophr. Res. 127 (1–3), 46–57.

Bouveyron, Charles, Brunet-Saumard, Camille, March 2014. Model-based clustering of high-dimensional data: A review. Computational Statistics & Data Analysis 71, 52–78. https://doi.org/10.1016/j.csda.2012.12.008.

Brodersen, Kay H., Deserno, Lorenz, Schlagenhauf, Florian, Lin, Zhihao, Penny, Will D., Buhmann, Joachim M., Stephan, Klaas E., 2014. Dissecting psychiatric spectrum disorders by generative embedding. NeuroImage 4, 98–111.

Brown, R., Colter, N., Corsellis, J.A., Crow, T.J., Frith, C.D., Jagoe, R., Johnstone, E.C., Marsh, L., 1986. Postmortem evidence of structural brain changes in schizophrenia. Differences in brain weight, temporal horn area, and parahippocampal gyrus compared with affective disorder. Arch. Gen. Psychiatry 43 (1), 36–42.

Carpenter Jr., W.T., Bartko, J.J., Carpenter, C.L., Strauss, J.S., 1976. Another view of schizophrenia subtypes. A report from the international pilot study of schizophrenia. Arch. Gen. Psychiatry 33 (4), 508–516.

Carter, C.S., Botvinick, M.M., Cohen, J.D., 1999. The contribution of the anterior cingulate cortex to executive processes in cognition. Rev. Neurosci. 10 (1), 49–57.

Clementz, Brett A., Sweeney, John A., Hamm, Jordan P., Ivleva, Elena I., Ethridge, Lauren E., Pearlson, Godfrey D., Keshavan, Matcheri S., Tamminga, Carol A., 2016. Identification of distinct psychosis biotypes using brain-based biomarkers. Am. J. Psychiatry 173 (4), 373–384.

Collins, D. Louis, Louis Collins, D., Neelin, Peter, Peters, Terrence M., Evans, Alan C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18 (2), 192–205.

Crow, T.J., Taylor, G.R., Tyrrell, D.A.J., 1986. Two syndromes in schizophrenia and the viral hypothesis. Progress in Brain Research, pp. 17–27.

Davatzikos, Christos, 2018. Machine learning in neuroimaging: progress and challenges. NeuroImage (October) https://doi.org/10.1016/j.neuroimage.2018.10.003.

Davatzikos, Christos, Shen, Dinggang, Gur, Ruben C., Wu, Xiaoying, Liu, Dengfeng, Fan, Yong, Hughett, Paul, Turetsky, Bruce I., Gur, Raquel E., 2005. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry 62 (11), 1218–1227.

Dinga, Richard, Schmaal, Lianne, Penninx, Brenda, Tol, Marie Jose van, Veltman, Dick, Velzen, Laura van, Wee, Nic van der, Marquand, Andre, 2018. Evaluating the Evidence for Biotypes of Depression: Attempted Replication of Drysdale et al. 2017. https://doi.org/10.1101/416321.

Dinga, Richard, Schmaal, Lianne, Penninx, Brenda W.J.H., van Tol, Marie Jose, Veltman, Dick J., van Velzen, Laura, Mennes, Maarten, et al., 2019. Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017), 2019. Neuroimage Clin. 22, 101796. https://doi.org/10.1016/j.nicl.2019.101796.

Dollfus, S., Everitt, B., Ribeyre, J.M., Assouly-Besse, F., Sharp, C., Petit, M., 1996. Identifying subtypes of schizophrenia by cluster analyses. Schizophr. Bull. 22 (3), 545–555.

Dwyer, Dominic B., Cabral, Carlos, Kambeitz-Ilankovic, Lana, Sanfelici, Rachele, Kambeitz, Joseph, Calhoun, Vince, Falkai, Peter, Pantelis, Christos, Meisenzahl, Eva, Koutsouleris, Nikolaos, 2018. Brain subtyping enhances the neuroanatomical discrimination of schizophrenia. Schizophr. Bull. February. https://doi.org/10.1093/schbul/sby008.

