



OPEN

Machine learning classification of schizophrenia patients and healthy controls using diverse neuroanatomical markers and Ensemble methods

Geetha Soujanya Chilla¹✉, Ling Yun Yeow¹, Qian Hui Chew², Kang Sim^{2,3} & K. N. Bhanu Prakash^{1,3}✉

Schizophrenia is a major psychiatric disorder that imposes enormous clinical burden on patients and their caregivers. Determining classification biomarkers can complement clinical measures and improve understanding of the neural basis underlying schizophrenia. Using neuroanatomical features, several machine learning based investigations have attempted to classify schizophrenia from healthy controls but the range of neuroanatomical measures employed have been limited in range to date. In this study, we sought to classify schizophrenia and healthy control cohorts using a diverse set of neuroanatomical measures (cortical and subcortical volumes, cortical areas and thickness, cortical mean curvature) and adopted Ensemble methods for better performance. Additionally, we correlated such neuroanatomical features with Quality of Life (QoL) assessment scores within the schizophrenia cohort. With Ensemble methods and diverse neuroanatomical measures, we achieved classification accuracies ranging from 83 to 87%, sensitivities and specificities varying between 90–98% and 65–70% respectively. In addition to lower QoL scores within schizophrenia cohort, significant correlations were found between specific neuroanatomical measures and psychological health, social relationship subscale domains of QoL. Our results suggest the utility of inclusion of subcortical and cortical measures and Ensemble methods to achieve better classification performance and their potential impact of parsing out neurobiological correlates of quality of life in schizophrenia.

Psychotic spectrum disorders such as schizophrenia affect individuals in multiple domains including cognitive domains, interpersonal relationships and daily psychosocial functioning¹. Diagnosis of these disorders is carried out through detailed history taking, mental status examination, clinical examination and laboratory investigations whenever appropriate to rule out organic causes^{2,3}. Apart from the use of clinical rating scales to assess the severity of psychopathology, there are increasing efforts in the identification of genetic, biochemical, or imaging biomarkers^{4–7} that could aid in diagnosis, treatment and prognosis of illnesses and their subtypes. Elucidation of such underlying biomarkers could complement extant clinical measures and provide information regarding neural substrates underlying illness status.

Of note, neuroimaging studies have revealed abnormalities involving structural and functional cerebral changes within schizophrenia involving cortical (such as frontal region), subcortical regions (such as hippocampus, thalamus) and network connectivity alterations^{8–14}. Specifically, there is increasing interest in the employment of structural neuroimaging features to improve the diagnosis of schizophrenia using machine learning methods^{15–17}. Guo et al. employed features from amygdaloid and hippocampal subregions to differentiate between healthy controls and schizophrenia patients¹⁵. They carried out feature selection using sequential backward elimination and utilized Support Vector Machine Classifier (SVC)/(SVM) from which they reported an accuracy of 81.75% with sensitivity of 84.21%. In another study¹⁶, authors Yassin et al. carried out classification on a dataset consisting of 64 schizophrenia patients and 106 healthy controls using subcortical volumes and cortical thickness

¹Institute of Bioengineering and Bioimaging, Agency for Science, Technology and Research, Singapore, Singapore 138667. ²Institute of Mental Health, Singapore, Singapore 539747. ³These authors jointly supervised this work: Kang Sim and K. N. Bhanu Prakash. ✉email: venkatangs@ibb.a-star.edu.sg; bhanu@ibb.a-star.edu.sg

Summary of Quality of Life domain scores for HC and SZ												
	Overall		General health		Physical health		Psychological health		Social relationships		Environment	
	HC	SZ	HC	SZ	HC	SZ	HC	SZ	HC	SZ	HC	SZ
Mean ± SD	72 ± 15.36	58.50 ± 23.33	70 ± 16.95	58.66 ± 26.33	72.6 ± 10.45	60.91 ± 15.72	67.32 ± 11.88	55.77 ± 18.28	68.28 ± 16.24	53.18 ± 21.04	71.8 ± 10.62	59.24 ± 17.2
	p < 0.001† ***		p = 0.0027† **		p < 0.001† ***		p < 0.001† ***		p < 0.001† ***		p < 0.001† ***	
Median (IQR)	75 (75-75)	50 (50-75)	75 (50-75)	75 (50-75)	69 (69-81)	63 (50-69)	69 (59.5-75)	56 (44-69)	75 (56-75)	50 (44-69)	75 (63-81)	63 (50-69)
Range	25-100	0-100	25-100	0-100	44-94	19-100	31-88	6-100	19-100	0-100	38-94	13-100
Kurtosis (SE)	4.37 (0.55)	2.66 (0.39)	3.58 (0.55)	2.36 (0.39)	2.7 (0.55)	2.69 (0.39)	3.83 (0.55)	2.89 (0.39)	3.16 (0.55)	2.89 (0.39)	3.6 (0.55)	2.95 (0.39)
Skewness (SE)	-0.64 (0.28)	-0.33 (0.20)	-0.52 (0.28)	-0.46 (0.20)	-0.14 (0.28)	0.05 (0.20)	-0.81 (0.28)	0.11 (0.20)	-0.11 (0.28)	-0.21 (0.20)	-0.48 (0.28)	-0.06 (0.20)
Significance test of normality	p < 0.001† ***	p < 0.001† ***	p < 0.001† ***	p < 0.001† ***	p = 0.0021† **	p = 0.013† *	p < 0.001† ***	p = 0.002† **	p < 0.001† ***	p < 0.001† ***	p = 0.0067† **	p = 0.024† *
Cronbach's alpha	-	-	-	-	0.57	0.74	0.77	0.81	0.82	0.73	0.76	0.86

Figure 1. Summary of QoL items and domain scores prior to transformations—QoL scores are significantly higher in the healthy controls than patients with schizophrenia in all items. p-value (* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$) was calculated by [†]Mann–Whitney U-test, and ^aShapiro–Wilk's test.

features. Highest accuracies of 76.4% were achieved using subcortical volumes as features and a random forest classifier, 70.5% using cortical thickness as features and a decision tree analysis and 70.5% using both subcortical volumes and cortical thickness features and logistic regression as a classifier. Xiao et al. carried out classification on 163 first-episode drug-naïve schizophrenia patients and 163 healthy controls. Using cortical thickness and cortical surface area, they achieved accuracy and sensitivity in the range of 81–85% and 77–83% respectively¹⁷.

