Multisite schizophrenia classification by integrating structural magnetic resonance imaging data with polygenic risk score

Ke Hu [a,b], Meng Wang [a,b], Yong Liu [a,b,c], Hao Yan [d,e], Ming Song [a,b], Jun Chen [f], Yunchun Chen [g], Huaning Wang [g], Hua Guo [h], Ping Wan [h], Luxian Lv [i,j], Yongfeng Yang [i,j], Peng Li [d,e], Lin Lu [d,e], Jun Yan [d,e], Huiling Wang [f,k], Hongxing Zhang [i,j,l], Dai Zhang [d,e,m], Huawang Wu [n], Yuping Ning [n], Tianzi Jiang [a,b,c,o,p,*], Bing Liu [q,r,*]

[a] Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[c] CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[d] Peking University Sixth Hospital/Institute of Mental Health, Beijing, China
[e] Key Laboratory of Mental Health, Ministry of Health (Peking University), Beijing, China
[f] Department of Radiology, Renmin Hospital of Wuhan University, Wuhan, China
[g] Department of Psychiatry, Xijing Hospital, The Fourth Military Medical University, Xi'an, China
[h] Zhumadian Psychiatric Hospital, Zhumadian, China
[i] Department of Psychiatry, Henan Mental Hospital, The Second Affiliated Hospital of Xinxiang Medical University, Xinxiang, China
[j] Henan Key Lab of Biological Psychiatry, Xinxiang Medical University, Xinxiang, China
[k] Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan, China
[l] Department of Psychology, Xinxiang Medical University, Xinxiang, China
[m] Center for Life Sciences/PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China
[n] The Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital), Guangzhou, China
[o] Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[p] Queensland Brain Institute, University of Queensland, Brisbane, Australia
[q] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[r] Chinese Institute for Brain Research, Beijing, China

ARTICLE INFO

ABSTRACT

Previous brain structural magnetic resonance imaging studies reported that patients with schizophrenia have brain structural abnormalities, which have been used to discriminate schizophrenia patients from normal controls. However, most existing studies identified schizophrenia patients at a single site, and the genetic features closely associated with highly heritable schizophrenia were not considered. In this study, we performed standardized feature extraction on brain structural magnetic resonance images and on genetic data to separate schizophrenia patients from normal controls. A total of 1010 participants, 508 schizophrenia patients and 502 normal controls, were recruited from 8 independent sites across China. Classification experiments were carried out using different machine learning methods and input features. We tested a support vector machine, logistic regression, and an ensemble learning strategy using 3 feature sets of interest: (1) imaging features: gray matter volume, (2) genetic features: polygenic risk scores, and (3) a fusion of imaging features and genetic features. The performance was assessed by leave-one-site-out cross-validation. Finally, some important brain and genetic features were identified. We found that the models with both imaging and genetic features as input performed better than models with either alone. The average accuracy of the classification models with the best performance in the cross-validation was 71.6%. The genetic feature that measured the cumulative risk of the genetic variants most associated with schizophrenia contributed the most to the classification. Our work took the first step toward considering both structural brain alterations and genome-wide genetic factors in a large-scale

multisite schizophrenia classification. Our findings may provide insight into the underlying pathophysiology and risk mechanisms of schizophrenia.

1. Introduction

Schizophrenia (SZ) is a chronic, persistent, and extremely harmful mental illness, with some combination of major symptoms, including hallucinations, delusions, and disordered thinking, which begin in late adolescent or young adulthood. Since the current diagnosis of schizophrenia is mainly based on a patient's clinical symptoms and observed behaviors, objective quantitative biomarkers for the diagnosis of schizophrenia are urgently needed.

The development of in vivo neuroimaging technology provides promise for the objective diagnosis of schizophrenia. High-resolution structural magnetic resonance imaging (sMRI) is particularly promising because sMRI images are relatively stable across different scanners and different status. Convergent evidence based on sMRI studies has shown that in patients with schizophrenia comprehensive brain structural abnormalities exist; in particular, gray matter volume has been shown to decrease in the prefrontal cortex (Dickey et al., 2004; Zhou et al., 2005), temporal cortex (Sun et al., 2009; Yamasue et al., 2004), anterior cingulate cortex (Ellison-Wright and Bullmore, 2010; Ellison-Wright et al., 2008), and thalamus (Ellison-Wright and Bullmore, 2010; Ellison-Wright et al., 2008). Group-level statistical analyses of brain structural abnormalities in SZ patients have provided important insights into schizophrenia pathology and indicated that these brain structural features could potentially be used to classify schizophrenia patients (Zarogianni et al., 2013). Pattern recognition approaches in the area of artificial intelligence can automatically discriminate SZ patients from normal controls (NCs) at the individual level by learning meaningful patterns from the data (Kambeitz et al., 2015), and several machine learning (ML) studies have yielded promising results (Rozycki et al., 2018; Vieira et al., 2020; Xiao et al., 2019). However, most of the existing models were built on relatively small samples or data from a single site, which limited the ability to apply them in clinical situations. Models trained with small samples tend to overfit and yield unstable and inconsistent results (Nieuwenhuis et al., 2012; Schnack and Kahn, 2016), and models that only use data from a single site have poor generalizability (Pinaya et al., 2016; Schnack et al., 2014; Vieira et al., 2020). For example, Pinaya et al. reported that the same ML model that was able to distinguish between patients with SZ and NCs with an accuracy of 74% showed poor generalizability (56%) when applied to an additional independent cohort of individuals (Pinaya et al., 2016). A review (Schnack and Kahn, 2016) based on the observations from published sMRI-ML studies on SZ showed that when the classification models were applied to independent validation samples, accuracy can drop as much as 10–15%. While studies with small sample size (n) can reach 90% and higher accuracy, above $n/2 = 50$ the maximum accuracy achieved steadily drops to below 70% for $n/2 > 150$. Therefore, large-sample multisite structural neuroimaging studies for schizophrenia using a ML-based diagnostic model and leave-site-out validation are needed.

Previous evidence has shown that schizophrenia is highly heritable (Cardno and Gottesman, 2000; Kety et al., 1994), and it is generally recognized that multiple susceptibility genes interact with environmental factors to comprise the major etiology of schizophrenia (Haller et al., 2014). A large-scale genome-wide association study (GWAS) of schizophrenia (Consortium, 2014) analyzed the association of tens of millions of genome-wide genetic loci with schizophrenia and further confirmed that schizophrenia is a polygenic disease that is subject to the cumulative effects of common variations with very small effects. Polygenic risk scores (PGRSs), by integrating all common genetic variant effects into a single risk metric, can capture the polygenic nature of complex disorders and could be useful features for schizophrenia

identification by measuring the intrinsic genetic effects for each individual (Anderson et al., 2019). However, to the best of our knowledge, most existing studies on SZ classification and identification only focused on various brain imaging features, and genetics-related features were not considered. Although two previous studies combined brain imaging features and genetics-related features in SZ classification (Pettersson-Yeo et al., 2013; Yang et al., 2010), they both used small samples from a single site ($n = 19$ and $n = 40$, respectively), which could not guarantee the reliability and generalization of the model. In addition, they used a small number of SNPs ($n = 26$ and $n = 367$, respectively), and fewer genetic variations were considered. The development of GWAS has enabled research on psychiatric genomics to move toward statistical methods to integrate additive effects from across the entire genome into a single metric (Anderson et al., 2019). Therefore, our goal was to take the first step toward considering both structural brain alterations and genome-wide genetic factors in a large-scale multisite schizophrenia classification. We hypothesize that integrating neuroimaging data with genetic features may improve the performance of schizophrenia classification. To test this hypothesis, a large-sample schizophrenia classification study that utilized multisite sMRI data and PGRS features was needed.

