



RESEARCH ARTICLE

Morphological fingerprinting: Identifying patients with first-episode schizophrenia using auto-encoded morphological patterns

Huaiqiang Sun^{1,2} | Guoting Luo¹ | Su Lui^{1,2} | Xiaoqi Huang^{1,2}  |
John Sweeney^{1,3} | Qiyong Gong^{1,4} 

¹Huaxi MR Research Center (HMRC),
Department of Radiology, West China
Hospital of Sichuan University, Chengdu,
China

²Functional and Molecular Imaging Key
Laboratory of Sichuan Province, West China
Hospital of Sichuan University, Chengdu,
China

³Department of Psychiatry and Behavioral
Neuroscience, University of Cincinnati College
of Medicine, Cincinnati, Ohio, USA

⁴Department of Radiology, West China
Xiamen Hospital of Sichuan University,
Xiamen, China

Correspondence

Qiyong Gong, Department of Radiology, West
China Xiamen Hospital of Sichuan University,
699 Jinyuan Xi Road, Jimei District, 361021,
Xiamen, Fujian, China.
Email: qiyonggong@hmrc.org.cn

Funding information

China Association for Science and Technology,
Grant/Award Number: YESS20160060;
National Natural Science Foundation of China,
Grant/Award Numbers: 81974278, 81621003,
81761128023, 81820108018, 82027808;
National Key R&D Program of China,
Grant/Award Number: 2022YFC2009900;
NIH/NIMH, Grant/Award Number:
R01MH112189-01

Abstract

Although a large number of case-control statistical and machine learning studies have been conducted to investigate structural brain changes in schizophrenia, how best to measure and characterize structural abnormalities for use in classification algorithms remains an open question. In the current study, a convolutional 3D auto-encoder specifically designed for discretized volumes was constructed and trained with segmented brains from 477 healthy individuals. A cohort containing 158 first-episode schizophrenia patients and 166 matched controls was fed into the trained autoencoder to generate auto-encoded morphological patterns. A classifier discriminating schizophrenia patients from healthy controls was built using 80% of the samples in this cohort by automated machine learning and validated on the remaining 20% of the samples, and this classifier was further validated on another independent cohort containing 77 first-episode schizophrenia patients and 58 matched controls acquired at a different resolution. This specially designed autoencoder allowed a satisfactory recovery of the input. With the same feature dimension, the classifier trained with autoencoded features outperformed the classifier trained with conventional morphological features by about 10% points, achieving 73.44% accuracy and 0.8 AUC on the internal validation set and 71.85% accuracy and 0.77 AUC on the external validation set. The use of features automatically learned from the segmented brain can better identify schizophrenia patients from healthy controls, but there is still a need for further improvements to establish a clinical diagnostic marker. However, with a limited sample size, the method proposed in the current study shed insight into the application of deep learning in psychiatric disorders.

KEYWORDS

autoencoder, classification, deep learning, feature engineering, machine learning, schizophrenia

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Since Johnstone et al. first reported significant increased ventricular size in patients with long-term ill schizophrenia from CT images (Johnstone et al., 1976), brain structural changes in schizophrenia have been shown in a large number of imaging studies, which have iteratively better characterized with the development of medical imaging and image analysis techniques. A meta-analysis conducted by Radua et al. reported gray matter volume decrements clustered in the insula, operculum, the superior temporal gyrus, and the medial frontal and anterior cingulate cortices in patients with first-episode psychosis in 25 case-control voxel-based morphometry (VBM) studies (Radua et al., 2012). Subsequently developed surface-based analysis, which is more optimal for the analysis of highly curved gray matter structure than VBM, showed reduced cortical thickness in the prefrontal, temporal and parietal cortices in both long-term ill and first-episode schizophrenia patients (van Erp et al., 2018; Sugihara et al., 2017; Takayanagi et al., 2020; Yan et al., 2019).

In addition to gray matter volume and thickness, other morphological features such as gyrification (Schaer et al., 2008) and curvature have also been implicated in schizophrenia by means of surface-based morphometry. Compared to cortical thickness, these measures are more influenced by genetic and embryonic developmental factors, and are more reflective of neural connectivity established during brain maturation (Mota & Herculano-Houzel, 2012). It has been reported that Gyrification Index (GI) is altered in the bilateral insula, temporal pole and left orbitofrontal cortex in patients with schizophrenia (Spalthoff et al., 2018). However, findings of cortical gyrification in first-episode schizophrenia remain uncertain and variable, as both increased and decreased local GI was observed in regions of the frontal cortex (Matsuda & Ohi, 2018).

Although informative, analyses using conventional univariate group statistics ignore potential interaction among brain regions or voxels. The combination of neuroimaging features and machine learning algorithms are capable of analyzing brain measures in a multivariate way to detect patterns in brain morphology that are different between patients and healthy controls. Most machine learning studies in neuroimaging use classical features, such as cortical thickness, as input to the algorithm. The using of predefined features makes classification models more readily interpretable by identifying brain regions and features that contribute significantly to classification. However, a recent multicenter machine learning study in schizophrenia found that, although having better classification performance than other type of features, utilizing regional cortical thickness as an input to the classifier could only achieve 55%–70% classification accuracy (Winterburn et al., 2019).

Although previous studies demonstrated that neuroanatomical alterations in neuropsychiatric disorders tend to be subtle and widely distributed spatially (Ellison-Wright et al., 2008), it remains an open question how best to measure and describe the structural abnormalities as the cerebral cortex of the human brain has a complex morphological structure that consists of both folded and smooth surfaces. We thus hypothesize that there may be a specific “pattern” in the brain of

schizophrenic patients at the higher level, which may not be fully represented by the analysis of single or even multiple morphological measures widely used in analyses. Therefore, a high-level representation may be needed to abstractly characterize brain morphology in order to achieve improved classification precision.

In the current study, we developed a deep convolutional autoencoder (CAE) network to encode the volume of segmented brain and in order to reconstruct the volume as closely as possible to the input. The encoded latent vector is treated as the “morphological fingerprint” and fed into a machine learning algorithm to build a schizophrenia discrimination model. Finally, the performance of the discrimination model built on morphological fingerprint was compared with the model built on conventional brain morphological features in terms of accuracy, sensitivity, specificity, and area under the ROC curve (AUC).