While there have been efforts to differentiate between patients with schizophrenia and healthy controls within subject cohorts, there are limited studies which employed a wider range of neuroanatomical measures for such classification. Even when more than one set of measures were used with machine learning algorithms, classification performance may not necessarily increase¹⁶. In this regard, Ensemble methods are multiple classifier systems where individual weak classifiers are combined to generate a more robust classification system. Outputs from multiple base learning algorithms are voted or stacked through an algorithm in training to generate an Ensemble classifier which can then classify new data. Based on extant data and possible benefit of Ensemble methods in strengthening classifiers, we hypothesize that employing a diverse set of neuroanatomical measures with Ensemble classification methods will improve classification performance. Hence, in a bid to improve on the accuracy and sensitivity of classification between schizophrenia and healthy controls, we employed a wider range of neuroanatomical features (cortical thickness, surface area, volume, mean curvature, subcortical volumes) with Ensemble methodology to improve overall performance of classification. We further correlated neuroimaging measures with quality of life measures to gain further insights into the relationship between neuroimaging measures and the functional status of our subjects.

Methods

Subject recruitment and study details. Patients with schizophrenia ($n = 158$) were recruited from Institute of Mental Health, Singapore. Confirmation of the diagnosis was made for all patients by psychiatrists based on information obtained from clinical history, existing medical records, interviews with significant others as well as administration of the Structured Clinical Interview for DSM-IV Disorders-Patient Version (SCID-P)¹⁸. There was no history of any significant neurological illness such as seizure disorder, head trauma or cerebrovascular accident for the patients. Healthy controls were recruited from the community by advertisements. Control subjects ($n = 76$) were free of any Axis I psychiatric disorder as determined by the SCID-Patient version (SCID-NP)¹⁹ and had no history of any major neurological, medical illnesses, substance abuse or psychotropic medication use. Written, informed consent was acquired from all the participants after a detailed explanation of the study procedures. The study protocol was approved by the Institutional Review Boards of both Institute of Mental Health and the National Neuroscience Institute, Singapore. All methods were performed in accordance with the relevant guidelines and regulations.

MR Imaging was carried out for patients and healthy controls on 3 T Philips Achieva scanner (Philips Medical Systems, Eindhoven, The Netherlands) using parallel imaging (SENSE). Axial T1 MPRAGE volumes were acquired with a matrix size of 256×256 and a resolution of $0.8984 \times 0.8984 \times 1 \text{ mm}^3$, with at least 180 slices covering the brain.

Quality of Life (QoL) for subjects was assessed using the World Health Organization Quality of Life assessment—Brief Form (WHOQOL-BREF)²⁰, which is a 26-item, 5-point self-rated questionnaire. It assesses subjective QoL in four domains, namely physical health (7-items), psychological (6-items), social relationships (3-items), and environment (8-items), with the 2 remaining items assessing overall perception of QoL and overall health satisfaction. After reverse-scoring for items 3, 4 and 26, raw scores within each domain were standardized to 0–100 range to obtain a domain score. A higher score indicates better subjective QoL. A summary of QoL for all domains and items in healthy controls and schizophrenia cohort is given in Fig. 1.

Image processing. To extract neuroanatomical measures, image data has been converted to NIfTI format using dcm2nii²¹ given the retrospective nature of the data. Subcortical segmentation and cortical surface reconstruction has been carried out using Freesurfer 6.0.0 software^{22–40}. In this reconstruction process, subcortical regions were segmented using Gaussian Classifier Atlas⁴¹ from which 55 subcortical features including sub-region volumes, white matter and non-white matter subcortical hypo-intensities were obtained. Cortical parcellation was carried out using Desikan–Killiany atlas⁴² through which 71 cortical volume features, 73 corti-

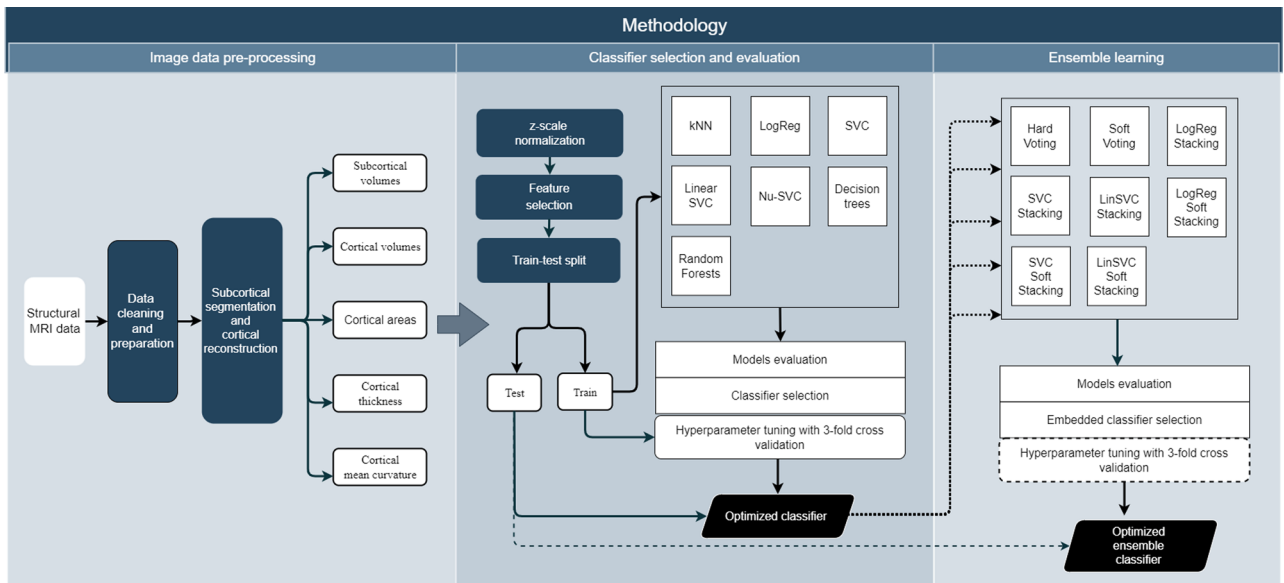


Figure 2. Overview of classification methodology.

cal surface area features, 71 cortical mean curvature features and 73 cortical thickness features were obtained. Reconstruction process on the entire dataset was performed using GNU parallel⁴³ on a high performance computing platform at National Supercomputing Centre (NSCC), Singapore.