In this study, we classified schizophrenia by combining sMRI and genetic data from 1010 participants from eight independent sites across China. The overall workflow is shown in Fig. 1. The raw sMRI and GWAS data were collected using the same protocol for each participant. A recent study (Vieira et al., 2020) showed that GM volumes (GMV) features tended to yield higher accuracies than surface-based features in traditional machine learning and that, compared with voxel-based cortical thickness features, the accuracies of GMV features showed smaller differences between logistic regression (LR) and support vector machine (SVM) models. Therefore, we chose the more stable GMV features as the primary brain imaging metric in our study. For simplicity and biological interpretability, the region of interest (ROI)-based GMV of 246 brain regions were extracted for each individual based on the Brainnetome atlas (http://atlas.brainnetome.org/) after sMRI preprocessing. Meanwhile, step-wise PGRSs were calculated for each individual after performing quality control and preprocessing of the GWAS data. To better test the comparability and reproducibility of our study, we used three common traditional machine learning methods and easily acquired features. SVM, LR, and ensemble learning methods were evaluated in this classification task. Leave-one-site-out CV was used to test the performance of the models. Finally, some important brain and genetic features were identified.

2. Methods

All the code used in this study is publicly available at https://github.com/BingLiu-Lab/SZ-classification. To protect participant privacy, which the genetic data used in our study may involve, the data is currently not available for public download. For specific research needs, the corresponding author can be contacted to discuss data sharing.

2.1. Subjects

The participants were recruited using the same protocol at each of eight sites: Peking University Sixth Hospital (PKUH6), Beijing Huilongguan Hospital (HLG), Henan Mental Hospital GE scanning site (HMG), Henan Mental Hospital Siemens scanning site (HMS), Xijing Hospital (XJ), Guangzhou Brain Hospital (GB), Renmin Hospital of Wuhan University (RWU), and Zhumadian Psychiatric Hospital (ZMD). MRI data acquired from Henan Mental Hospital were separated into two

sites because two different MR scanners were used.

All SZ patients had a consistent diagnosis of schizophrenia confirmed by experienced psychiatrists according to the Diagnosis and Statistic Manual of Mental Disorders, fourth edition (DSM-IV) criteria for schizophrenia or schizophreniform disorder using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Patient Edition (SCID-I/P). The exclusion criteria for the patients included a history of somatic or neurologic disorders, serious medical illness, substance dependence, pregnancy, electroconvulsive therapy within the last six months, or a diagnosis of any other axis I disorder. The Positive and Negative Syndrome Scale (PANSS) total test scores for all the patients with SZ were higher than or equal to 60, and the scores for at least 3 positive items were higher than 4. The NCs, who had no current or previous axis I psychiatric disorders, were recruited by advertisement from the local community near each site and were carefully screened using the SCID-I, non-patient edition. None of the NCs had any personal history of psychotic illness nor any family history of psychosis in their first-, second-, or third-degree relatives. The study at each site was approved by the local ethical review board. After a complete description of the study was provided to the participants, written informed consent was obtained for each patient. The patients' legal guardians supervised the consent process. For patients with insight deficits, written informed consent from their legal guardians was also obtained. All the participants were right-handed and had no contraindications to MRI scanning.

The recruited sample contained 690 SZ patients and 619 healthy subjects, for a total of 1,309 subjects, for all of whom we had brain scan images. Out of these subjects, 1231 also had genetic data. All the images were carefully reviewed by four examiners to exclude those with artifacts such as motion, ghosting, low signal-to-noise ratio, or insufficient gray/white matter contrast. After extensive quality checking of the brain imaging data, 34 subjects were excluded. After conducting a genotype quality control of the genetic data, 67 subjects were excluded, and the exclusion process is described in the following content. In the end, only

508 SZ patients and 502 NCs, for a total of 1010 subjects with both qualified brain images and genetic data were used in our study.

2.2. Data acquisition

2.2.1. sMRI data acquisition

All sMRI data were obtained from 3.0 T magnetic resonance (MR) scanners, including Siemens Trio Trim 3.0 T scanners at four sites (HLG, PKUH6, HMS, and XJ), GE Signa 3.0 T scanners at three sites (RWU, HMG, and ZMD), and a Philips Achieva 3.0 T scanner at one site (GB). Uniform scanning protocols for all eight sites were set up by an experienced expert to ensure equivalent and high-quality data acquisition. A T1-weighted brain volume (BRAVO) MRI sequence was performed using a protocol with a matrix size of 256×256, resolution of 1×1 mm^2, inversion time of 1100 ms, and slice thickness of 1 mm. A total of 192 sagittal slices were acquired on the Siemens scanners and 188 sagittal slices on the GE scanners and Philips scanner.

2.2.2. Genetic data acquisition and preprocessing

The detailed genetic data acquisition and preprocessing methods were described in our prior work (Liu et al., 2020), and in the current paper, a similar description was re-used to maintain the clarity and scientific integrity of the present study.

Anticoagulant venous blood samples were acquired from all participants in this study. We used the EZgene Blood gDNA Miniprep Kit to extract genomic DNA from the whole blood of each subject and obtained whole-genome genotype data based on Illumina Human OmniZhongHua-8 BeadChips. Then we performed the genotype quality control using PLINK version 1.0.7 on a Linux system (Purcell et al., 2007). First, we removed two individuals whose missing genotype rates were greater than 0.05. Then, we removed 19 individuals with gender discrepancies. We also used the pairwise identity by descent estimate to identify individuals who could possibly be related. For pairs of
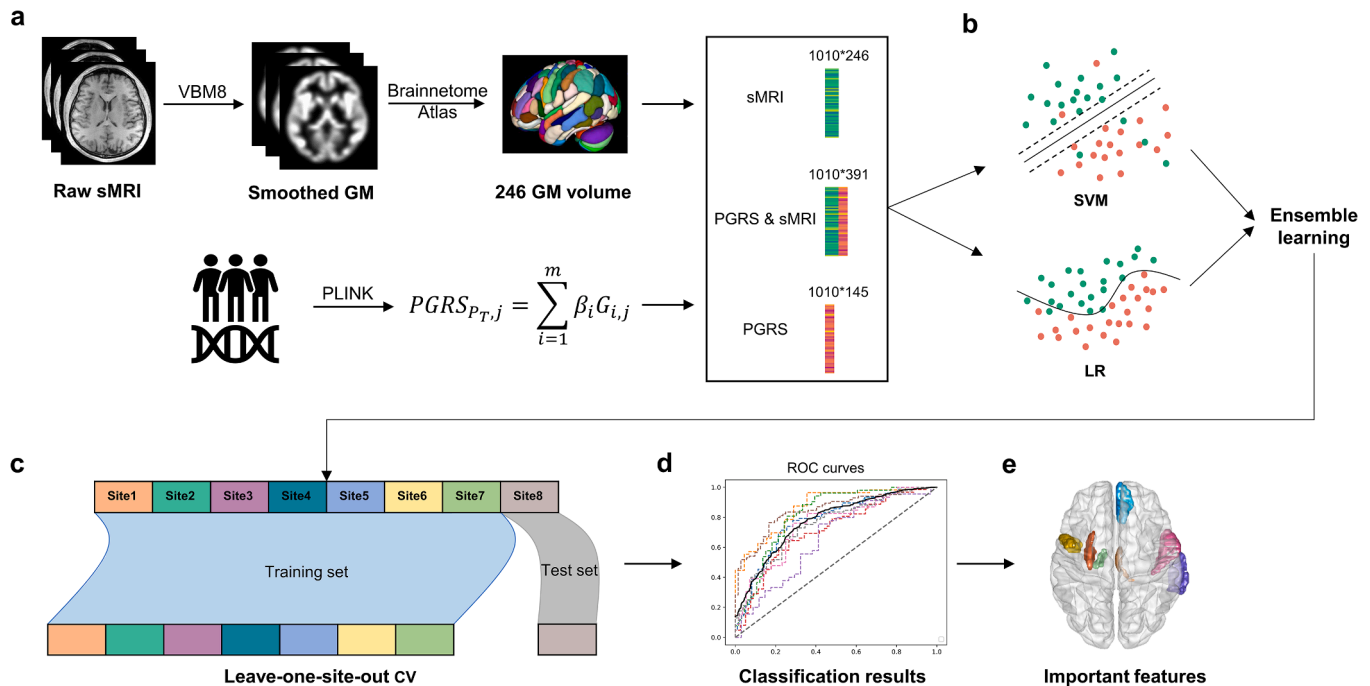