2 | METHOD

2.1 | Datasets

The study was approved by the ethics committee of West China Hospital and written informed consent was obtained from all participants. The study included three datasets (Table 1). Dataset 1, which contains 477 healthy volunteers, was used for training of the CAE. Dataset 2, which contains 158 first-episode schizophrenia patients and 166 demographically matched healthy controls, was used for training the classification model and internal validation. Dataset 3, which contains an independent cohort of 77 first-episode schizophrenia patients and 58 matched healthy controls, was treated as an independent validation set. Illness duration of all patients was less than 2 years, with illness onset evaluated by the Nottingham Onset Schedule using the information provided by patients, family members, and other sources when available. Healthy controls in Dataset 1 used for CAE development did not overlap with healthy individuals of Dataset 2 or 3 (Table 1).

Diagnoses of schizophrenia were determined using the Structured Interview for the DSM-IV (SCID-P). Severity of psychopathology was evaluated using the Positive and Negative Syndrome Scale (PANSS). Psychiatric evaluations and MR scans were performed prior to any medication treatment.

Healthy controls were recruited by poster advertisement, and did not differ significantly from the corresponding schizophrenia group on age, sex, and years of education. All controls were screened using the SCID-NP to confirm the lifetime absence of psychiatric illnesses. Control subjects reported no known history of psychiatric illness in first-degree relatives. The following exclusion criteria applied to all of the above groups: any neurological disorder, lifetime drug or alcohol abuse or dependency, pregnancy, or significant systemic illness such as hepatitis or cardiovascular disease. Brain MR images were inspected by an experienced neuroradiologist to check image quality and exclude patients with visible cerebral abnormalities of neuroradiological significance.

TABLE 1 Demographic and imaging information of three datasets.

Dataset	Content	Scanner	Resolution	Age	Gender	Educational year	GAF score	PANSS score			Usage
								Positive	Negative	General	
1	477 HC	GE EXCITE 3 T (194 subjects)	1 mm isotropic	30.47 ± 15.00 (23)	247 M/230F	14.15 ± 2.93(15)	NA	NA	NA	NA	Autoencoder training
		Siemens Trio Tim 3 T (283 subjects)									
2	158 SCZ 166 HC	GE EXCITE 3 T	1 mm isotropic	23.68 ± 7.42	74 M/84F	14.86 ± 3.41(15)	30.10 ± 11.13	18.73 ± 7.97	24.71 ± 6.36	45.89 ± 9.82	model training and internal test
3	77 SCZ 58 HC	Siemens Trio Tim 3 T	0.8 mm isotropic	23.62 ± 6.03 24.37 ± 4.48 24.21 ± 5.65	87 M/79F 36 M/41F 28 M/30F	15.05 ± 3.73(15) 14.32 ± 3.13(15) 14.68 ± 2.72(15)	NA 32.02 ± 12.58 NA	NA 18.15 ± 7.34 NA	NA 23.24 ± 6.57 NA	NA 44.29 ± 9.41 NA	External test

Abbreviations: HC, healthy controls; SCZ, Schizophrenia.

2.2 | Image acquisition

194 subjects of Dataset 1 were scanned on a GE EXCITE 3 T MRI scanner (Milwaukee, USA). High-resolution T1 weighted volumes were obtained using a spoiled gradient recall (SPGR) sequence (TR = 8.53 ms, TE = 3.4 ms, TI = 400 ms, Flip angle = 12°, resolution = 1 mm isotropic) and an 8-channel phase-array head coil. The other 283 subjects of Dataset 1 were scanned on a Siemens Tim Trio 3 T MRI scanner (Erlangen, German). High-resolution T1 weighted (T1w) volumes were obtained using a MPRAGE sequence (TR = 1900 ms, TE = 2.26 ms, TI = 900 ms, Flip angle = 9°, resolution = 1 mm isotropic) and a 12-channel phase-array head coil.

Subjects in Dataset 2 were scanned on the same GE EXCITE 3 T MRI scanner with identical imaging parameters as individuals scanned with this scanner in Dataset 1.

Subjects in Dataset 3 were scanned on the same Siemens Tim Trio 3 T MRI scanner as individuals in Dataset 1, but using a different coil (32-channel phase array coil) and imaging parameters (MPRAGE, TR = 2400 ms, TE = 2.01 ms, TI = 1000 ms, Flip angle = 8°, resolution = 0.8 mm isotropic).

2.3 | Image preprocessing and segmentation

All T1w volumes were spatially normalized through iterative rigid registration. Initially, all the T1 volumes were rigid registered to the MNI152 template using flirt tool provided in the fsl package (Jenkinson et al., 2012). Then all the registered T1w volumes were averaged to form an initial study specific template. In the next iteration, all the original T1w volumes registered to the newly generated study specific template and then averaged again to form a new study specific template. The iteration stopped when the study-specific template generated in the current iteration had not changed compared to the one generated in the previous iteration.

All the registered T1w volumes from the last iteration were further processed with Freesurfer's recon-all processing pipeline with the Desikan-Killiany-Tourville (DKT) atlas to generate the segmented brains (Fischl, 2012). The DKT atlas defined 68 cortical regions (34 per hemisphere) and 14 subcortical structures (seven per hemisphere). The quality of segmentation was visual inspected by overlaying the segmented brain on the T1 weighted image (Figure S1). The brain surface was reconstructed from a segmented brain and the following measures were computed: 1) Volume of 14 subcortical regions normalized by the participant's total cerebral volume (14 features); 2) mean cortical thickness, mean curvature, gauss curvature, cortical folding index and cortical curvature index from 68 cortical regions (68*5 = 340 features). The quality of labeled surface mesh and parametric surface mesh generated by Freesurfer (FS) were visual inspected to avoid topological defeat and apparent calculation error (Figure S2). A total of 354 (340 + 14) shape-related features that represent cerebral morphometry were extracted from each T1w volume for classical machine learning.

Next, all cortical regions and subcortical structures were merged into one label respectively, making the segmented brain contains only

five components: background, cortical gray matter, white matter, cerebrospinal fluid, and subcortical gray matter. The label-merged volumes were used as input for the autoencoder.

2.4 | The architecture of convolutional autoencoder and network training

The architecture of the CAE is illustrated in Figure 1. Briefly, the encoder contained four 3D convolutional layers and one fully connection layer. The number of channels linearly increases from 64 to 512. The pooling layer was replaced by setting the stride in the convolutional layer to two to avoid loss of information (Du et al., 2019). The decoder contains one fully connection layer, which restores the latent vector to volume, and four transposed convolution layers with linearly reduced channels to upsample volume to the size of the input. The length of the latent vector was initially set to 354 to match the length of feature vector obtained from the FS pipeline. The effect of the length of the latent vector on the classification performance was also evaluated by setting the length to 708, 177, 88, 44, and 22 (double, 1/2, 1/4, 1/8 and 1/16 of 354).