In addition to visual inspection of images, quality assurance of data has been carried out through histogram based and quantile–quantile plots using generated neuroanatomical features to ensure that no significant abnormalities were included in the analyzed dataset.

Machine learning based classification. A total of 5 measures, namely cortical and subcortical volume, cortical surface area, cortical mean curvature and cortical thickness were obtained for classification of patients as mentioned in the previous section. In the first set of analyses, these measure sets were independently employed for classification and in the second set of analyses, all measures were merged or used in Ensemble for classification, as explained in detail in “[Study design](#)”. For all analyses, data was standardized, feature selection and train-test splitting were carried out before classifier selection and hyper-parameter tuning.

Standardization, feature selection and train-test split. All features were standardized to zero mean and unit variance using *StandardScaler* of Sklearn library. Feature selection was then carried out where most important features were selected using *SelectFromModel* of Sklearn library, using an SVC estimator. For training and test datasets generation, data was split in the ratio 70:30. Class proportions of 1:2 between healthy control cohort and schizophrenia cohort were maintained in training and test datasets and class balancing was employed with all estimators and classifiers wherever applicable. Final composition of the training and test dataset was 163 and 71 samples respectively.

Classifier selection and hyper-parameter tuning. To achieve best classification performance, initial classifier selection was done on its baseline performance and then was further optimized using best hyper-parameters. Multiple classifiers including k-Nearest Neighbors, Logistic regression, SVM classifiers (SVC with radial basis function kernel, Linear SVC, Nu-SVC), Decision trees, Random forests were tested for classifier selection process using F1 score and Area under Curve (AUC) as performance metrics. Once the base classifier was selected, tuning was carried out through exhaustive search over parameter space using *GridSearchCV*. A threefold cross-validation was done on training dataset during optimization and classification performance was evaluated on accuracy, AUC, F1 and recall scores. Classifier was then refitted on the training dataset with the parameters that resulted in best cross-validated AUC score. Final classification performance was evaluated using accuracy, sensitivity, specificity, F1 and AUC scores.

For Ensemble classification, input classifiers trained on measure subsets were fused using voting or stacking classifiers. Ensemble classification was done in a manner similar to that of initial base classifier, with initial classifier chosen from a total of 8 Ensemble classifiers. Baseline performance of hard voting, soft voting, hard stacking, and soft stacking, using three different estimators—logistic regression, SVC, Linear SVC and Nu-SVC was tested before choosing initial Ensemble classifier. If stacking classifiers were chosen for Ensemble, they were further tuned for best performance using *GridSearchCV* and threefold cross-validation. Overview of the methodology is seen in Fig. 2.

Study	Measures used	Total number of available features	Threshold value for feature selection	Number of selected features
1	Subcortical volumes (SV)	55	0.283	23
2	Cortical volumes (CV)	70	0.266	28
3	Cortical areas (CA)	72	0.294	26
4	Cortical thickness (CT)	72	0.389	36
5	Cortical mean curvature (CMC)	70	0.329	28
6	All measures (SV + CV + CA + CT + CMC)	331	0.074	139
7	Ensemble with 3 inputs (SV, CV, CA + CT + CMC)	(a) SV—55 (b) CV—70 (c) CA + CT + CMC—210	-	(a) SV—23 (b) CV—28 (c) CA + CT + CMC—85
8	Ensemble with 5 inputs (SV, CV, CA, CT, CMC)	(a) SV—55 (b) CV—70 (c) CA—72 (d) CT—72 (e) CMC—70	-	(a) SV—23 (b) CV—28 (c) CA—26 (d) CT—36 (e) CMC—28

Table 1. Overview of analyses performed—measures, number of available features and selected number of features used.

Correlation of QoL with neuroimaging features. To evaluate relationship between QoL and key neuroimaging features in schizophrenia cohort, we correlated ΔQoL with $\Delta\text{Features}$, where ΔQoL is difference between QoL for a patient to mean value for healthy control cohort ($\text{QoL} - \text{meanQoL}_{\text{HC}}$) and $\Delta\text{Features}$ is the difference between neuroimaging feature value for patient and mean value of the same feature for the healthy control cohort ($\text{Feature}_{\text{patient}} - \text{meanFeature}_{\text{HC}}$). This ΔQoL and $\Delta\text{Features}$ correlation testing was carried out using Spearman's rank-order correlation and was corrected for multiple comparisons using Bonferroni method.

Study design. A total of 8 analyses were carried out using Subcortical Volumes (SV), Cortical Volumes (CV), Cortical Areas (CA), Cortical Thickness (CT) and Cortical Mean Curvature (CMC) as feature sets for classification of schizophrenia and healthy controls. An overview of these analyses with feature subsets used, total number of available features and number of selected features is given in Table 1.

Classification using independent measures. The first set of analyses, analyses 1–5 of Table 1, were conducted using an independent feature set for their classification performance. Classifiers were trained on individual subcortical or cortical measures and optimized as described in “Machine learning based classification”. Additionally, for this set of analyses, correlation of selected features with QoL has been carried out, as given in “Correlation of QoL with neuroimaging features”.

Classification using all measures. In the second set of 3 analyses, analyses 6–8 of Table 1, all subcortical and cortical measures were used, either by merging feature sets or through Ensemble methods. For direct comparison with Ensemble methods, we carried out analysis 6 of Table 1, where all subcortical and cortical measures were merged, redundant variables were removed and feature selection and classification were carried out in a manner similar to independent measures classification. In analysis 7, we employed three different input classifiers for Ensemble classification, where input classifiers were trained on subcortical volumes, cortical volumes and remaining cortical measures respectively as shown in Table 1. In analysis 8, five different input classifiers trained on independent measure sets, from analyses 1–5, are used for Ensemble classification.

Results

In our dataset of 234 patients and healthy controls, training set consisted of 163 samples (53 HC, 110 SZ) and testing set consisted of 71 samples (23 HC, 48 SZ). Demographics of patients and controls in this training and test datasets, with group QoL item scores is given in Fig. 3.

Classification using independent measures and QoL correlation. When independent measure sets like SV, CA, CV etc. were used, classification accuracy and sensitivity were above 70% and F1 score greater than 0.70 was achieved. Specificity in general was lower between 55 and 65%. Of all independent measures tested, using cortical thickness measures resulted in higher classification accuracy, sensitivity and comparable specificity and F1 scores to other neuroanatomical measures. Among the classifiers, SVM based classifiers and Logistic regression classifiers gave the highest classification performance compared to other classifiers. Results of independent measures-based classification is shown in Table 2, with corresponding ROC curves shown in Fig. 4 from (a) to (e). Key neuroanatomical features selected are given in Supplementary Data A.1–A.5.