Fig. 1. Multisite schizophrenia classification with brain imaging and genetic data. a. Data preprocessing and feature extraction. As shown, different features were extracted and integrated: 1) Genetic features: PGRSs, 2) sMRI features: GM volume of 246 brain regions, 3) Fusion of PGRSs and GM volume features. b. The SVM, LR, and ensemble learning models predicting individual diagnostic status. c. The leave-one-site-out CV for assessing the accuracy of the models. Parameter optimization was performed on the training set and performance was evaluated on the test set. d. Model evaluation. The average classification performance of the eight models in the leave-one-site-out CV was used as the final result. e. Important features that contributed significantly to the classification were identified and ranked in order of importance. Abbreviations: CV: cross-validation; GM: gray matter; LR: logistic regression; PGRS: polygenic risk score; SVM: support vector machine; sMRI: structural magnetic resonance imaging.

individuals who had more similar genotypes than expected by chance in a random sample (based on pairwise identity by state value, PI_HAT in PLINK $\geq$ 0.25), we removed the one with the greater missing rate, and a total of 46 subjects were excluded. Next, we used single nucleotide polymorphism (SNP)-level filtering to remove SNPs with missing genotype rates greater than 0.05, a minor allele frequency<0.01, or a significant departure from Hardy–Weinberg equilibrium (P < .001). To avoid the confounding effects of population stratification (Walton et al., 2014), we performed a principal component analysis using EIGEN-STRAT 5.0.2 on the Linux system (Patterson et al., 2006; Price et al., 2006) on a linkage disequilibrium pruned set of autosomal SNPs. This set was obtained by carrying out linkage disequilibrium pruning with PLINK (r^2 < 0.05) and removing five long-range linkage disequilibrium regions with the HapMap phase 3 reference data-set (Thorisson et al., 2005). After obtaining 10 principal components, we excluded the outliers, that is, the samples greater than 6 s.d. Ungenotyped SNPs were imputed using SHAPEIT version 2 (r790) (Delaneau et al., 2012) and IMPUTE2 (Howie et al., 2009) on the Linux system with the 1000 Genomes Phase 1 reference data-set. Further analyses focused on autosomal SNPs with imputation quality scores greater than 0.8.

2.3. Feature extraction

2.3.1. Brain imaging features

T1-weighted images were processed using statistical parametric mapping (SPM8, https://www.fil.ion.ucl.ac.uk/spm/software/spm8/) and voxel-based morphometry toolbox version 8 (VBM8, http://dbm.neuro.uni-jena.de/vbm/). The sMRI images were segmented into gray matter (GM), white matter, and cerebrospinal fluid images. Next, the preprocessed GM images were smoothed with a 6 mm full width at half maximum isotropic Gaussian kernel. Finally, the GM images were mapped to 246 brain regions of the fine-grained Brainnetome atlas (Fan et al., 2016). The choice of brain parcellation was based on a recent study (Lee et al., 2021) that demonstrated the advantage of the Brainnetome atlas for this type of study. Then we calculated the average gray matter volume of each brain region for each subject. Thus, each subject had corresponding brain imaging features with 246-dimensional gray matter volumes.

2.3.2. PGRS features

The genetic features for each subject were obtained by calculating the PGRS for SZ in PLINK. The PGRS computation method was developed by the International Schizophrenia Consortium (Consortium, 2009), as described in full detail in Walton's paper (Walton et al., 2013). In brief, the PGRS was calculated by summing the number of risk alleles weighted by the strength of the association of each SNP with schizophrenia, which was measured by the risk allele effect size (natural log of the odds ratio) of each SNP reported by a *meta*-analysis of a GWAS comprising a large number of Chinese individuals (Li et al., 2017). The subjects included in this study were independent of the participants from the *meta*-analysis. To obtain multiple independent PGRSs as genetic features for each individual, we calculated step-wise PGRSs using the SNP inclusion threshold from the *meta*-analysis GWAS statistical summary. The step length of the [0,0.05] interval was 0.001, and the step length of the [0.05,1] interval was 0.01. In this way, 145 PGRSs for each participant were obtained with the following different SNP inclusion thresholds: $0 \leq P_T < 0.001$, $0.001 \leq P_T < 0.002$, ……, $0.048 \leq P_T < 0.049$, $0.049 \leq P_T < 0.05$, $0.05 \leq P_T < 0.06$, $0.06 \leq P_T < 0.07$, ……, $0.98 \leq P_T < 0.99$, $0.99 \leq P_T < 1$. Therefore, each subject's data had corresponding genetic features for the 145-dimensional PGRSs.

2.4. Model construction and evaluation

After feature extraction, we constructed the classification models by using several traditional ML methods and different input features to differentiate the SZ patients from the NCs at an individual level. To ensure the generalization and reliability of the proposed models and to avoid the possibility of overfitting the models, we evaluated the performance of our models using leave-one-site-out CV, with one of eight sites used as independent test data and the remaining seven as training data. In the end, the average performance of the models validated at eight different independent sites was used as the final classification result of the model.

2.4.1. Experimental setup

Three machine learning methods, support vector machine (SVM), logistic regression (LR), and ensemble learning, which have frequently been used in previous neuroimaging research, were used to form the classification models. In addition, to investigate the ability of different features to differentiate SZ patients from NCs, we set up three sets of experiments for each classification model and utilized different input features. The input features were: (1) the brain imaging features of the average GMV of the 246 brain regions, (2) the genetic features of the 145 step-wise PGRSs, (3) a fusion of brain imaging and genetic features. The model training and test procedures of the three sets of experiments were the same so that the classification results could be compared fairly.

2.4.2. Model training and testing

During model training and testing, we used a nested CV framework to ensure that the data used for the hyper-parameter tuning and the data used to test the model were strictly independent. The inner 10-fold CV was used to select the optimal parameters, and an outer leave-one-site-out CV was used to test the performance of the models with the optimal parameters on a completely independent dataset.