Instead of using the traditional MSE loss when training the proposed CAE, we used the same strategy as used for training deep neural networks for image segmentation purposes (Ma et al., 2021). After the last transpose convolution, a volume of five channels with the same size as the original input was generated and normalized using the softmax function. The voxel values in each channel represent their probability of belonging to the five components. The five-channel probability volume was then fed into a specially designed loss function (\mathcal{L}), which is a hybrid of focal loss (\mathcal{L}_{Focal}) and Hausdorff distance (HD) loss (\mathcal{L}_{HD}).

$$\mathcal{L} = \mathcal{L}_{Focal} + \mathcal{L}_{HD}$$

$$\mathcal{L}_{Focal}(p_t) = -(1 - p_t)^2 \log(p_t)$$

$$p_t = \begin{cases} p, & \text{if } y = \text{class label} \\ 1 - p, & \text{otherwise} \end{cases}$$

$$\mathcal{L}_{HD} = \frac{1}{|\Omega|} \sum_{\Omega} \left[(S - G)^2 \circ (G_{DTM}^2 + S_{DTM}^2) \right]$$

Where $p_t \in [0, 1]$ is the model's estimated probability for the class with y equal to the predefined label. Ω denotes the grid on which the image is defined, which means that max is with respect to all pixels. G_{DTM} and S_{DTM} denote the distance transform maps of ground truth G (Figure 2c) and predicted segmentation S , respectively. The \circ sign denotes the Hadamard (i.e. voxel-wise) product. Minimizing this hybrid loss function maximizes the recovery of the boundaries of each tissue (Figure 2d).

This CAE was trained for 500 epochs and each data batch included 32 resized label-merged volumes for parameter optimization via Adam (Adaptive Moment estimation)-based iteration (Kingma & Ba, 2014). More detail about the design of the network is provided in Supplement eMethod. Our implementation of the proposed CAE and hybrid loss training in PyTorch is available at <https://github.com/MAI-Lab-West-China-Hospital/Convolutional-autoencoder-for-labeled-volume>

2.5 | Training and evaluation the classification model via automated machine learning

Stratified hold-out partitioning was performed on Dataset 2 with ratio of 4:1 to generate training and test sets, and the ratio of patients to controls in both training and test sets was similar to that of the entire dataset.

An automated machine learning (AutoML) pipeline implemented in the open source package PyCaret was used to build classification models with feature vectors from both freesurfer (FS) and CAE. This pipeline first performs an exhaustive search in the pool of all available classification algorithms with default hyperparameters. The best possible classification algorithm was determined through nested 10-fold cross-validation. The selected classifier was further fine-tuned using different hyperparameter combinations inside a grid search cross-validation. The hyperparameter configuration producing the best classification performance was chosen, and then the model was fit to the entire training set to build the final model. The final model was tested on both the internal (Dataset 2) and external test set (Dataset 3). The entire workflow is presented in Figure 3. Evaluation of classification accuracy, specificity, sensitivity and AUC, were used to evaluate classifier performance.

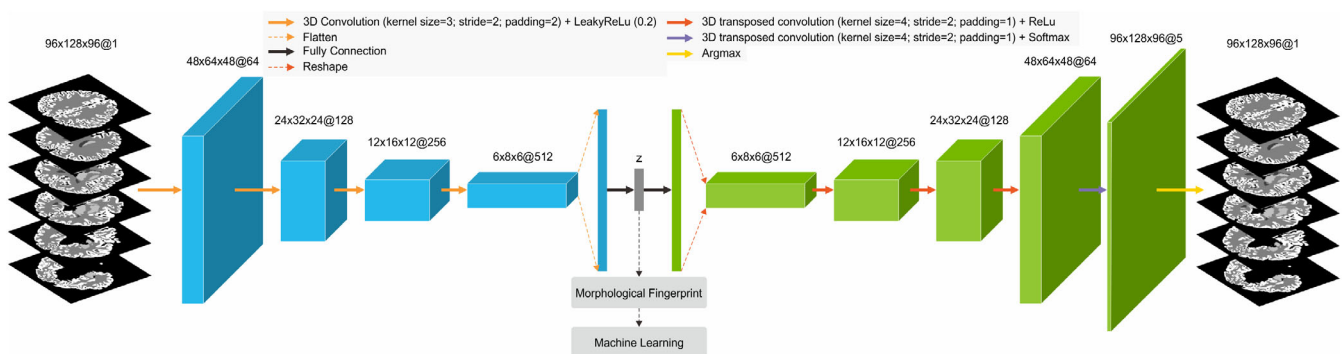


FIGURE 1 The architecture of the proposed 3D convolutional autoencoder.

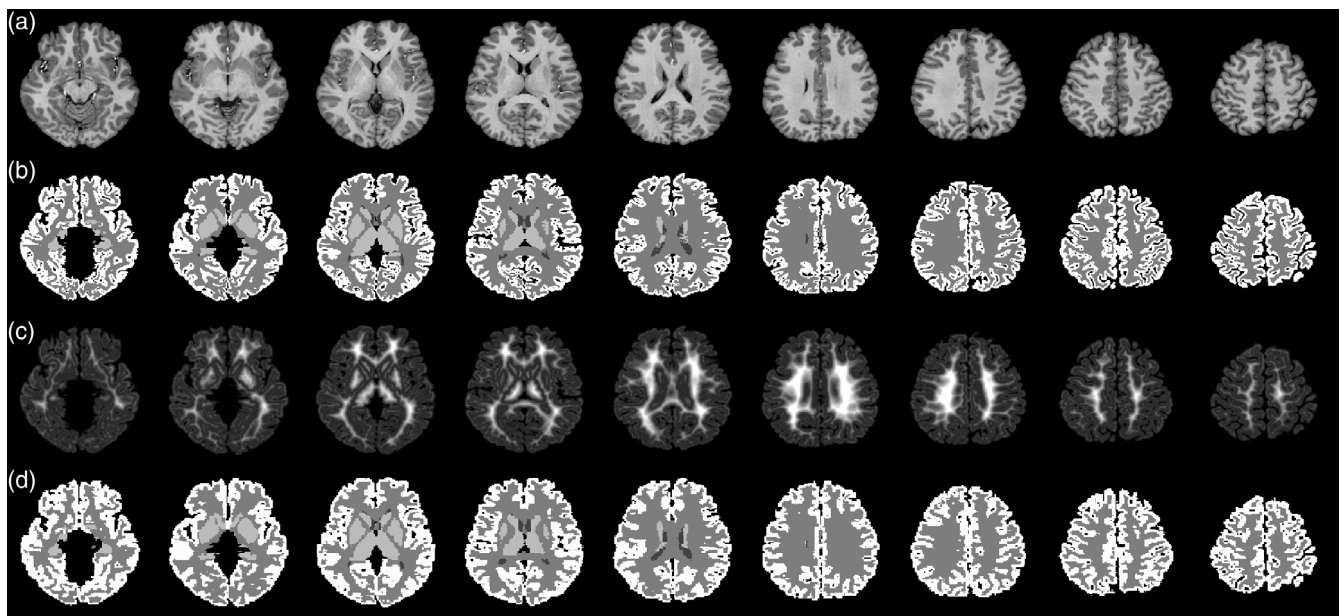


FIGURE 2 A representative case demonstrating the input and output of proposed CAE: Skull removed T1w volume (a); tissue segmentation of a (b); distance transform of b (c); CAE recovered b (d).