Ellegood, J., Anagnostou, E., Babineau, B.A., Crawley, J.N., Lin, L., Genestine, M., DiCicco-Bloom, E., et al., 2015 Feb. Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. Mol. Psychiatry 20 (1), 118–125. https://doi.org/10.1038/mp.2014.98 (Epub 2014 Sep 9).

Eskildsen, Simon F., Coupé, Pierrick, Fonov, Vladimir, Manjón, José V., Leung, Kelvin K., Guizard, Nicolas, Wassef, Shafik N., Østergaard, Lasse Riis, Collins, D. Louis, Alzheimer's Disease Neuroimaging Initiative, 2012. BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59 (3), 2362–2373.

Felice Reddy, L., Green, M.F., Rizzo, S., Sugar, C.A., Blanchard, J.J., Gur, R.E., Kring, A.M., Horan, W.P., 2014 Oct. Behavioral approach and avoidance in schizophrenia: an evaluation of motivational profiles. Schizophr. Res. 159 (1), 164–170. https://doi.org/10.1016/j.schres.2014.07.047 (Epub 2014 Aug 19).

Fornito, Alexander, Yücel, Murat, Wood, Stephen, Stuart, Geoffrey W., Buchanan, Jo-Anne, Proffitt, Tina, Anderson, Vicki, Velakoulis, Dennis, Pantelis, Christos, 2004. Individual differences in anterior cingulate/paracingulate morphology are related to executive functions in healthy males. Cereb. Cortex 14 (4), 424–431.

Garreta, Raul, Moncecchi, Guillermo, 2013. Learning Scikit-learn: Machine Learning in Python. Packt Publishing Ltd.

Glahn, David C., Laird, Angela R., Ellison-Wright, Ian, Thelen, Sarah M., Robinson, Jennifer L., Lancaster, Jack L., Bullmore, Edward, Fox, Peter T., 2008. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. Biol. Psychiatry 64 (9), 774–781.

Gould, Ian C., Shepherd, Alana M., Laurens, Kristin R., Cairns, Murray J., Carr, Vaughan J., Green, Melissa J., 2014. Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia: A support vector machine learning approach. Published online 2014 Sep 18. Neuroimage Clin. 6, 229–236. https://doi.org/10.1016/j.nicl.2014.09.009.

Grossman, Linda S., Harrow, Martin, Rosen, Cherise, Faull, Robert, Strauss, Gregory P., 2008. Sex differences in schizophrenia and other psychotic disorders: a 20-year longitudinal study of psychosis and recovery. Compr. Psychiatry 49 (6), 523–529.