We used common grid search strategies and 10-fold CV in the training set to find the optimal parameters. The parameter search range for each model was as follows. For the SVM, we used a linear kernel and optimized its C parameter value (a constant determining the tradeoff between training error and model flatness) by CV in the range of {10^{-6}, 10^{-5}, …, 10}. For the LR, we chose "liblinear" as the solver parameter and optimized its C parameter by cross-validating its value in the range of {0.001, 0.01, 0.1, 1, 10, 100} with a penalty parameter in {'l1', 'l2'}. In addition, when defining classifiers, we took class imbalance into account by adding the parameter class_weight = 'balanced' that weighted the penalty parameter (C) based on the samples in each class. As a result, the more samples a certain class had, the smaller its penalty parameter, 'C'. In this way, the learning bias problem caused by the imbalance of input samples could be well balanced. The other parameters were set to their default values.

After applying SVM and LR for the classification analysis, we used an ensemble learning method to improve the effect of the classification model. Ensemble learning is a strategy for improving existing classification models. It learns by training multiple homogeneous weakly supervised models on the basis of existing classification models. Integrating the learning results of each classification model according to appropriate rules can give full play to the advantages of each classification model to obtain a unified integrated learning model, thereby achieving the effect of improving the classification ability of the model and making the results more stable and stronger. In this study, after using SVM and LR to estimate the probability of each subject's having SZ, we took the weighted average of the probabilities given by these two models as the final probability. The specific training details were as follows: We assumed that the SVM and LR classifiers predicted that the probability of subject x having SZ was, respectively, $h_1(x)$, $h_2(x)$ and that each classifier had a weight w_1 or w_2, where w_i is the weight of the ith classifier h_i. First, under the condition that $w_i \geq 0$ and $w_1 + w_2 = 1$, we set the following 101 wt combinations (w_1, w_2): (0, 1), (0.01, 0.99), (0.02, 0.98), …, (0.98, 0.02), (0.99, 0.01), (1, 0). Second, we used these 101 wt combinations to obtain a weighted average of the probabilities given by the SVM and LR models so we could get 101 different models. The final prediction probability was:

$$H(x) = w_1 h_1(x) + w_2 h_2(x) \tag{1}$$

Third, we used $H(x)$ to predict the labels of the subjects and assessed the accuracy of 100 models. We finally selected the model with the highest accuracy as the best model.

2.4.3. Model evaluation

The performance of the classification models was evaluated by assessing accuracy, f1 score, and area under the receiver operating characteristic curve (AUC). Since eight different models were obtained by using leave-one-site-out CV, we averaged the performance of the eight models as the final result. The statistical significance of the accuracy was determined by permutation testing with 1000 permutations, in which the subjects were randomly assigned to one of the classes (SZ patients/NCs), so that the labels no longer matched the data in any meaningful way and repeated the process 1000 times. This resulted in a distribution of accuracies reflecting the null hypothesis that the classifier did not exceed chance. The number of times the classifier's performance was greater than or equal to the true accuracy was divided by 1000 to determine a *p*-value. A *p*-value lower than 0.05 was considered statistically significant.

2.4.4. Model comparison

First, we selected the optimal algorithm from the three ML algorithms (SVM, LR, and ensemble learning) for subsequent analysis. Since different evaluation metrics reflected different aspects of the models, in order to combine the results from the three metrics (accuracy, f1 score, and AUC) and evaluate the performance of models more comprehensively, we used a percentile-based scoring method (range from 0 to 100 for each criterion) proposed by Lee et al. (2021). The method can be expressed using the following equation:

$$P_{i,j} = \frac{(S_{i,j} - minV_{i,j})}{maxV_j - minV_j} \times 100, where V_j = \{S_{i,j} | i = 1, 2, \cdots, n\} \tag{2}$$

Where $S_{i,j}$ is the raw performance scores of the ith model for the jth metric, V_j is the vectorized performance scores $S_{i,j}$ across models; and $P_{i,j}$ is the percentile-based normalized score of the ith model for the jth metric. Since there were three feature sets, three algorithms, and eight sites for independent validation, the total number of models n was 72. The equation assigned 100 to the model with the highest performance score and 0 to the model with the lowest performance score for each criterion. Then, the normalized percentile scores of the three criteria were summed to one composite score (possible range: 0–300). We selected the algorithm that had the highest composite score for further model comparison.

Then, additional experiments and analyses were performed to compare the performance of models with different input features. During performing the leave-one-site-out CV, we randomly selected 50% of the pre-divided training set for training models and 50% of the pre-divided test set for validation. In other words, half of the data from one site was randomly selected as the test set, and half of the data from the remaining seven sites were randomly selected as the training set. In this process, stratified random sampling was used to ensure that the proportion of classes in the data did not change. The details of the experiment were the same as in section 2.4.2 except for stratified random sampling. Considering variations in the data, this random selection of samples was repeated 20 times for each model to get a more accurate estimate of model performance, resulting in 160 models for each type of input feature. Finally, we used a nonparametric Wilcoxon signed ranks test (Demšar, 2006) to compare the performance of models with different input features.

2.4.5. Identifying important brain imaging and genetic features

After training the ensemble learning model using brain imaging features and genetic data as predictors in section 2.4.2, we identified the brain and genetic features that were important in the model. In the trained SVM or LR models, the importance of a feature was estimated by its coefficient. For each of the eight models in the leave-one-site-out CV method, we used the best weight combination learned above to weight the coefficients of the features in the SVM and LR models. The Friedman test was performed to investigate whether the feature weights have significant differences between the eight models. In addition, we examined the test–retest reliability of the feature weights between these eight models by calculating the intraclass correlation coefficient (ICC). When the ICC was calculated using SPSS software (version 18.0, SPSS), we chose the following parameters: "two-way random/mixed model", "absolute agreement", and "average measure". Finally, we used the mean value of the coefficients of the 8 models to represent the importance of each feature and rank the weights.

3. Results

Classification performance of different models

Structural brain imaging data from 508 SZ patients and 502 healthy controls were obtained using 8 independent MRI scanners (Table 1). For each individual, we extracted the gray matter volume of 246 brain regions and calculated 145 step-wise PGRSs. Based on these features, we constructed 9 different classification models by using 3 different features (only imaging features, only genetic features, and both features combined) and 3 different machine learning methods (SVM, LR, and ensemble learning). Then we systematically evaluated the classification performance of the different models by a leave-one-site-out CV strategy.

The detailed classification performance for each model can be found in Table 2. Our analyses showed that the ensemble learning was the optimal algorithm among the three ML algorithms (Supplementary figure 1) and was used for subsequent analysis. When both the PGRS and brain sMRI features were included as predictors, the ensemble learning model had an average accuracy of 71.6% in discriminating SZ from NC. When only PGRS or brain sMRI features were included as predictors, the ensemble learning model had an average accuracy of 59.8% and 69.9%, respectively. The results of model comparison showed that the performance of the models with both PGRS and brain sMRI features as inputs were significantly better than those of the models with only PGRS features (one-tailed Wilcoxon signed ranks test, $P < .0001$) or only brain sMRI features as inputs (one-tailed Wilcoxon signed ranks test, $P < .0001$) (Supplementary figure 2).

3.1. Important genetic and brain features for classification

There was no statistically significant difference in the distribution of feature weights among the eight ensemble learning models with GMV and PGRSs as inputs (Friedman test, $P = .9745$). Since the $ICC = 0.928$, the test–retest reliability of the feature weights was excellent (Koo and Li, 2016). Thus, it was reasonable to select important features based on their average weights. Then, we listed the top ten features that contributed the most to the classification performance in the ensemble learning model in which both the PGRSs and the brain sMRI features were used. In total, three PGRSs and seven sMRI features were included (Table 3). For the top 10 features listed in Table 3, we listed the ranking of their importance scores across all left-out sites (see Supplementary Table 2). In general, the importance of the top features was relatively stable across the models. Interestingly, we found that the PGRS with a threshold $0 \leq P_T < 0.001$ was identified as the most important feature contributing to schizophrenia classification. This validated our hypothesis that genetic features may also be very important for the identification of schizophrenia. We also found that the most important brain gray matter volume features mainly focused on the brain regions located at the superior and middle temporal gyrus, orbital gyrus, precentral gyrus, basal ganglia, and thalamus. The ranking of all the features in the ensemble learning model can be found in Supplementary Table 1.