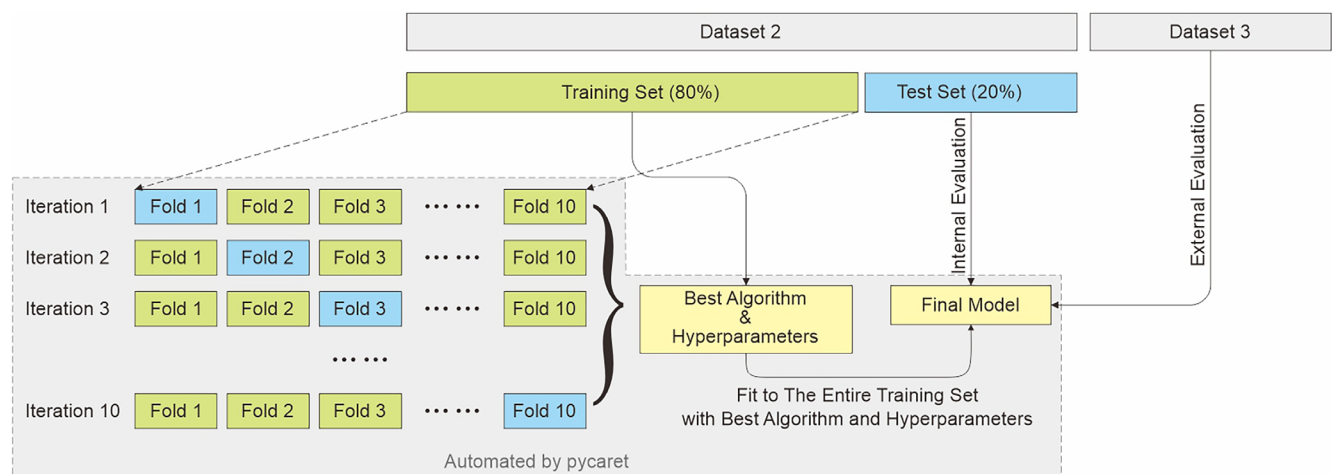


FIGURE 3 The workflow of training and evaluation the classification model via automated machine learning.

3 | RESULTS

3.1 | The quality of decoded volume from CAE

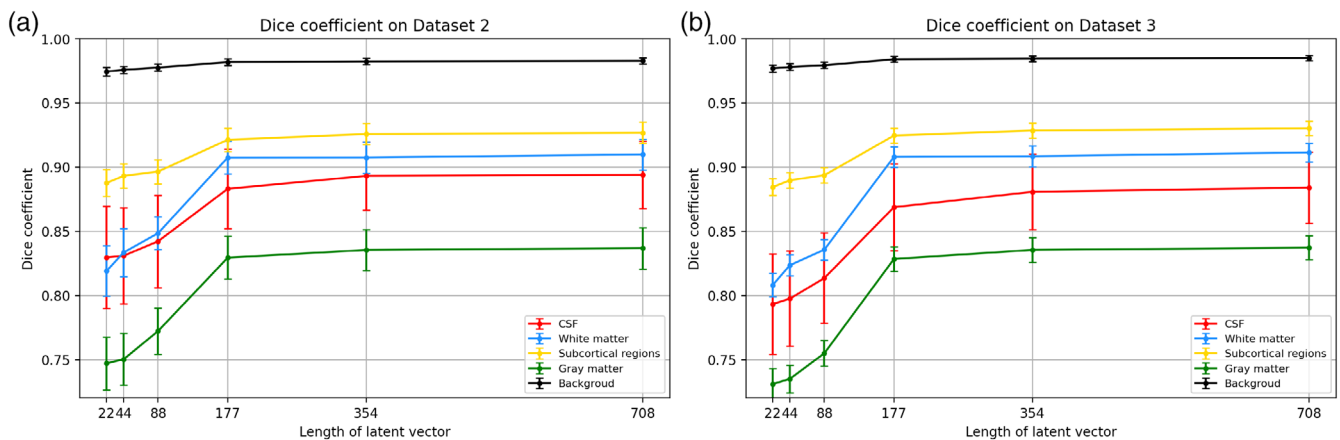
As the type of output volumes and loss functions of our proposed CAE are similar to those of semantic segmentation in deep neural networks, the DICE coefficient, which is commonly used in assessment the quality of image segmentation, was used to evaluate the quality of the volumes recovered from the CAE. Since the background component has the largest percentage, the background component unsurprisingly achieved the highest DICE score in dataset 2 and 3. Within the cerebrum, the DICE scores for each

component were subcortical nuclei, white matter, cerebrospinal fluid, and gray matter in descending order (Table 2). Visually, the subcortical nuclei and cerebrospinal fluid obtained relatively good recovery due to their simpler shapes. The boundaries between the white matter and gray matter were also well recovered. However, some boundaries of the pial surface lost certain details after decoding, but the gross shape of the sulcus and fundus were still well preserved.

There is an approximate exponential relationship between the length of latent vector and DICE score. When the length was set to 177, the DICE score of each component is almost at a plateau (Figure 4).

TABLE 2 The Dice coefficient of five decoded component in dataset 2 and 3.