Figure 5 shows the Spearman's correlation between ΔQoL items and $\Delta\text{Features}$ from analyses 1–5, color-coded by Spearman's rank correlation coefficient value, ρ , and significance level highlighted in asterisks. A weak to moderate correlation was found between ΔQoL and $\Delta\text{Features}$, with significant correlation ($p < 0.05$) for several features ranging from a ρ of ± 0.2 to ± 0.4 . As shown in Fig. 5a, there was no significant correlation between any

	Training set		Testing set		Total	
	HC	SZ	HC	SZ	HC	SZ
<i>n</i>	163		71		234	
Age (Mean ± SD)	32.94 ± 9.41		32.82 ± 9.13		32.9 ± 9.31	
Gender (Male Female)	110 53		44 27		154 80	
Ethnicity (Chinese Malay Indian Others)	140 10 11 2		65 3 1 2		205 13 12 4	
Handedness (Right Left ambidextrous)	152 10 1		62 9 0		214 19 1	
<i>n</i>	53	110	23	48	76	158
Illness Duration in Years (Mean ± SD)	-	6.33 ± 7.74	-	6.88 ± 6.45	-	6.5 ± 7.36
Medication (CPZ eq. dosage)	-	221.45 ± 206.73	-	168.75 ± 112.34	-	205.44 ± 184.5
Overall QoL (Mean ± SD)	70.19 ± 15.69	58.18 ± 23.49	76.09 ± 14.06	59.24 ± 23.19	72 ± 15.36	58.50 ± 23.33
General health (Mean ± SD)	70.19 ± 15.69	58.88 ± 25.76	69.57 ± 19.88	58.15 ± 27.92	70 ± 16.95	58.67 ± 26.33
Physical health QoL (Mean ± SD)	72.79 ± 11.09	61.41 ± 15.96	72.17 ± 9.05	59.74 ± 15.26	72.6 ± 10.45	60.91 ± 15.72
Psychological health QoL (Mean ± SD)	67.13 ± 12.52	56.97 ± 18.23	67.74 ± 10.53	52.98 ± 18.27	67.32 ± 11.88	55.77 ± 18.28
Social relationships QoL (Mean ± SD)	69.15 ± 17.35	56.39 ± 21.56	66.3 ± 13.56	45.72 ± 17.88	68.28 ± 16.24	53.18 ± 21.04
Environment QoL (Mean ± SD)	72.48 ± 9.86	59.72 ± 16.66	70.26 ± 12.26	58.11 ± 18.53	71.8 ± 10.62	59.24 ± 17.2

Figure 3. Demographics of patients and controls in training and test datasets.

Study	Measures used	Classifier	Parameters tuned	Accuracy	Sensitivity	Specificity	F1	AUC
1	Subcortical volumes	Support Vector Classifier	Regularization parameter C, Max iterations	72%	79%	57%	0.79	0.68
2	Cortical volumes	Nu-Support Vector Classifier	Regularization parameter C, Max iterations, Kernel	73%	81%	57%	0.73	0.69
3	Cortical surface areas	Support Vector Classifier	Regularization parameter C, Max iterations	73%	77%	65%	0.74	0.72
4	Cortical thickness	Support Vector Classifier	Regularization parameter C, Max iterations	75%	81%	61%	0.75	0.71
5	Cortical mean curvature	Logistic Regression	Regularization parameter C, Max iterations, penalty, solver	70%	73%	65%	0.71	0.69

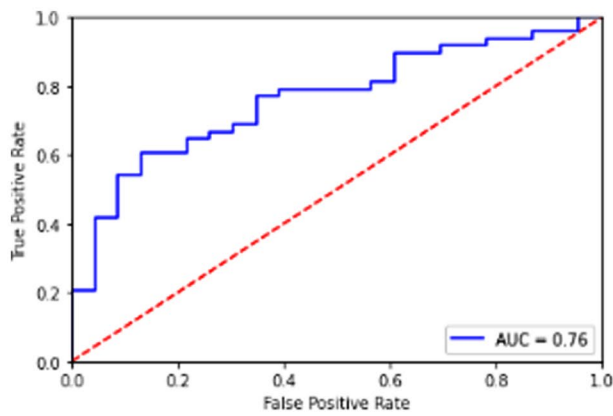
Table 2. Classification performance using independent measures.

of the subcortical volume features and Δ QoL. Several cortical volume features were identified to have significant correlations with Δ QoL, as shown in Fig. 5b. Negative correlations were found between left pars triangularis volume and right transverse temporal volume with overall QoL and left pars triangularis, left middle temporal and superior frontal volumes with social relationships domain. Figure 5c shows Δ QoL correlations with cortical surface features, where only the right rostral middle frontal surface area was negatively correlated with the social relationships domain. In cortical thickness measures shown in Fig. 5d, the left pars triangularis region negatively correlated with the psychological health domain of QoL while the right precentral region negatively correlated with overall QoL. Two positive correlations were found with cortical mean curvature features, between left fusiform mean curvature and psychological health domain and left parahippocampal curvature with overall QoL, as shown in Fig. 5e.

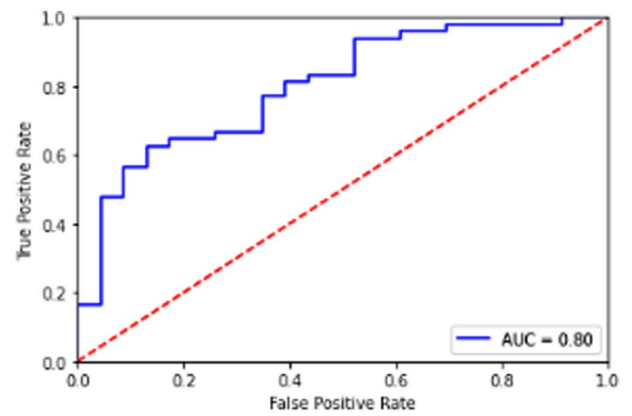
Classification using all measures. For analyses 6–8, where all neuroanatomical measures were used either by prior merging or through Ensemble classification, an increase in classification performance was observed. Compared to direct classification using merged features, Ensemble classification resulted in increased accuracy and sensitivity, although decrease in specificity was observed. Hard voting using three input classifiers—SVM trained on subcortical volumes, Nu-SVM trained on cortical volumes and logistic regression classifier trained on remaining cortical measures (areas, thickness and mean curvature) gave an accuracy of 87%, sensitivity of 90% and specificity of 70%. However, when Ensemble classification was carried out using five input classifiers from analyses 1–5 where each was trained on independent features sets, accuracy and sensitivity increased to 87% and 98% but specificity reduced to 65%. Results from classification for these analyses are given in Table 3 below with ROC plots for analyses 6 and 8 are given in Fig. 6.