Table 1

Demographic characteristics of the participants from each site.

Site	Normal controls			Schizophrenia				All		
	Number	Male/Female	Age (y)	Number	Male/Female	Age (y)	PANSS	Number	Male/Female	Age (y)
1	98	52/46	25.77 ± 5.39	77	49/28	26.78 ± 6.32	78.23 ± 9.52	175	101/74	26.21 ± 5.82
2	45	16/29	26.78 ± 5.25	56	22/34	27.82 ± 6.61	78.04 ± 6.20	101	38/63	27.36 ± 6.04
3	66	34/32	30.69 ± 6.87	52	30/22	29.22 ± 7.45	88.15 ± 11.52	118	64/54	30.14 ± 7.20
4	79	40/39	28.66 ± 5.92	62	28/34	27.38 ± 5.47	82.18 ± 8.47	141	68/73	28.10 ± 5.74
5	34	18/16	29.14 ± 4.51	45	23/22	26.16 ± 4.70	89.69 ± 15.69	79	41/38	27.45 ± 4.82
6	72	43/29	26.09 ± 4.81	85	59/26	27.66 ± 5.88	85.92 ± 15.54	157	102/55	26.94 ± 5.46
7	56	31/25	24.46 ± 4.46	46	14/32	26.76 ± 6.28	87.60 ± 11.39	102	45/57	25.50 ± 5.45
8	52	21/31	31.53 ± 5.51	85	49/36	29.41 ± 7.54	85.48 ± 13.21	137	70/67	30.22 ± 6.90
Summary	502	255/247	27.69 ± 5.91	508	274/234	27.78 ± 6.48	84.07 ± 12.56	1010	529/481	27.73 ± 6.20

Table 2

Accuracies (f1 score/AUC) for each feature set and algorithm across all sites using leave-one-site-out cross-validation. * $P < .05$; **$P < .001$ (permutation test).

		PGRS	sMRI	sMRI & PGRS
Site 1	SVM	0.549 (0.533/ 0.593)	0.714** (0.688/ 0.790)	0.709** (0.683/ 0.711)
	LR	0.549 (0.515/ 0.597)	0.714** (0.702/ 0.804)	0.737** (0.723/ 0.776)
	Ensemble	0.560 (0.539/ 0.595)	0.731** (0.715/ 0.802)	0.743** (0.727/ 0.778)
Site 2	SVM	0.634* (0.667/ 0.674)	0.762** (0.782/ 0.843)	0.733** (0.727/ 0.870)
	LR	0.614* (0.642/ 0.676)	0.772** (0.785/ 0.830)	0.723** (0.726/ 0.862)
	Ensemble	0.614* (0.642/ 0.676)	0.792** (0.807/ 0.837)	0.743** (0.740/ 0.870)
Site 3	SVM	0.661* (0.636/ 0.687)	0.653** (0.696/ 0.794)	0.737** (0.756/ 0.826)
	LR	0.644* (0.604/ 0.715)	0.653** (0.701/ 0.773)	0.737** (0.760/ 0.809)
	Ensemble	0.670* (0.661/ 0.687)	0.661* (0.706/ 0.794)	0.763** (0.781/ 0.824)
Site 4	SVM	0.603* (0.576/ 0.663)	0.596* (0.537/ 0.630)	0.681** (0.622/ 0.715)
	LR	0.617* (0.585/ 0.672)	0.596* (0.551/ 0.640)	0.667** (0.612/ 0.710)
	Ensemble	0.631* (0.623/ 0.663)	0.617* (0.578/ 0.637)	0.688** (0.639/ 0.715)
Site 5	SVM	0.570 (0.622/ 0.560)	0.608* (0.587/ 0.703)	0.570 (0.564/ 0.697)
	LR	0.506 (0.562/ 0.536)	0.633* (0.613/ 0.731)	0.595 (0.610/ 0.658)
	Ensemble	0.582 (0.629/ 0.560)	0.633* (0.613/ 0.718)	0.608* (0.617/ 0.661)
Site 6	SVM	0.605* (0.613/ 0.601)	0.745** (0.740/ 0.851)	0.739** (0.729/ 0.838)
	LR	0.592* (0.632/ 0.622)	0.758** (0.747/ 0.865)	0.739** (0.729/ 0.852)
	Ensemble	0.612* (0.630/ 0.618)	0.764** (0.758/ 0.865)	0.758** (0.753/ 0.851)
Site 7	SVM	0.520 (0.495/ 0.530)	0.677* (0.593/ 0.755)	0.686* (0.680/ 0.760)
	LR	0.520 (0.495/ 0.525)	0.686* (0.628/ 0.757)	0.696** (0.659/ 0.741)
	Ensemble	0.549 (0.549/ 0.530)	0.706** (0.643/ 0.755)	0.726** (0.708/ 0.758)
Site 8	SVM	0.569 (0.634/ 0.638)	0.679** (0.732/ 0.711)	0.693** (0.741/ 0.752)
	LR	0.555 (0.616/ 0.595)	0.664** (0.697/ 0.734)	0.672** (0.724/ 0.766)
	Ensemble	0.569 (0.609/ 0.638)	0.686** (0.736/ 0.711)	0.701** (0.745/ 0.753)
Average	SVM	0.589 (0.597/ 0.618)	0.679 (0.669/ 0.759)	0.693 (0.688/ 0.771)
	LR	0.575 (0.581/ 0.617)	0.685 (0.678/ 0.767)	0.696 (0.693/ 0.772)
	Ensemble	0.598 (0.610/ 0.621)	0.699 (0.695/ 0.765)	0.716 (0.714/ 0.776)

Abbreviations: LR: logistic regression; PGRS: polygenic risk score; SVM: support vector machine; sMRI: structural magnetic resonance imaging.

Table 3

The weights and ranking of the top ten most important features.

Ranking	Features	Descriptions	Weight
1	PGRS	$0 \leq P_{T} < 0.001$	0.233
2	GMV	STG_R_6_3	0.231
3	GMV	Tha_L_8_2	0.165
4	GMV	MTG_R_4_1	0.114
5	GMV	BG_L_6_6	0.111
6	GMV	Tha_R_8_4	0.097
7	GMV	OrG_R_6_1	0.089
8	PGRS	$0.37 \leq P_{T} < 0.38$	0.089
9	PGRS	$0.65 \leq P_{T} < 0.66$	0.086
10	GMV	PrG_L_6_6	0.083

Abbreviations: BG: basal ganglia; GMV: gray matter volume; L: left hemisphere; MTG: middle temporal gyrus; OrG: orbital gyrus; PGRS: polygenic risk score; PrG: precentral gyrus; R: right hemisphere; STG: superior temporal gyrus; Tha: thalamus.