Dataset (length of latent vector)	Dice coefficient				
	Background	CSF	White matter	Subcortical regions	Gray matter
2 (length $z = 22$)	0.97 ± 0.0032	0.82 ± 0.040	0.82 ± 0.020	0.89 ± 0.0105	0.75 ± 0.021
2 (length $z = 44$)	0.98 ± 0.0027	0.83 ± 0.037	0.83 ± 0.019	0.89 ± 0.0095	0.75 ± 0.020
2 (length $z = 88$)	0.98 ± 0.0029	0.84 ± 0.036	0.85 ± 0.013	0.90 ± 0.0095	0.77 ± 0.018
2 (length $z = 177$)	0.98 ± 0.0025	0.88 ± 0.031	0.91 ± 0.013	0.92 ± 0.0091	0.83 ± 0.017
2 (length $z = 354$)	0.98 ± 0.0024	0.89 ± 0.026	0.91 ± 0.012	0.93 ± 0.0082	0.84 ± 0.016
2 (length $z = 708$)	0.98 ± 0.0024	0.89 ± 0.026	0.91 ± 0.012	0.93 ± 0.0083	0.84 ± 0.016
3 (length $z = 22$)	0.98 ± 0.0027	0.79 ± 0.039	0.81 ± 0.0090	0.88 ± 0.0064	0.73 ± 0.012
3 (length $z = 44$)	0.98 ± 0.0027	0.80 ± 0.037	0.82 ± 0.0083	0.89 ± 0.0061	0.74 ± 0.011
3 (length $z = 88$)	0.98 ± 0.0023	0.81 ± 0.035	0.84 ± 0.0081	0.89 ± 0.0060	0.75 ± 0.010
3 (length $z = 177$)	0.98 ± 0.0023	0.87 ± 0.034	0.91 ± 0.0080	0.92 ± 0.0058	0.83 ± 0.010
3 (length $z = 354$)	0.98 ± 0.0022	0.88 ± 0.029	0.91 ± 0.0081	0.93 ± 0.0061	0.84 ± 0.0096
3 (length $z = 708$)	0.99 ± 0.0021	0.88 ± 0.028	0.91 ± 0.0072	0.93 ± 0.0056	0.84 ± 0.0093

**FIGURE 4** The Dice coefficient of images recovered from different lengths of latent vector on dataset 2 (a) and dataset 3 (b).

3.2 | Machine learning with classical features from Freesurfer

The AutoML pipeline identified the gradient boosting classifier (GBC) as the best model algorithm for the training dataset. After fine-tuning, an average accuracy of $62.61\% \pm 8.20\%$ was achieved in 10-fold cross-validation. The hyperparameters were then fixed and a final model was trained on the entire training dataset. The final model achieved an accuracy of 64.06% (95%CI: 51.10% – 75.68%), specificity of 60.61% , sensitivity of 67.74% and AUC of 0.68 on the internal test dataset and an accuracy of 60.00% (95%CI: 51.22% – 68.33%), specificity of 52.31% , sensitivity of 67.14% and AUC of 0.63 on the external test dataset.

The top 10 features significant contributed to the classification are listed in Table 3 (also plotted in Figure 6a).

3.3 | Machine learning with morphological fingerprint from CAE

The AutoML pipeline popped up extraTrees (Geurts et al., 2006) as the best model algorithm for the training dataset. After fine-tuning, an average accuracy of $72.10\% \pm 9.10\%$ was achieved in 10-fold cross-validation. The hyperparameters were then fixed and a final model was trained on the entire training dataset.

The final model achieved an accuracy of 73.44% (95%CI: 60.91% – 83.70%), specificity of 63.64% , sensitivity of 83.87% and AUC of 0.80 on the internal test dataset and an accuracy of 71.85% (95%CI: 63.47% – 79.25%), specificity of 63.08% , sensitivity of 80.00% and AUC of 0.77 on the external test dataset. The ROC curve comparison of two classifier on both internal and external test dataset is presented in Figure 5.

TABLE 3 The top 10 features significant contributed to the classification from the optimal model build on FS features.

Hemisphere	Region	Measure	Relative importance score
Left	Inferior temporal cortex	Curvature index	6.49778
Left	Pars opercularis cortex	Gauss curvature	6.07964
Left	Caudal anterior cingulate cortex	Curvature index	4.36896
Left	Superior temporal cortex	Gauss curvature	4.29484
Right	Entorhinal cortex	Thickness	4.22146
Right	Frontal pole cortex	Curvature index	3.85126
Right	Pars opercularis cortex	Curvature index	3.64841
Left	Inferior parietal cortex	Folding index	3.63153
Right	Fusiform cortex	Curvature index	3.48298
Right	Caudal middle frontal cortex	Folding index	3.47759

Although the features from CAE cannot directly correspond to the anatomical structure, we still can still infer the correspondence between CAE features and neuroanatomy from the decoded volume. We set the first 10 features that contribute significantly to the classification in the feature vector of length 354 to 0, fed them into the decoder to get the recovered images, subtracted them from the images recovered from the untouched features in order to get the difference images, and then superimposed the difference images onto the brain surface (Figure 6b and the workflow is illustrated in Figure S3). Differential areas (different between groups) were found throughout the brain, but with a tendency to concentrate in orbito-frontal cortex and temporal lobe.

3.4 | Effect of fingerprint length on discriminative performance

When doubling the length of latent vector to 708, extraTrees was the best classification algorithm. An average accuracy of $69.98\% \pm 9.39\%$ were achieved in training phase, and an accuracy of 75.00% (95%CI: 62.60%–84.98%), specificity of 69.70%, sensitivity of 80.65% and AUC of 0.80 on the internal test dataset and an accuracy of 68.15% (95%CI: 59.58%–75.90%), specificity of 66.15%, sensitivity of 70.00% and AUC of 0.74 on the external test dataset.

When halving the length of latent vector to 177, extraTrees was still the best classification algorithm. An average accuracy of $70.53\% \pm 8.04\%$ were achieved in training phase, and an accuracy of 71.88% (95%CI: 59.24%–82.40%), specificity of 69.70%, sensitivity of 74.19% and AUC of 0.80 on the internal test dataset and an accuracy of 68.89% (95%CI: 60.36%–76.57%), specificity of 67.69%, sensitivity of 70.00% and AUC of 0.72 on the external test dataset.

The performance of classifiers trained on shorter feature vectors are summarized in Table 4 and Figure 5. A positive correlation can be observed between length of feature vector and the performance of classifier trained on it, especially when the length is below 177.

4 | DISCUSSION

Deep learning is very good at discovering complex patterns of features in data, and has made great progress in the field of image recognition, with 90.88% accuracy already achieved on ImageNet (Deng et al., 2009). Features used are distinct from expert-designed features, as they are learned automatically from the data. Such characteristics make deep learning very promising for image-based diagnosis in mental disorders, as changes in the brains of patients with mental disorders tend to be modest, widely distributed and to have a complex correlational structure. In the current study, we utilized a novel CAE based method for brain morphological feature extraction. The volume of segmented brain was compressed by the CAE to approximately 0.03% of original size to form the morphological fingerprint. As cortical thickness, volume and gyrification index are all derived from the segmented brain, the morphological fingerprint covers all these features implicitly. With the same feature dimension, the classifier trained with the features extracted by this method outperformed the classifier trained with traditional features by about 10 percentage points in discriminating first-episode schizophrenia patients from healthy controls.