Discussion and conclusion

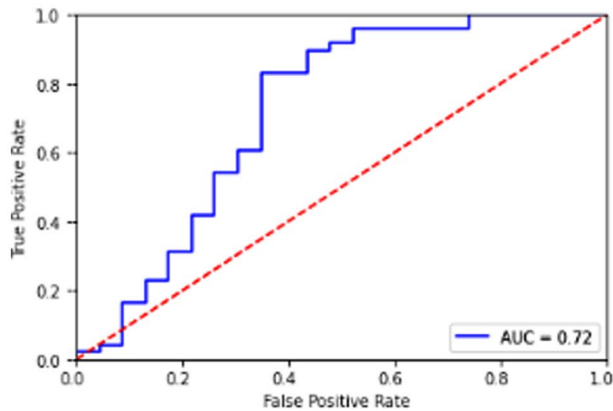
In this study, we investigated classification of subjects with schizophrenia and healthy controls using diverse neuroanatomical measures, including subcortical and cortical structure volumes, cortical surface areas, mean curvature and thickness of cortical structures. Some of these neuroanatomical measures have not been studied so far and hence their role and utility in classifying the two cohorts is unclear. Hence our work on classification using these independent feature sets, provides a baseline for future studies in this direction. From our results,



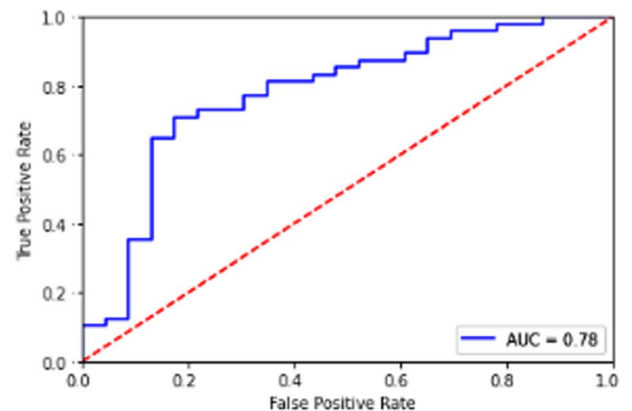
(a) Subcortical volumes - SVC



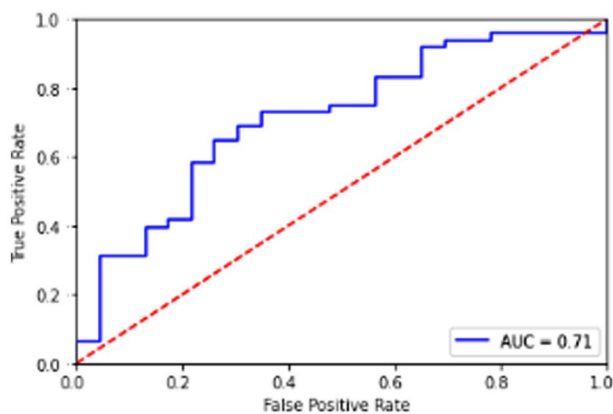
(b) Cortical volumes - NuSVC



(c) Cortical surface areas - SVC



(d) Cortical thickness - SVC



(e) Cortical mean curvature - LogReg

Figure 4. ROC plots for classification using independent feature sets.

classification performance was comparable between independent measure sets, with accuracy, sensitivity and specificity ranging from 70–73%, 73–81% and 57–61% respectively. Among all the measures, employing cortical thickness as the feature set resulted in slightly higher accuracy and sensitivity. In these single measure set based classification, SVM-based classifiers and logistic regression classifiers consistently gave better classification performances compared to other tested classifiers for measures, which was also reported by Yassin et al.¹⁶ Additionally, we also evaluated classification performance using Ensemble classification using all available measures. Employing a diverse set of measures however resulted in much improved accuracy, sensitivity and specificity, with ranges of 77–87%, 79–98% and 65–74% respectively. In Ensemble classification with 5 different input classifiers, one from each measure set, and 141 of all available 339 neuroanatomical features were employed. This resulted in highest accuracy and sensitivity of 87% and 98% respectively. These 141 features were further correlated with QoL scores of patients with schizophrenia which revealed a weak to moderate correlation. With overall QoL,