4. Discussion

By including large multisite samples ($N = 1010$) and using the novel strategy of integrating structural imaging data with polygenic risk scores, we provided potential biomarkers for the identification of schizophrenia. We constructed and evaluated nine classification models by combining different features or ML methods. The results obtained by using leave-one-site-out CV showed that the ensemble learning model with both the sMRI and genetic features had the best classification performance, indicating the necessity of integrating brain imaging with genetic information in schizophrenia classification.

Identifying schizophrenia patients based on quantitative biomarkers is important for translational neuroscience (Woo et al., 2017). Based only on our current classification results, the diagnostic utility of biological classifiers will not outperform that of psychiatrists using prescribed ICD/DSM criteria, but the search for biological classifiers and biomarkers may provide insight into the pathophysiology, etiology, genetic risk, and neural mechanisms of disease. Predicting diagnostic status and identifying diagnostic biomarkers based on reliable biological classifiers may be useful first steps toward the goal of developing biomarkers for predicting illness outcomes and treatment responses.

Based on neuroimaging data, a number of existing studies have attempted to use machine learning methods to distinguish patients with schizophrenia from healthy individuals (Kambeitz et al., 2015; Orru et al., 2012; Wolfers et al., 2015; Zarogianni et al., 2013). However, most of the current studies may have published overestimated accuracies due to the use of small and homogeneous samples (Schnack and Kahn, 2016), and many models may have low generalizability and limited clinical applicability because of not being validated using data from independent sites (Vieira et al., 2020). Therefore, to improve the clinical utility of the proposed machine learning model, our current study aimed to identify reliable and reproducible biomarkers. Several considerations and considerable effort have been made to improve this situation in our current research. First, the number of subjects in our

investigation was relatively large, which can help to guarantee the effectiveness of the classification and the heterogeneity of the data, to a certain extent. Second, we used the data from eight different sites and adopted a leave-one-site-out CV strategy, which is an excellent way to improve the robustness and generalizability of the model. This may also explain why the performance of our models did not seem to be as good as the performance reported in previous studies, in which the performance was not evaluated in independent samples. Third, the high dimensionality issue is one of the challenges in fusing information from neuroimaging data with genetic data (Zhou et al., 2019). One sMRI image usually contains millions of voxels, while the genetic data of one subject has thousands of SZ-related SNPs. To address the high dimensionality issue, we applied the ROI-based approach to extract a higher level of information by using brain anatomical priors and calculated the PGRS to integrate all the common genetic variant effects into a single risk metric. However, since the processing pipelines for the two modalities of features were not similar, future work should explore more matched feature reduction methods to alleviate the possible feature mismatch problem. Fourth, to the best of our knowledge, our study was the first attempt to integrate sMRI data and polygenic risk scores as features in a large-scale multisite schizophrenia classification, allowing us to achieve better results than using either type of feature alone.

Our classification results indicated that including both brain and PGRS features as model predictors improved the model's performance compared with only including either brain or PGRS features in the models. Genetic features were important for distinguishing SZ from NC. A random prediction usually yields an accuracy of about 50%, but using PGRS features as inputs increased the average accuracy of the ensemble learning models to 59.8% and even to 67.0% at site 3 (Table 2). In comparison, the ensemble learning models with brain features alone as inputs had a better performance in predicting SZ, with an average accuracy of 69.9%. Integrating brain and PGRS features as inputs in the models further increased the models' prediction performance by a moderate amount: The average accuracy of the ensemble learning models was 71.6%. In terms of model training, ensemble learning has the advantage of using a weighted combination rule to combine different classifiers, leading to a stacked generalization model (Polikar, 2012), thereby achieving slightly better classification performance. In addition, when identifying the important features by integrating the imaging with the genetic information, we found that the weight of the PGRS with a threshold $0 \leq P_T < 0.001$ was the largest. In fact, the PGRS with threshold $0 \leq P_T < 0.001$ measured the cumulative risk of the genetic variants most associated with schizophrenia based on previous GWAS evidence (Walton et al., 2013). Although no prior information was used in the classification model, this PGRS feature still obtained the greatest weight, further indicating that genetic features can indeed provide important information for classification.

By ranking all the features according to their weights, we found that the important brain regions that contributed to the differentiation between SZ patients and NC included the superior temporal gyrus (STG), middle temporal gyrus (MTG), orbital gyrus and precentral gyrus in the frontal lobe, basal ganglia, and thalamus. These alterations in gray matter volume are consistent with previous studies and may reflect the loss of neurons, neuropil areas, or interconnections between regions, leading to functional deficits in schizophrenia. For example, the reduction in STG volume has been widely observed in schizophrenia (Ellison-Wright and Bullmore, 2010; Ellison-Wright et al., 2008; Haijma et al., 2013; Vita et al., 2012). In particular, a significantly more pronounced decrease in GM volume was detected in the STG and STG subregions (the Heschl's gyrus (HG) and planum temporale (PT)) in patients with SZ (Vita et al., 2012). However, several other cortical subregions did not exhibit such a large reduction in volume over time. This seems to be consistent with our findings that the STG is the brain region that contributes the most to classification. Anatomically, the superior part of the STG includes the primary and secondary auditory sensory cortex (HG and PT), and the STG has been found to be involved in early auditory

processing common to speech and nonspeech stimuli (Binder et al., 2000). Changes in STG volume have been related to positive symptom severity, especially thought disorder (posterior STG) (Hirayasu et al., 1998; Shenton et al., 1992; Vita et al., 1995) and hallucinations (anterior and middle STG) (Barta et al., 1990; Flaum et al., 1995; Levitan et al., 1999; Menon et al., 1995); HG volume abnormalities have been correlated with the severity of formal thought disorder (Rajarethinam et al., 2000; Yamasaki et al., 2007), and PT alterations have been correlated with the language and thought disorder of schizophrenia (Shenton et al., 1992; Yamasaki et al., 2007). In contrast, the MTG has received less attention, but there are still some studies that have reported a reduction in the MTG volume in schizophrenia (Ellison-Wright et al., 2008; Honea et al., 2005; Kuroki et al., 2006). The MTG is known to play an important role in both dorsal and visual pathways (Sewards and Sewards, 2002) as well as in multimodal and higher sensory processing (Doniger et al., 2002). Since functional deficits in language, memory (Nestor et al., 1998), and visual spatial perception have all been reported in schizophrenia, the gray matter abnormality in the MTG may be an important anatomical substrate. In addition, the basal ganglia structures (caudate, putamen, and globus pallidus) and thalamus are also the focus of research in schizophrenia, and their volume abnormalities have been reported in many previous studies (Ellison-Wright and Bullmore, 2010; Ellison-Wright et al., 2008; Glahn et al., 2008; Haijma et al., 2013; Van Erp et al., 2014; Van Erp et al., 2016). Functionally, the basal ganglia structures play an important role in cognitive, sensory, and motor processing (Keshavan et al., 1998), and the thalamus is involved in attention and information processing and serves as a 'filter' for gating the input of sensory signals (Fuster, 2015). Other studies suggest that executive functioning deficits in schizophrenia may be mediated by basal ganglia-thalamocortical circuitry disruptions (Camchong et al., 2006). Accordingly, these important brain features that we identified as discriminating between the patients with SZ and the NCs may contribute to the symptoms of schizophrenia and provide insight into the underlying neural mechanisms of schizophrenia.