Although promising, applications of autoencoder approaches to neuroimaging data obtained from patients with psychiatric disorders are still rare (Chen et al., 2020). Pinaya et al. used a deep autoencoders approach to identify abnormal brain structural patterns in neuropsychiatric disorders. However, they used classical features, such as cortical thickness, as the input of autoencoder, which does not fully exploit the ability of deep learning to learn features automatically from images (Pinaya et al., 2019). Vyškovský et al. applied a stacked autoencoder classifier to a group of voxels selected by group-wise statistics to discriminate schizophrenic patients from normal controls (Vyškovský et al., 2022). The use of univariate feature selection ignores the spatial relationships between voxels and also does not take full advantage of the representation learning capabilities of deep learning. Yamaguchi et al. used the image as an input to the autoencoder, however, only gray matter was considered. In addition, the input images, which were preprocessed with smoothing, lost most of the

TABLE 4 Summary of performances of models with different feature source and dimension.

Feature source	Feature dimension	Best algorithm	Internal test				External test				
			Training accuracy	Accuracy (95% CI)	Sensitivity	Specificity	AUC	Accuracy(95% CI)	Sensitivity	Specificity	AUC
FS	354	Gradient boosted classifier	62.61% ± 8.20%	64.06% (51.10%–75.68%)	67.74%	60.61%	0.68	60.00% (51.22%–68.33%)	67.14%	52.31%	0.63
CAE	22	Decision tree	61.17% ± 9.04%	60.94% (47.93%–72.90%)	54.84%	66.67%	0.61	53.33% (44.56%–61.96%)	62.86%	43.08%	0.56
CAE	44	Decision tree	64.24% ± 10.11%	62.50% (49.51%–74.30%)	64.52%	60.61%	0.65	62.22% (53.48%–70.42%)	68.57%	55.38%	0.67
CAE	88	Extra trees	65.53% ± 9.41%	70.31% (57.58%–81.09%)	61.29%	78.79%	0.74	65.93% (57.28%–73.86%)	57.14%	75.38%	0.71
CAE	177	Extra trees	70.53% ± 8.04%	71.88% (59.24%–82.40%)	74.19%	69.70%	0.80	68.89% (60.36%–76.57%)	70.00%	67.69%	0.72
CAE	354	Extra trees	72.10% ± 9.10%	73.44% (60.91%–83.70%)	83.87%	63.64%	0.80	71.85% (63.47%–79.25%)	80.00%	63.08%	0.77
CAE	708	Extra trees	69.98% ± 9.39%	75.00% (62.60%–84.98%)	80.65%	69.70%	0.80	68.15% (59.58%–75.90%)	70.00%	66.15%	0.74

Note: Shade represents maxima in each metric.

key morphological information, such as local curvature and folding index (Yamaguchi et al., 2021).

In contrast, five main components were simultaneously involved in our current study. Even so, the amount of foreground and background voxels is still highly unbalanced, especially for subcortical nuclei and cerebrospinal fluid. The focal loss, which lessens the contribution of background voxels and highlights foreground voxels, was employed as the base loss function to prevent the loss function from being dominating by enormous background voxels during training (Lin et al., 2020). In addition, it is of utmost importance for our purposes to accurately recover the boundaries of each tissue type. In conventional U-net network, skip connections was used in the encoder-decoder architecture to preserve fine-grained details in the prediction. However, in our task, the use of skip connection results in a lack of detail in the encoded feature vector. In order to encode all the information into the feature vector while preserving as much detail as possible, we discarded the skip connection and added the HD loss item to the loss function. HD loss used the distance map to generate a gradient to make the boundary stops at the interface of two tissue (Karimi & Salcudean, 2020).

It is worth noting that even when we employed the AutoML to seek the optimal classification algorithm and hyperparameters, but accuracies achieved, whether based on FS features or CAE features, are lower than most previous classification studies. Clinical heterogeneity in disease stage (ours were early course patients) or medication (our samples were medication naïve) may be an influencing factor. Most previous studies recruited later course patients and treated patients, both of which features have been associated with increased case-control differences. In particular, while there may be widely distributed accelerated aging effects in long term ill individuals (Zhang et al., 2015), antipsychotic effects may lead to focal change such as in striatum that may be used to increase case-control differentiation (Li et al., 2018).

In addition, some prior studies had the problem called “double dipping” due to incorrect use of feature selection (Arbabshirani et al., 2017; Ball et al., 2020). Since in most prospective neuroimaging studies, the sample size of patients is often less than the number of features. Thus, feature selection allows a more specific focus on the relevant features and helps to improve the accuracy of the classification. Caution should be taken that feature selection can only be performed on the training set, otherwise information leakage would happen. However, some studies performed feature selection before training/test data split, making information from the test data leak into the model training process, which violates the need for complete data separation between model training and test data sets. Using error in analysis can lead to an overestimation of classification performance by about 15%–30%, depending on the number of involved study participants and features (Ball et al., 2020). Double dipping can be easily avoided by skipping feature selection. In current study, the best models that best fit the data are all tree-based ensemble models, which have built-in feature selection capabilities (Guan et al., 2014), making an external feature selection step not necessary.

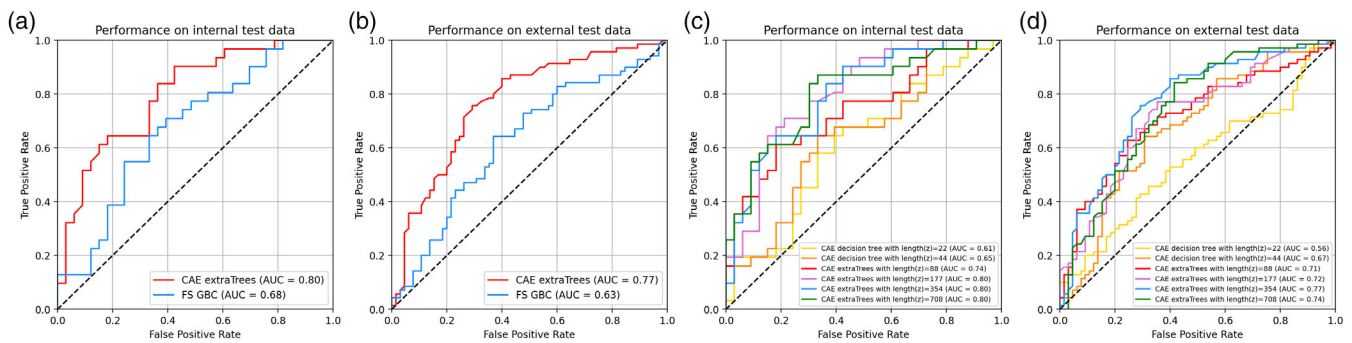


FIGURE 5 ROC curve of models built on autoencoded features (indicated by CAE) and traditional anatomical features (indicated by FS) on internal (a) and external (b) validation dataset; ROC curve of autoencoded features based models with different feature lengths on internal (c) and external (d) validation dataset.