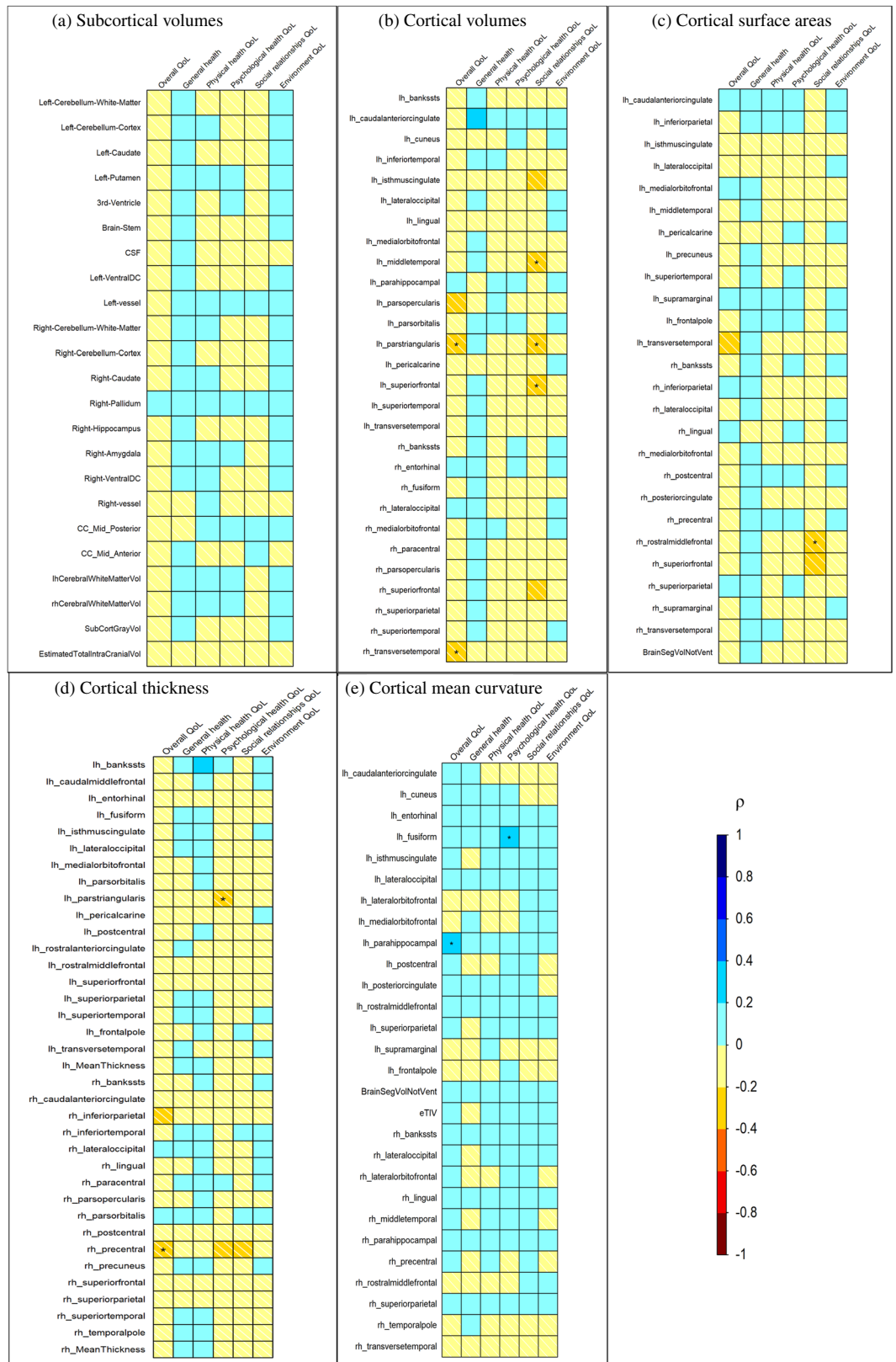


Figure 5. Spearman's rank-order correlation of the Δ QoL items and Δ Features, with feature sets from studies 1–5 i.e., (a) subcortical volumes, (b) cortical volumes, (c) cortical surface areas, (d) cortical thickness and (e) cortical mean curvature (*p-value < 0.05).

Study and description	Classifier(s)	Parameters tuned	Accuracy	Sensitivity	Specificity	F1	AUC
6—All subcortical and cortical measures	Logistic Regression	Regularization parameter C, Max iterations, penalty, solver	77%	79%	74%	0.83	0.77
7—Ensemble with 3 inputs (a) SV (b) CV (c) CA + CT + CMC	Ensemble—Hard Voting (a) Support Vector Classifier (b) Nu-Support Vector Classifier (c) Logistic Regression		83%	90%	70%	0.83	0.80
8—Ensemble with 5 inputs (a) SV (b) CV (c) CA (d) CT (e) CMC	Ensemble—Soft Stacking (a) Support Vector Classifier (b) Nu-Support Vector Classifier (c) Support Vector Classifier (d) Support Vector Classifier (e) Logistic Regression	Regularization parameter C, Max iterations, penalty, solver	87%	98%	65%	0.87	0.82

Table 3. Classification performance using all measures.

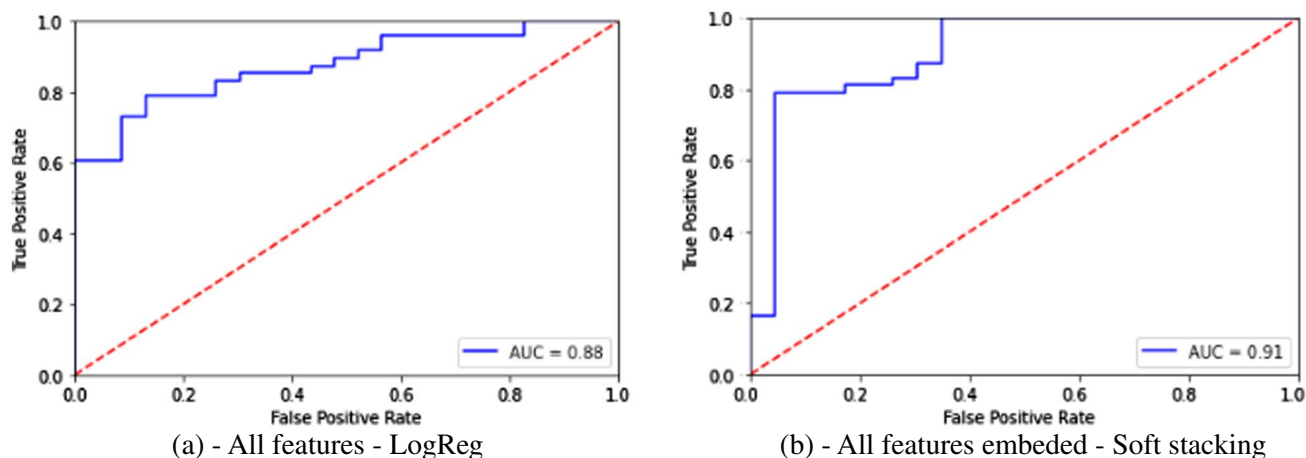


Figure 6. ROC plots for classification using all feature sets.

volumes of left pars triangularis and right transverse temporal regions, thickness of right precentral region were negatively correlated while mean curvature of left parahippocampal region was positively correlated. With the psychological health domain of QoL, thickness of left pars triangularis was negatively correlated and mean curvature of left fusiform was positively correlated. Volumes of left pars triangularis, left middle temporal and superior frontal regions, and surface area of right rostral middle frontal region negatively correlated with social relationship domain of QoL.