There are several improvements that can be considered for future work. First, since the brain features only included the gray matter volume in the present study, a better classification ability may be achievable by including additional types of imaging data such as brain dynamics, structural connectome data, etc. We will evaluate the incremental gain of other modalities of features in the future using a similar design. Second, the models used in this study were traditional "shallow" machine learning techniques, but their performance has been surpassed by advanced deep learning in many applications. Future studies could try to use deep learning models to extract complex latent features from the original data through consecutive nonlinear transformations and may be able to capture the interactions and joint effect of the brain imaging and genetic features. Third, all the data used in this study were from the Han Chinese population, and the models and results of the current study need to be independently validated on other multi-site schizophrenia data from other ethnicities. Fourth, there may be redundancy between features in this study. When we calculated the step-wise PGRSs, there was no overlap between the SNP inclusion thresholds, which we think can reduce the correlation between PGRSs to a certain extent. However, we have to admit that even though we have tried to do so, there was still some correlation between the PGRSs. A possible reason may be that the SNPs in different inclusion thresholds are not completely independent. In addition, with the development of imaging genetics, most aspects of brain structure and function have shown some level of heritability (Thompson et al., 2001, 2013; Arnatkeviciute et al., 2021). In particular, a recent GWAS of 19,692 individuals identified genetic variants influencing regional brain volumes (Zhao et al., 2019), 46 of which were associated with mental health disorders, including schizophrenia. In a post hoc analysis, we performed a Pearson's correlation analysis and found no significant association between PGRSs and brain structure features used in the study ($-0.1371 < r < 0.1267$, $P > 1.40\mathrm{e}{-6}$ $(0.05/246/145)$, not reaching significance after

Bonferroni correction for multiple testing). Thus, we consider that the redundancy between genetics and brain structure features may be relatively low. However, given the complexity of the relationship between genetics and brain features, future studies still need to decompose the genetic and phenotype variances and further explore the interactions between genetics and brain imaging data.

5. Conclusion

In summary, we trained ML models on a large number of multisite subjects to classify SZ patients and NCs. By using the a previously unused strategy of integrating sMRI data with PGRS, our models performed well on the SZ classification task, revealing the importance of utilizing multiple data streams to classify complex diseases and showing that genetic features can provide important information about SZ. Our study provided potential biomarkers for identifying individuals with SZ and may help to guide future research utilizing machine learning in identifying SZ.

CRediT authorship contribution statement

T.J. and B.L. led the project. B.L. and K.H. were responsible for the study concept and the design of the study. K.H. and B.L. also analyzed the data, created the figures and wrote the manuscript. M.W., Y.L., H.Y., M.S. and T.J. participated in discussions of the results and the manuscript. J.C., Y.C., H.N.W., H.G., P.W., L.Lv, Y.Y., P.L., L.Lu, J.Y., H.L.W., H.Z., D.Z., H.W. and Y.N. contributed to the data acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2021.102860.

References

Anderson, J.S., Shade, J., Diblasi, E., et al., 2019. Polygenic risk scoring and prediction of mental health outcomes. Current opinion in psychology [J] 27, 77–81.

Arnatkeviciute, A., Fulcher, B., Bellgrove, M., et al., 2021. Where the genome meets the connectome: understanding how genes shape human brain connectivity. PsyArXiv.

Barta P E, Pearlson G D, Powers R E, et al. 1990. Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia. The American journal of psychiatry [J].

Binder, J.R., Frost, J.A., Hammeke, T.A., et al., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cerebral cortex [J] 10, 512–528.

Camchong, J., Dyckman, K.A., Chapman, C.E., et al., 2006. Basal ganglia-thalamocortical circuitry disruptions in schizophrenia during delayed response tasks. Biological psychiatry [J] 60, 235–241.

Cardno, A.G., Gottesman, I.I., 2000. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. American journal of medical genetics [J] 97 (1), 12–17.

Consortium I S, 2009. Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. Nature [J] 460, 748.

Consortium S W G O T P G 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature [J], 511: 421-427.

Delaneau, O., Marchini, J., Zagury, J.-F., 2012. A linear complexity phasing method for thousands of genomes. Nature methods [J] 9 (2), 179–181.

Dickey, C.C., Salisbury, D.F., Nagy, A.I., et al., 2004. Follow-up MRI study of prefrontal volumes in first-episode psychotic patients. Schizophrenia research [J] 71, 349.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research [J] 7, 1–30.

Doniger, G.M., Foxe, J.J., Murray, M.M., et al., 2002. Impaired visual object recognition and dorsal/ventral stream interaction in schizophrenia. Archives of general psychiatry [J] 59, 1011–1020.

Ellison-Wright, I., Bullmore, Ed., 2010. Anatomy of bipolar disorder and schizophrenia: a meta-analysis. Schizophrenia research [J] 117 (1), 1–12.

Ellison-Wright, I., Glahn, D.C., Laird, A.R., et al., 2008. The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. American Journal of Psychiatry [J] 165, 1015–1023.

Fan, L., Li, H., Zhuo, J., et al., 2016. The human brainnetome atlas: a new brain atlas based on connectional architecture. Cerebral cortex [J] 26, 3508–3526.

Flaum, M., O'leary D S, Swayze Ii V W, et al., 1995. Symptom dimensions and brain morphology in schizophrenia and related psychotic disorders. Journal of Psychiatric Research [J] 29, 261–276.

Fuster, J., 2015. The prefrontal cortex [M]. Academic Press.

Glahn, D.C., Laird, A.R., Ellison-Wright, I., Thelen, S.M., Robinson, J.L., Lancaster, J.L., Bullmore, E., Fox, P.T., 2008. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. Biological psychiatry [J] 64 (9), 774–781.

Haijma, S.V., Van Haren, N., Cahn, W., et al., 2013. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. Schizophrenia bulletin [J] 39, 1129–1138.

Haller C S, Padmanabhan J L, Lizano P, et al. 2014. Recent advances in understanding schizophrenia. F1000prime reports [J], 6.

Hirayasu, Y., Shenton, M.E., Salisbury, D.F., Dickey, C.C., Fischer, I.A., Mazzoni, P., Kisler, T., Arakaki, H., Kwon, J.S., Anderson, J.E., Yurgelun-Todd, D., Tohen, M., McCarley, R.W., 1998. Lower left temporal lobe MRI volumes in patients with first-episode schizophrenia compared with psychotic patients with first-episode affective disorder and normal subjects. American Journal of Psychiatry [J] 155 (10), 1384–1391.

Honea, R., Crow, T.J., Passingham, D., et al., 2005. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. American Journal of Psychiatry [J] 162, 2233–2245.

Howie, B.N., Donnelly, P., Marchini, J., Schork, N.J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genet [J] 5 (6), e1000529.

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

Keshavan, M.S., Rosenberg, D., Sweeney, J.A., et al., 1998. Decreased caudate volume in neuroleptic-naive psychotic patients. American Journal of Psychiatry [J] 155, 774–778.

Kety, S.S., Wender, P.H., Jacobsen, B., et al., 1994. Mental illness in the biological and adoptive relatives of schizophrenic adoptees: replication of the Copenhagen study in the rest of Denmark. Archives of general psychiatry [J] 51, 442–455.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine [J] 15 (2), 155–163.

Kuroki, N., Shenton, M.E., Salisbury, D.F., et al., 2006. Middle and inferior temporal gyrus gray matter volume abnormalities in first-episode schizophrenia: an MRI study. American Journal of Psychiatry [J] 163, 2103–2110.

Lee, J.-J., Kim, H.J., Čeko, M., et al., 2021. A neuroimaging biomarker for sustained experimental and clinical pain. Nature medicine [J] 27, 174–182.

Levitan, C., Ward, P.B., Catts, S.V., 1999. Superior temporal gyral volumes and laterality correlates of auditory hallucinations in schizophrenia. Biological psychiatry [J] 46 (7), 955–962.