Haren, Neeltje E.M. van, Schnack, Hugo G., Cahn, Wiepke, Heuvel, Martijn P. van den, Lepage, Claude, Collins, Louis, Evans, Alan C., Pol, Hilleke E. Hulshoff, Kahn, René S., 2011. Changes in cortical thickness during the course of illness in schizophrenia. Arch. Gen. Psychiatry 68 (9), 871–880.
Hirayasu, Y., McCarley, R.W., Salisbury, D.F., Tanaka, S., Kwon, J.S., Frumin, M., Snyderman, D., et al., 2000. Planum temporale and Heschl gyrus volume reduction in schizophrenia: a magnetic resonance imaging study of first-episode patients. Arch. Gen. Psychiatry 57 (7), 692–699.
Ho, Beng-Choon, Andreasen, Nancy C., Ziebell, Steven, Pierson, Ronald, Magnotta, Vincent, 2011. Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. Arch. Gen. Psychiatry 68 (2), 128–137.
Honea, Robyn, Crow, Tim J., Passingham, Dick, Mackay, Clare E., 2005. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. Am. J. Psychiatry 162 (12), 2233–2245.
Kabani, N., Le Goualher, G., MacDonald, D., Evans, A.C., 2001. Measurement of cortical thickness using an automated 3-D algorithm: a validation study. NeuroImage 13 (2), 375–380.
Kambeitz, Joseph, Kambeitz-Ilankovic, Lana, Leucht, Stefan, Wood, Stephen, Davatzikos, Christos, Malchow, Berend, Falkai, Peter, Koutsouleris, Nikolaos, 2015. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. Neuropsychopharmacology 40 (7), 1742–1751.
Keator, David B., Erp, Theo G.M. van, Turner, Jessica A., Glover, Gary H., Mueller, Bryon A., Liu, Thomas T., Voyvodic, James T., et al., 2016. The function biomedical informatics research network data repository. NeuroImage 124 (Pt B), 1074–1079.
Kim, June Sic, Singh, Vivek, Lee, Jun Ki, Lerch, Jason, Ad-Dab'bagh, Yasser, MacDonald, David, Lee, Jong Min, Kim, Sun I., Evans, Alan C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. NeuroImage 27 (1), 210–221.
Kolmogorov, A., 1992. On the empirical determination of a distribution function. Springer Series in Statistics, pp. 106–113.
Koutsouleris, Nikolaos, Gaser, Christian, Jäger, Markus, Bottlender, Ronald, Frodl, Thomas, Holzinger, Silvia, Schmitt, Gisela J.E., et al., 2008. Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. NeuroImage 39 (4), 1600–1612.
Kuperberg, Gina R., Broome, Matthew R., McGuire, Philip K., David, Anthony S., Eddy, Marianna, Ozawa, Fujiro, Goff, Donald, et al., 2003. Regionally localized thinning of the cerebral cortex in schizophrenia. Arch. Gen. Psychiatry 60 (9), 878–888.
Lemaitre, Herve, Goldman, Aaron L., Sambataro, Fabio, Verchinski, Beth A., Meyer-Lindenberg, Andreas, Weinberger, Daniel R., Mattay, Venkata S., 2012. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? Neurobiol. Aging 33 (3) 617.e1–9.
Lerch, Jason P., Evans, Alan C., 2005. Cortical thickness analysis examined through power analysis and a population simulation. NeuroImage 24 (1), 163–173.
Li, Ying, Li, Wen-Xiu, Xie, Dong-Jie, Yi, Wang, Cheung, Eric F.C., Chan, Raymond C.K., 2018. Grey matter reduction in the caudate nucleus in patients with persistent negative symptoms: an ALE meta-analysis. Schizophr. Res. 192 (February), 9–15.
Liddle, Peter F., 1987. The symptoms of chronic schizophrenia. Br. J. Psychiatry 151 (02), 145–151.
MacDonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage 12 (3), 340–356.
Makowski, C., Bodnar, M., Shenker, J.J., Malla, A.K., Joober, R., Chakravarty, M.M., Lepage, M., 2017. Linking persistent negative symptoms to amygdala-hippocampus structure in first-episode psychosis. Transl. Psychiatry 7 (8), e1195.