Compared to single measure based classification models from our own analyses as well as from literature, we observed that employing multiple measures increased classification performance. Specifically, using Ensemble methods resulted in much higher accuracy and sensitivity compared to direct classification from measures. In Ensemble classification, a new classifier is generated with inputs from various base classifiers. Such a learning process performed better than any single input classifier and allowed for increased classification performance, reducing bias and variance. We attributed improved performance of our classification to usage of Ensemble methods as well as utilization of multiple neuroanatomical measures for classification. Although sensitivity–specificity trade-off was observed as we increased the number of input classifiers, specificity achieved was still comparable to those obtained from single measure classification analyses within this study. Further studies in this direction on different datasets could employ hybrid and ensemble machine learning or deep learning methods which have been shown to improve classification performance⁴⁴.

In our study, we employed neuroanatomical features from both the left and right hemisphere separately for feature selection. Among these measures, certain regions have been identified from feature selection to be important in both the hemispheres. Pericalcarine region seemed to play a key role in Ensemble classification, with its surface area, cortical mean curvature and mean thickness measures selected and employed. This was followed by volume and thickness features of medial orbitofrontal and superior temporal regions, volume and surface area features of transverse temporal region and mean curvature and thickness features of rostral middle frontal regions. Among other features that contributed to classification, in both left and right hemispheres are volumes of pars opercularis, superior frontal, lateral occipital, banks of superior temporal sulcus, surface areas of isthmus of cingulate and supramarginal region, curvature of caudal anterior cingulate, precentral, superior parietal, parahippocampal, temporal pole regions and insula and thickness of pars orbitalis, inferior parietal and postcentral regions. Among subcortical measures, volumes of putamen in left hemisphere, amygdala, hippocampus and pallidum in right hemisphere, caudate, mid anterior and mid posterior regions of corpus callosum, cerebral white matter volume, ventral diencephalon and subcortical gray volume were identified as important features. Volumes of brain stem, cerebellum cortex, cerebellum white matter, brain segmentation, CSF, right and

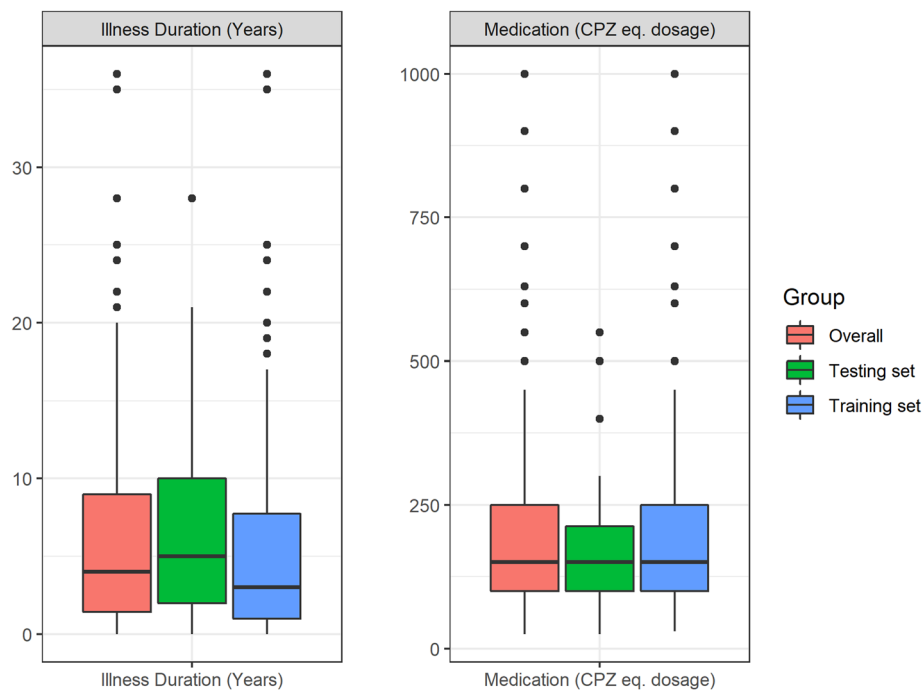


Figure 7. Illness duration (in years) and medication status (chlorpromazine equivalent dosage) distribution in patient dataset.

left vessel, 3rd Ventricle and total intracranial volume were also among key selected features for classification. However, it is important to note that our dataset consists of patients with varying illness and medication status, as shown in Fig. 7. When we further analysed the Ensemble results after excluding 13 subjects receiving > 500 CPZ eq mg/day within our modest sample, we found that whilst there is a mild decrease in accuracy (87 to 81%), sensitivity (98 to 89%) and AUC (0.82 to 0.77) for Ensemble with 5 inputs, the specificity and F1 remained the same. The overall pattern of gains in employing the Ensemble methods specifically Ensemble with 5 inputs remained. Additional studies using a larger and more normally distributed dataset can evaluate the utility and role of neuroanatomical markers at each stage of illness and treatment.

There have been very few studies which examined the correlations between neuroanatomical measures and quality of life assessments. For instance, our findings of negative correlation between social relationship subscale of QoL and cortical volumes (temporal, frontal regions) were consistent with those within a study by Ubukata et al.⁴⁵ Using voxel-based morphometry and Japanese version of the Schizophrenia Quality of Life Scale (JSQLS), they found that the psychosocial subscale QoL score is negatively correlated with gray matter volume in bilateral middle frontal gyrus, left midbrain, left postcentral gyrus, left inferior temporal gyrus, left inferior frontal gyrus, right middle occipital gyrus, and right cerebellum. Motivation/energy subscale QoL score was found to be negatively correlated with gray matter volume in the left superior frontal sulcus, left parahippocampal gyrus, left inferior temporal gyrus, right fusiform gyrus, right amygdala, right lingual gyrus, bilateral middle frontal gyrus, right superior temporal gyrus, right postcentral gyrus, and left middle temporal gyrus. Of note, the clinical factors subscale score was negatively correlated with GM volume in the left inferior frontal gyrus, left precentral gyrus, right middle frontal gyrus, left fusiform gyrus, and left inferior temporal gyrus. Another study⁴⁶ which carried out correlation of features with objective Quality of Life Scale (QLS) reported that instrumental role category score from the four subscales was correlated with the right anterior insula.

Several limitations are to be noted. First, we tested this classification system on a modest cross sectional data set. Second, further efforts to assess the utility of this classification system within a longitudinal dataset would allow better understanding and optimization of the classifiers over the time-course of illness. Third, we did not examine the use of the classifiers in differentiating subtypes of the illness by specific psychopathology, functional course or treatment response. Future studies applying and extending the current parameter-tuning may want to focus on parcellating heterogeneity of illness pertaining to specific symptomatology such as hallucinations, delusions, first rank symptoms using classification systems or even incorporating other clinical modalities such as cognitive functioning, treatment variables and functional factors within such classifiers in a larger dataset drawn from different and larger cohorts of subjects.