Li, Z., Chen, J., Yu, H., et al., 2017. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nature genetics [J] 49, 1576.

Liu, Shu, Li, Ang, Liu, Yong, et al., 2020. Polygenic effects of schizophrenia on hippocampal grey matter volume and hippocampus–medial prefrontal cortex functional connectivity. The British Journal of Psychiatry 216 (5), 267–274.

Menon, R.R., Barta, P.E., Aylward, E.H., et al., 1995. Posterior superior temporal gyrus in schizophrenia: grey matter changes and clinical correlates. Schizophrenia research [J] 16, 127–135.

Nestor, P.G., Akdag, S.J., O'donnell B F, et al., 1998. Word recall in schizophrenia: a connectionist model. American Journal of Psychiatry [J] 155, 1685–1690.

Nieuwenhuis, M., Van Haren, N.E., Pol, H.E.H., et al., 2012. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. Neuroimage [J] 61, 606–612.

Orru, G., Pettersson-Yeo, W., Marquand, A.F., et al., 2012. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience & Biobehavioral Reviews [J] 36, 1140–1152.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS genet [J] 2 (12), e190. https://doi.org/10.1371/journal.pgen.0020190.

Pettersson-Yeo, W., Benetti, S., Marquand, A.F., Dell'Acqua, F., Williams, S.C.R., Allen, P., Prata, D., McGuire, P., Mechelli, A., 2013. Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. Psychological medicine [J] 43 (12), 2547–2562.

Pinaya, W.H., Gadelha, A., Doyle, O.M., et al., 2016. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. Scientific reports [J] 6, 1–9.

Polikar, R., 2012. In: Ensemble Machine Learning. Springer US, Boston, MA, pp. 1–34. https://doi.org/10.1007/978-1-4419-9326-7_1.

Price, A.L., Patterson, N.J., Plenge, R.M., et al., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics [J] 38, 904–909.

Purcell, S., Neale, B., Todd-Brown, K., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics [J] 81, 559–575.

Rajarethinam, R., Dequardo, J., Nalepa, R., et al., 2000. Superior temporal gyrus in schizophrenia: a volumetric magnetic resonance imaging study. Schizophrenia research [J] 41, 303–312.

Martin Rozycki Theodore D Satterthwaite Nikolaos Koutsouleris Guray Erus Jimit Doshi Daniel H Wolf Yong Fan Raquel E Gur Ruben C Gur Eva M Meisenzahl Chuanjun Zhuo Hong Yin Hao Yan Weihua Yue Dai Zhang Christos Davatzikos Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals 44 5 2018 2018 1035 1044.

Schnack, H.G., Kahn, R.S., 2016. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. Frontiers in psychiatry [J] 7, 50.

Schnack, H.G., Nieuwenhuis, M., van Haren, N.E.M., Abramovic, L., Scheewe, T.W., Brouwer, R.M., Hulshoff Pol, H.E., Kahn, René.S., 2014. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. Neuroimage [J] 84, 299–306.

Sewards, T.V., Sewards, M.A., 2002. On the neural correlates of object recognition awareness: relationship to computational activities and activities mediating perceptual awareness. Consciousness and Cognition [J] 11 (1), 51–77.

Shenton, M.E., Kikinis, R., Jolesz, F.A., et al., 1992. Abnormalities of the left temporal lobe and thought disorder in schizophrenia: a quantitative magnetic resonance imaging study. New England Journal of Medicine [J] 327, 604–612.

Sun, J., Maller, J.J., Guo, L., et al., 2009. Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. Brain research reviews [J] 61, 14–32.

Thompson, P.M., Cannon, T.D., Narr, K.L., et al., 2001. Genetic influences on brain structure. Nature neuroscience [J] 4, 1253–1258.

Thompson, P.M., Ge, T., Glahn, D.C., et al., 2013. Genetics of the connectome. Neuroimage [J] 80, 475–488.

Thorisson, G.A., Smith, A.V., Krishnan, L., et al., 2005. The international HapMap project web site. Genome research [J] 15, 1592–1593.

Van Erp, T.G., Greve, D.N., Rasmussen, J., et al., 2014. A multi-scanner study of subcortical brain volume abnormalities in schizophrenia. Psychiatry Research: Neuroimaging [J] 222, 10–16.

Van Erp, T.G., Hibar, D.P., Rasmussen, J.M., et al., 2016. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Molecular psychiatry [J] 21, 547–553.

Vieira, S., Gong, Q.-Y., Pinaya, W.H., et al., 2020. Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence. Schizophrenia bulletin [J] 46, 17–26.

Vita, A., De Peri, L., Deste, G., Sacchetti, E., 2012. Progressive loss of cortical gray matter in schizophrenia: a meta-analysis and meta-regression of longitudinal MRI studies. Translational psychiatry [J] 2 (11), e190.

Vita, A., Dieci, M., Giobbio, G.M., et al., 1995. Language and thought disorder in schizophrenia: brain morphological correlates. Schizophrenia research [J] 15, 243–251.

Walton, E., Geisler, D., Lee, P.H., et al., 2014. Prefrontal inefficiency is associated with polygenic risk for schizophrenia. Schizophrenia bulletin [J] 40, 1263–1271.

Walton, E., Turner, J., Gollub, R.L., et al., 2013. Cumulative genetic risk and prefrontal activity in patients with schizophrenia. Schizophrenia bulletin [J] 39, 703–711.

Wolfers, T., Buitelaar, J.K., Beckmann, C.F., et al., 2015. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. Neuroscience & Biobehavioral Reviews [J] 57, 328–349.

Woo, C.-W., Chang, L.J., Lindquist, M.A., et al., 2017. Building better biomarkers: brain models in translational neuroimaging. Nature neuroscience [J] 20, 365.

Xiao, Y., Yan, Z., Zhao, Y., et al., 2019. Support vector machine-based classification of first episode drug-naïve schizophrenia patients and healthy controls using structural MRI. Schizophrenia research [J] 214, 11–17.

Yamasaki, S., Yamasue, H., Abe, O., et al., 2007. Reduced planum temporale volume and delusional behaviour in patients with schizophrenia. European archives of psychiatry and clinical neuroscience [J] 257, 318–324.

Yamasue, H., Iwanami, A., Hirayasu, Y., et al., 2004. Localized volume reduction in prefrontal, temporolimbic, and paralimbic regions in schizophrenia: an MRI parcellation study. Psychiatry Research: Neuroimaging [J] 131, 195–207.

Yang, H., Liu, J., Sui, J., et al., 2010. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. Frontiers in human neuroscience [J] 4, 192.

Zarogianni, E., Moorhead, T.W.J., Lawrie, S.M., 2013. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. NeuroImage: Clinical [J] 3, 279–289.

Zhao, B., Luo, T., Li, T., Li, Y., Zhang, J., Shan, Y., Wang, X., Yang, L., Zhou, F., Zhu, Z., Zhu, H., 2019. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. Nature genetics [J] 51 (11), 1637–1644.

Zhou, S.-Y., Suzuki, M., Hagino, H., et al., 2005. Volumetric analysis of sulci/gyri-defined in vivo frontal lobe regions in schizophrenia: precentral gyrus, cingulate gyrus, and prefrontal region. Psychiatry Research: Neuroimaging [J] 139, 127–139.

Zhou, T., Thung, K.H., Zhu, X., et al., 2019. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Human brain mapping [J] 40, 1001–1016.