Malla, Ashok K., Norman, Ross M.G., Williamson, Peter, Cortese, Leonard, Diaz, Fernando, 1993. Three syndrome concept of schizophrenia: A factor analytic study. Schizophr. Res. 10 (2), 143–150. https://doi.org/10.1016/0920-9964(93)90049-O.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60.
McGrath, John, Saha, Sukanta, Chant, David, Welham, Joy, 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol. Rev. 30 (May), 67–76.
Mutlu, A. Kadir, Schneider, Maude, Debbané, Martin, Badoud, Deborah, Eliez, Stephan, Schaer, Marie, 2013. Sex differences in thickness, and folding developments throughout the cortex. NeuroImage 82 (November), 200–207.
Narr, Katherine L., Toga, Arthur W., Szeszko, Philip, Thompson, Paul M., Woods, Roger P., Robinson, Delbert, Sevy, Serge, Wang, Yungping, Schrock, Karen, Bilder, Robert M., 2005. Cortical thinning in cingulate and occipital cortices in first episode schizophrenia. Biol. Psychiatry 58 (1), 32–40.
Nenadic, Igor, Sauer, Heinrich, Gaser, Christian, 2010. Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. NeuroImage 49 (2), 1153–1160.
Nenadic, Igor, Yotter, Rachel A., Sauer, Heinrich, Gaser, Christian, 2015. Patterns of cortical thinning in different subgroups of schizophrenia. Br. J. Psychiatry J. Ment. Sci. 206 (6), 479–483.
Nesvåg, Ragnar, Lawyer, Glenn, Varnäs, Katarina, Fjell, Anders M., Walhovd, Kristine B., Frigessi, Arnoldo, Jönsson, Erik G., Agartz, Ingrid, 2008. Regional thinning of the cerebral cortex in schizophrenia: effects of diagnosis, age and antipsychotic medication. Schizophr. Res. 98 (1–3), 16–28.
Patel, Krishna R., Cherian, Jessica, Gohil, Kunj, Atkinson, Dylan, 2014. Schizophrenia: overview and treatment options. P T 39 (9), 638–645.
Ruigrok, Amber N.V., Salimi-Khorshidi, Gholamreza, Lai, Meng-Chuan, Baron-Cohen, Simon, Lombardo, Michael V., Tait, Roger J., Suckling, John, 2014. A meta-analysis of sex differences in human brain structure. Neurosci. Biobehav. Rev. 39 (February), 34–50.
Selemon, L.D., Zecevic, N., 2015. Schizophrenia: a tale of two critical periods for prefrontal cortical development. Transl. Psychiatry 5 (August), e623.
Shenton, Martha E., Dickey, Chandlee C., Frumin, Melissa, McCarley, Robert W., 2001. A review of MRI findings in schizophrenia. Schizophr. Res. 49 (1–2), 1–52.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87–97.
Smirnov, Nikolai V., 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bull. Math. Univ. Moscou 2 (2), 3–14.
Smith, Stephen M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
Struyf, Jan, Dobrin, Seth, Page, David, 2008. Combining gene expression, demographic and clinical data in modeling disease: a case study of bipolar disorder and schizophrenia. BMC Genomics 9 (1), 531.
Sumich, Alex, Chitnis, Xavier A., Fannon, Dominic G., O'Ceallaigh, Séamus, Doku, Victor C., Faldrowicz, Abi, Sharma, Tonmoy, 2005. Unreality symptoms and volumetric measures of Heschl's gyrus and Planum temporal in first-episode psychosis. Biol. Psychiatry 57 (8), 947–950.
Sun, Jinhua, Maller, Jerome J., Guo, Lanting, Fitzgerald, Paul B., 2009. Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. Brain Res. Rev. 61 (1), 14–32.
Tamminga, Carol A., Pearlson, Godfrey, Keshavan, Matcheri, Sweeney, John, Clementz, Brett, Thaker, Gunvant, 2014. Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. Schizophr. Bull. 40 (Suppl. 2), S131–S137 (March).
Tamnes, Christian K., Østby, Ylva, 2018. Morphometry and development: changes in brain structure from birth to adult age. Neuromethods, pp. 143–164.