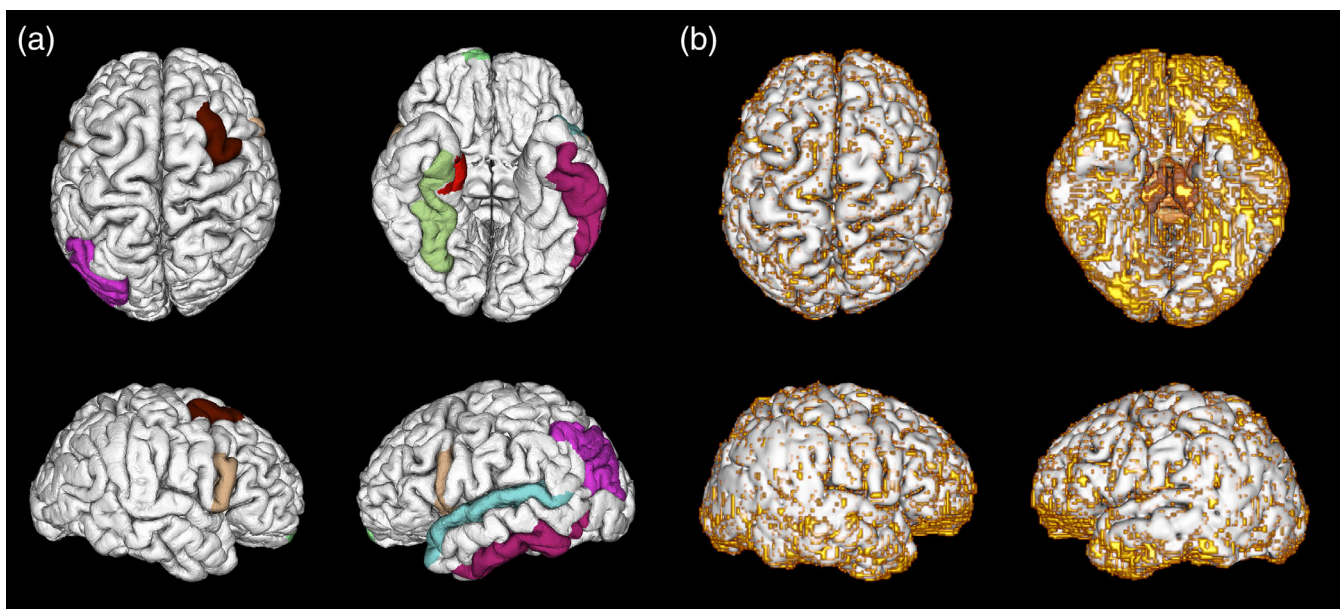


FIGURE 6 Significant brain regions contribution to classification derived from traditional anatomical features based model (a) and autoencoded features based model (b). The colored regions in subplot a indicate that at least one morphological metric in this region contributes significantly to the classification. The highlight areas in subplot b are the significant regions derived from autoencoded features based model, which is widespread throughout the brain, but with a tendency to concentrate in orbitofrontal cortex and temporal lobe.

External validation is crucial before deploying predictive/classification models in clinical practice (Kelly et al., 2019). However, the results of previous studies were mostly from the same cohort, or even just from cross-validation without setting up separate validation datasets, which rendered findings vulnerable to the risk of overfitting (Arbabshirani et al., 2017). In the current study, we utilized both internal and external validation sets. The external validation set differed from the training data in terms of population, imaging device and acquisition parameters, providing a relatively conservative approach to external validation. The performance on the external validation set is slightly lower than that on the internal validation set, whether based on FS features or CAE features. Since differences in imaging resolution have been standardized in the process of spatial normalization and structural images are relatively insensitive to differences in

imaging equipment, possible explanations may be the heterogeneity of the patient population and patient sample characteristics.

Another noteworthy point is that most previous classification studies have incorporated only features such as cortical thickness and surface area, whereas our study found that the shape of the local cortex may be a more valuable discriminatory feature. Previous studies have also raised concern that the results of VBM studies may have misidentified differences in local cortical shape as changes in gray matter volume (Davatzikos, 2004). As cortical thickness is susceptible to age and antipsychotic drugs, there may be important advantages to examining brain shape relative to local volume features in studies of mental disorders. In this regard, the current study also provides a novel and flexible approach for brain shape representation.

There are several limitations to this study. All subjects involved in current study were assigned to a binary case-control label. This dichotomous classification may limit the practical application of the model in the clinical practice, as spectrum concept was introduced in DSM-5 (Tandon et al., 2013). Further, comparing individuals with other psychiatric disorders, such as bipolar disorder, is needed to evaluate the validity and generalization of the proposed methodology to address clinically important diagnostic questions. During the assessment of feature length on classification performance, an expected linear relationship was not observed between feature length and performance, therefore a systematic evaluation of the relationship among length of encoded features, the quality of recovered image and classification performance is needed.

5 | CONCLUSION

In summary, the current study used features automatically learned from brain tissue segmentation maps from MRI images to discriminate schizophrenia patients and healthy controls. The discriminative accuracy was improved over that achieved using conventional cortical features on the same dataset and with the same feature dimension, but still were at a level below that needed for clinical application as diagnostic markers. However, our analyses provide evidence that efforts to identify valid markers for case-control differentiation may benefit from our encoding based approach which provided superior performance.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (81974278, 81621003, 81761128023, 81820108018 and 82027808), National Key R&D Program of China (2022YFC200900), NIH/NIMH (R01MH112189-01) and Young Elite Scientists Sponsorship Program (YESS20160060) by China Association for Science and Technology.

CONFLICT OF INTEREST

Dr. Sweeney consults to VeraSci. Other authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The code for the proposed CAE and training the network was available on Github (<https://github.com/MAI-Lab-West-China-Hospital/Convolutional-autoencoder-for-labeled-volume>).