Received: 15 June 2021; Accepted: 3 February 2022

Published online: 17 February 2022

References

- Ruderfer, D. M. *et al.* Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715. e16 (2018).
- Arbabshirani, M. R., Plis, S., Sui, J. & Calhoun, V. D. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* **145**, 137–165 (2017).
- Kumari, S. *et al.* 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). *J. Addict. Res. Ther.* **08**, 1000324 (2017).
- Sklar, P. & Sklar, P. Genetics of schizophrenia and bipolar disorder. In *Neurobiology of Mental Illness* 232–246 (2013). <https://doi.org/10.1093/med/9780199934959.003.0018>.
- Leboyer, M. & Jamain, S. 31.4 Genetic, immunological and biochemical markers of treatment response in schizophrenia. *Schizophr. Bull.* **44**, S51–S51 (2018).
- Jimenez, A. M. Biomarkers for psychosis. *Gen. Methods Biomark. Res. Appl.* **2–2**, 979–1008 (2015).
- Jollans, L. & Whelan, R. Neuromarkers for mental disorders: Harnessing population neuroscience. *Front. Psychiatry* **9**, 242 (2018).
- van Erp, T. G. M. *et al.* 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**, 644–654 (2018).
- Ho, N. F. *et al.* Progression from selective to general involvement of hippocampal subfields in schizophrenia. *Mol. Psychiatry* **22**, 142–152 (2017).
- Sun, Y., Collinson, S. L., Suckling, J. & Sim, K. Dynamic reorganization of functional connectivity reveals abnormal temporal efficiency in schizophrenia. *Schizophr. Bull.* **45**, 659–669 (2019).
- Gheiratmand, M. *et al.* Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. *NPJ Schizophr.* **3**, 22 (2017).
- Plis, S. M. *et al.* Deep learning for neuroimaging: A validation study. *Front. Neurosci.* **8**, 229 (2014).
- Kim, J., Calhoun, V. D., Shim, E. & Lee, J. H. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage* **124**, 127–146 (2016).
- Dazzan, P. Neuroimaging biomarkers to predict treatment response in schizophrenia: The end of 30 years of solitude?. *Dialogues Clin. Neurosci.* **16**, 491–503 (2014).
- Guo, Y., Qiu, J. & Lu, W. Support vector machine-based schizophrenia classification using morphological information from amygdaloid and hippocampal subregions. *Brain Sci.* **10**, 1–14 (2020).
- Yassin, W. *et al.* Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. *Transl. Psychiatry* **10**, 1–11 (2020).
- Xiao, Y. *et al.* Support vector machine-based classification of first episode drug-naïve schizophrenia patients and healthy controls using structural MRI. *Schizophr. Res.* **214**, 11–17 (2019).
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W. *Structured Clinical Interview for DSM-IV Disorders-Patient Version (SCID-P)*. (American Psychiatric Press, 1994).
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W. *Structured Clinical Interview for DSM-IV Axis I Disorders-non-Patient Version*. (American Psychiatric Press, 2002).
- World Health Organisation. *WHO-BREF: Introduction, administration, scoring and generic version of the assessment*. http://www.who.int/mental_health/media/en/76.pdf (1996).
- Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* **264**, 47–56 (2016). Accessed March 15, 2021
- Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **17**, 87–97 (1998).
- Fischl, B., Sereno, M. I., Tootell, R. B. H. & Dale, A. M. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8**, 272–284 (1999).
- Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
- Fischl, B., Sereno, M. I. & Dale, A. Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).
- Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex. *Neuroimage* **9**, 11050–11055 (1999).
- Fischl, B., Liu, A. & Dale, A. M. Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* **20**, 70–80 (2001).
- Rosas, H. D. *et al.* Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* **58**, 695–701 (2002).
- Kuperberg, G. R. *et al.* Regionally localized thinning of the cerebral cortex in schizophrenia. *Arch. Gen. Psychiatry* **60**, 878–888 (2003).
- Fischl, B. *et al.* Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).
- Fischl, B. *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, S69–S84 (2004).
- Salat, D. H. *et al.* Thinning of the cerebral cortex in aging. *Cereb. Cortex* **14**, 721–730 (2004).
- Ségonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**, 1060–1075 (2004).
- Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
- Jovicich, J. *et al.* Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* **30**, 436–443 (2006).
- Han, X. *et al.* Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180–194 (2006).
- Ségonne, F., Pacheco, J. & Fischl, B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans. Med. Imaging* **26**, 518–529 (2007).
- Reuter, M., Rosas, H. D. & Fischl, B. Highly accurate inverse consistent registration: A robust approach. *Neuroimage* **53**, 1181–1196 (2010).
- Reuter, M. & Fischl, B. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage* **57**, 19–21 (2011).
- Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).
- Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
- Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11050–11055 (2000).
- Tange, O. GNU Parallel 20210122 ('Capitol Riots'). (2020) <https://doi.org/10.5281/zenodo.4454976>.
- Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A. T. & Alinejad-Rokny, H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J. Biomed. Inform.* **113**, 103627 (2021).

45. Ubukata, S. *et al.* Regional gray matter reduction correlates with subjective quality of life in schizophrenia. *J. Psychiatr. Res.* **47**, 548–554 (2013).
46. Uwatoko, T. *et al.* Insular gray matter volume and objective quality of life in schizophrenia. *PLoS One* **10**, e0142018 (2015).

Acknowledgements

Authors would like to thank Dr. Kuan Jin Lee, Ms. Zhang Jiayi, Mr. Renick Lee and Nick Wilson from Fujitsu for their inputs and help during the project. We also extend our thanks to the personnel at NSCC. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (<https://www.nscg.sg>).

Author contributions

The following authors have made substantial contributions to conception and design of this project theme (G.C.), acquisition of data (K.S., Q.H.C.), analysis and interpretation of the data (G.C., Y.L.Y., B.P., Q.H.C., K.S.), drafting of the manuscript (G.C., Y.L.Y.), revision of the paper for important intellectual content (B.P., Q.H.C., K.S.), and have given final approval of the version to be published and agreed to be accountable for all aspects of the work (G.C., Y.L.Y., Q.H.C., B.P., K.S.).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-06651-4>.

Correspondence and requests for materials should be addressed to G.S.C. or K.N.B.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022