Tandon, Rajiv, Nasrallah, Henry A., Keshavan, Matcheri S., 2009. Schizophrenia, 'just the facts' 4. Clinical features and conceptualization. Schizophr. Res. 110 (1–3), 1–23.
Thompson, Paul M., Stein, Jason L., Medland, Sarah E., Hibar, Derrek P., Vasquez, Alejandro Arias, Renteria, Miguel E., Toro, Roberto, et al., 2014. and the Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. 8 (2), 153–182. https://doi.org/10.1007/s11682-013-9269-5.
Tohka, Jussi, Zijdenbos, Alex, Evans, Alan, 2004. Fast and robust parameter estimation for statistical partial volume models in brain MRI. NeuroImage 23 (1), 84–97.
Tustison, Nicholas J., Avants, Brian B., Cook, Philip A., Zheng, Yuanjie, Egan, Alexander, Yushkevich, Paul A., Gee, James C., 2010. N4ITK: improved N3 Bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289.
van Erp, T.G.M., Hibar, D.P., Rasmussen, J.M., Glahn, D.C., Pearlson, G.D., Andreassen, O.A., Agartz, I., et al., 2016 Apr. for the ENIGMA Schizophrenia Working Group. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Published online 2015 Jun 2. Mol. Psychiatry 21 (4), 547–553. https://doi.org/10.1038/mp.2015.63.
Van Erp, Theo G.M., Walton, Esther, Hibar, Derrek P., Schmaal, Lianne, Jiang, Wenhao, Glahn, David C., Pearlson, Godfrey D., et al., 2018. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuro imaging genetics through meta analysis (ENIGMA) consortium. Biol. Psychiatry 84 (9), 644–654.
Velthorst, E., Meyer, E.C., Giuliano, A.J., Addington, J., Cadenhead, K.S., Cannon, T.D., Cornblatt, B.A., et al., 2019 Feb. Neurocognitive profiles in the prodrome to psychosis in NAPLS-1. Schizophr. Res. 204, 311–319. https://doi.org/10.1016/j.schres.2018.07.038 (Epub 2018 Aug 2).
Walton, E., Hibar, D.P., van Erp, T.G.M., Potkin, S.G., Roiz-Santiañez, R., Crespo-Facorro, B., Suarez-Pinilla, P., et al., 2017. Positive symptoms associate with cortical thinning in the superior temporal gyrus via the ENIGMA schizophrenia consortium. Acta Psychiatr. Scand. 135 (5), 439–447.
Walton, E., Hibar, D.P., van Erp, T.G.M., Potkin, S.G., Roiz-Santiañez, R., Crespo-Facorro, B., Suarez-Pinilla, P., et al., 2018. Prefrontal cortical thinning links to negative symptoms in schizophrenia via the ENIGMA consortium. Psychol. Med. 48 (1), 82–94.
Wang, Lei, Kogan, Alex, Cobia, Derin, Alpert, Kathryn, Kolasny, Anthony, Miller, Michael I., Marcus, Daniel, 2013. Northwestern University Schizophrenia Data and Software Tool (NUSDAST). Front. Neuroinform. 7 (November), 25.
Ward, Joe H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58 (301), 236.
Winterburn, Julie L., Voineskos, Aristotle N., Devenyi, Gabriel A., Plitman, Eric, Fuente-Sandoval, Camilo de la, Bhagwat, Nikhil, Graff-Guerrero, Ariel, Jo, Knight, Mallar Chakravarty, M., 2017. Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. Schizophr. Res. December. https://doi.org/10.1016/j.schres.2017.11.038.
Wylie, Korey P., Tregellas, Jason R., 2010. The role of the insula in schizophrenia. Schizophr. Res. 123 (2–3), 93–104.
Yang, Zhi, Xu, Yong, Xu, Ting, Hoy, Colin W., Handwerker, Daniel A., Chen, Gang, Northoff, Georg, Zuo, Xi-Nian, Bandettini, Peter A., 2014. Brain network informed subject community detection in early-onset schizophrenia. Sci. Rep. 4 (July), 5549.
Zarogianni, Eleni, Moorhead, Thomas W.J., Lawrie, Stephen M., 2013. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. NeuroImage 3, 279–289.
Zijdenbos, Alex P., Forghani, Reza, Evans, Alan C., 2002. Automatic 'pipeline' analysis of 3-D MRsI data for clinical trials: application to multiple sclerosis. IEEE Trans. Med. Imaging 21 (10), 1280–1291.