ORCID

Xiaoqi Huang  <https://orcid.org/0000-0001-8686-5010>

Qiyong Gong  <https://orcid.org/0000-0002-5912-4871>

REFERENCES

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145, 137–165.

- Ball, T. M., Squeglia, L. M., Tapert, S. F., & Paulus, M. P. (2020). Double dipping in machine learning: Problems and solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5, 261–263.
- Chen, L., Xia, C., & Sun, H. (2020). Recent advances of deep learning in psychiatric disorders. *Precision Clinical Medicine*, 3, 202–213.
- Davatzikos, C. (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23, 17–20.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.
- Du, C., Gao, S., Liu, Y., & Gao, B. (2019). Multi-focus image fusion using deep support value convolutional neural network. *Optik (Stuttgart)*, 176, 567–578.
- Ellison-Wright, I., Glahn, D. C., Laird, A. R., Thelen, S. M., & Bullmore, E. (2008). The anatomy of first-episode and Chronic schizophrenia: An anatomical likelihood estimation meta-analysis. *The American Journal of Psychiatry*, 165, 1015–1023.
- van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., Pearlson, G. D., Yao, N., Fukunaga, M., Hashimoto, R., Okada, N., Yamamori, H., Bustillo, J. R., Clark, V. P., Agartz, I., Mueller, B. A., Cahn, W., de Zwarte, S. M. C., Hulshoff Pol, H. E., ... Orhan, F. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuroimaging genetics through meta analysis (ENIGMA) consortium. *Biological Psychiatry*, 84, 644–654.
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62, 774–781.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63, 3–42.
- Guan, D., Yuan, W., Lee, Y.-K., Najeebullah, K., & Rasel, M. K. (2014). A review of ensemble learning based feature selection. *IETE Technical Review*, 31, 190–198.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62, 782–790.
- Johnstone, E., Frith, C. D., Crow, T. J., Husband, J., & Kreef, L. (1976). Cerebral ventricular size and cognitive impairment in CHRONIC schizophrenia. *Lancet*, 308, 924–926.
- Karimi, D., & Salcudean, S. E. (2020). Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39, 499–513.
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 195.
- Kingma DP, Ba J (2014): Adam: A method for stochastic optimization.
- Li, W., Li, K., Guan, P., Chen, Y., Xiao, Y., Lui, S., Sweeney, J. A., & Gong, Q. (2018). Volume alteration of hippocampal subfields in first-episode antipsychotic-naïve schizophrenia patients before and after acute antipsychotic treatment. *NeuroImage: Clinical*, 20, 169–176.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 318–327.
- Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., & Martel, A. L. (2021). Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71, 102035.
- Matsuda, Y., & Ohi, K. (2018). Cortical gyrification in schizophrenia: Current perspectives. *Neuropsychiatric Disease and Treatment*, 14, 1861–1869.
- Mota, B., & Herculano-Houzel, S. (2012). How the cortex gets its folds: An inside-out, connectivity-driven model for the scaling of mammalian cortical folding. *Frontiers in Neuroanatomy*, 6, 3.
- Pinaya, W. H. L. L., Mechelli, A., & Sato, J. R. (2019). Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. *Human Brain Mapping*, 40, 944–954.
- Radua, J., Borgwardt, S., Crescini, A., Mataix-Cols, D., Meyer-Lindenberg, A., McGuire, P. K., & Fusar-Poli, P. (2012). Multimodal

- meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication. *Neuroscience and Biobehavioral Reviews*, 36, 2325–2333.
- Schaer, M., Cuadra, M. B., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J.-P. (2008). A surface-based approach to quantify local cortical Gyrification. *IEEE Transactions on Medical Imaging*, 27, 161–170.
- Spalthoff, R., Gaser, C., & Nenadić, I. (2018). Altered gyrification in schizophrenia and its relation to other morphometric markers. *Schizophrenia Research*, 202, 195–202.
- Sugihara, G., Oishi, N., Son, S., Kubota, M., Takahashi, H., & Murai, T. (2017). Distinct patterns of cerebral cortical thinning in schizophrenia: A neuroimaging data-driven approach. *Schizophrenia Bulletin*, 43, 900–906.
- Takayanagi, Y., Sasabayashi, D., Takahashi, T., Furuichi, A., Kido, M., Nishikawa, Y., Nakamura, M., Noguchi, K., & Suzuki, M. (2020). Reduced cortical thickness in schizophrenia and schizotypal disorder. *Schizophrenia Bulletin*, 46, 387–394.
- Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., Malaspina, D., Owen, M. J., Schultz, S., Tsuang, M., Van Os, J., & Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia Research*, 150, 3–10.
- Vyškovský, R., Schwarz, D., Churová, V., & Kašpárek, T. (2022). Structural MRI-based schizophrenia classification using autoencoders and 3D convolutional neural networks in combination with various pre-processing techniques. *Brain Sciences*, 12, 615.
- Winterburn, J. L., Voineskos, A. N., Devenyi, G. A., Plitman, E., de la Fuente-Sandoval, C., Bhagwat, N., Graff-Guerrero, A., Knight, J., & Chakravarty, M. M. (2019). Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. *Schizophrenia Research*, 214, 3–10.
- Yamaguchi, H., Hashimoto, Y., Sugihara, G., Miyata, J., Murai, T., Takahashi, H., Honda, M., Hishimoto, A., & Yamashita, Y. (2021). Three-dimensional convolutional autoencoder extracts features of structural brain images with a “diagnostic label-free” approach: Application to schizophrenia datasets. *Frontiers in Neuroscience*, 15, 652987.
- Yan, J., Cui, Y., Li, Q., Tian, L., Liu, B., Jiang, T., Zhang, D., & Yan, H. (2019). Cortical thinning and flattening in schizophrenia and their unaffected parents. *Neuropsychiatric Disease and Treatment*, 15, 935–946.
- Zhang, W., Deng, W., Yao, L., Xiao, Y., Li, F., Liu, J., Sweeney, J. A., Lui, S., & Gong, Q. (2015). Brain structural abnormalities in a Group of Never-Medicated Patients with Long-Term Schizophrenia. *The American Journal of Psychiatry*, 172, 995–1003.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sun, H., Luo, G., Lui, S., Huang, X., Sweeney, J., & Gong, Q. (2023). Morphological fingerprinting: Identifying patients with first-episode schizophrenia using auto-encoded morphological patterns. *Human Brain Mapping*, 44(2), 779–789. <https://doi.org/10.1002/hbm.26098>