Schizophrenia Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Structural and diffusion MRI based schizophrenia classification using 2D pretrained and 3D naive Convolutional Neural Networks

Mengjiao Hu [a,b], Xing Qian [b], Siwei Liu [b], Amelia Jialing Koh [b], Kang Sim [c,d], Xudong Jiang [e,1], Cuntai Guan [f,1], Juan Helen Zhou [b,g,h,i,*,1]

[a] NTU Institute for Health Technologies, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore, Singapore
[b] Center for Sleep and Cognition, Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[c] West Region, Institute of Mental Health (IMH), Singapore, Singapore
[d] Department of Research, Institute of Mental Health (IMH), Singapore, Singapore
[e] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore, Singapore
[f] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore
[g] Center for Translational Magnetic Resonance Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[h] Neuroscience and Behavioural Disorders Program, Duke-NUS Medical School, Singapore, Singapore
[i] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore, Singapore

ARTICLE INFO

Keywords:
Schizophrenia
MRI
Classification
Deep learning
Convolutional Neural Networks
Transfer learning

ABSTRACT

The ability of automatic feature learning makes Convolutional Neural Network (CNN) potentially suitable to uncover the complex and widespread brain changes in schizophrenia. Despite that, limited studies have been done on schizophrenia identification using interpretable deep learning approaches on multimodal neuroimaging data. Here, we developed a deep feature approach based on pre-trained 2D CNN and naive 3D CNN models trained from scratch for schizophrenia classification by integrating 3D structural and diffusion magnetic resonance imaging (MRI) data. We found that the naive 3D CNN models outperformed the pretrained 2D CNN models and the handcrafted feature-based machine learning approach using support vector machine during both cross-validation and testing on an independent dataset. Multimodal neuroimaging-based models accomplished performance superior to models based on a single modality. Furthermore, we identified brain grey matter and white matter regions critical for illness classification at the individual- and group-level which supported the salience network and striatal dysfunction hypotheses in schizophrenia. Our findings underscore the potential of CNN not only to automatically uncover and integrate multimodal 3D brain imaging features for schizophrenia identification, but also to provide relevant neurobiological interpretations which are crucial for developing objective and interpretable imaging-based probes for prognosis and diagnosis in psychiatric disorders.

1. Introduction

Schizophrenia is a potentially severe and chronic mental disorder that imposes great burdens on patients, their families and society. Early and accurate diagnosis of schizophrenia could facilitate treatment planning and improve the outcome of the illness. However, current reliance on clinical interviews and corroboration of clinical data make diagnosis of schizophrenia challenging due to the complex and heterogeneous symptom presentations which can vary with the course of the illness, especially at early onset (Del Barrio, 2016; Fanous et al., 2012; Kennedy et al., 2014). Therefore, it is important to establish an objective approach which yields an accurate diagnosis and leads to an appropriate treatment in turn.

Structural magnetic resonance imaging (sMRI) and diffusion MRI (dMRI) have been used to detect brain structural and microstructural abnormalities in patients with schizophrenia (Gong et al., 2019; Mitelman, 2019; Ott et al., 2019; Power et al., 2016; Shenton et al., 2001). Accumulating sMRI evidence suggests widespread grey matter reductions, especially in frontal, temporal, thalamic and striatal regions (Fornito et al., 2009; Haijma et al., 2013; Koelkebeck et al., 2019; Kuo and Pogue-Geile, 2019), cortical thinning in frontal, temporal, cingulate and insular regions (Takayanagi et al., 2020; Van Erp et al., 2018; Yan

* Corresponding author at: Tahir Foundation Building (MD1), 12 Science Drive 2, #13-05C, National University of Singapore, Singapore 117549, Singapore.
E-mail address: helen.zhou@nus.edu.sg (J.H. Zhou).
[1] Joint Senior Authors.

https://doi.org/10.1016/j.schres.2021.06.011
Received 22 December 2020; Received in revised form 11 May 2021; Accepted 18 June 2021

et al., 2019) and enlargement of ventricles (Gaser et al., 2004; Kuo and Pogue-Geile, 2019; Zheng et al., 2019) in schizophrenia patients compared to healthy controls. In parallel, dMRI or diffusion tensor imaging studies reported lower fractional anisotropy (FA) and higher mean diffusivity (MD) in multiple white matter tracts implicating frontal-striatal-thalamic circuits and the cingulum in schizophrenia and its prodrome (Di Biase et al., 2020; Kelly et al., 2018; Wang et al., 2016, 2019). However, the heterogeneity in both the effect sizes and the regional distribution of the brain alterations reported across studies have prevented the application of group-level findings to individual-level diagnosis (Arbabshirani et al., 2017; Di Biase et al., 2020; Kelly et al., 2018). Given the high-dimensional multimodal neuroimaging and clinical data, there is an increasing need to develop automatic and standardized multimodal probes for accurate and objective classification of schizophrenia.

Machine learning approaches have begun to demonstrate its potential in complementing the clinical diagnosis of psychiatric disorders using neuroimaging data (Arbabshirani et al., 2017; Rashid and Calhoun, 2020). Previous work on computer-aided classification of schizophrenia patients and healthy controls mainly focused on handcrafted feature-based machine learning approach, which requires feature extraction and reduction before classification. The most common features in previous studies are cortical thickness and voxel-based morphometry (via sMRI) (Chin et al., 2018; Nieuwenhuis et al., 2012; Salvador et al., 2017; Winterburn et al., 2019) as well as white matter microstructure such as FA (via dMRI) (Ardekani et al., 2011; Liang et al., 2019; Mikolas et al., 2018). Traditional classifiers like support vector machine (SVM), random forest and logistic regression are the most widely used methods (Arbabshirani et al., 2017; De Filippis et al., 2019; Winterburn et al., 2019). However, the subtle, mixed, and widespread brain anatomical changes in schizophrenia limit the performance of such handcrafted machine learning approaches. Pre-selected features might not be efficient and generalizable across different cohorts with differing duration of illness thus causing the accuracy to vary across datasets, features and hyperparameter settings. Handcrafted feature-based machine learning also has difficulty in uncovering new features to facilitate biological inferences (Arbabshirani et al., 2017; Rashid and Calhoun, 2020; Winterburn et al., 2019).

Convolutional Neural Network (CNN) has recently become a promising approach for medical image classification such as brain tumor, lung nodule and Alzheimer's disease (Iizuka et al., 2019; Khvostikov et al., 2018; Lin et al., 2018; Liu and Kang, 2017). As a data-driven deep learning method, CNN is capable of automatic feature learning which mitigates the subjectivity and variability in pre-selecting relevant features. This is especially important for psychiatric disorders like schizophrenia which has subtle, complex and widely distributed brain alterations (Lee et al., 2017). Deep model architecture with nonlinear layers also allows efficient mapping of complicated data patterns (Arbabshirani et al., 2017; Lee et al., 2017). Nevertheless, limited studies have applied CNN on structural neuroimaging data to differentiate patients with schizophrenia patients from healthy controls. A recent study applied a sequential 3D CNN model on sMRI data for classification and achieved an area under the receiver operating characteristic curve (ROC-AUC) of 0.96, but independent testing performance degraded significantly (Oh et al., 2020). Complicated 3D CNN architectures have not been investigated though. Studies using transfer learning approaches that utilize powerful pretrained 2D CNN networks to extract features from deep layers are also lacking. Of note, multimodal feature extraction from different neuroimaging modalities is critical for understanding the neural substrates underlying schizophrenia from complementary perspectives, potentially leading to higher classification performance (Lei et al., 2020; Lerman-Sinkoff et al., 2019; Salvador et al., 2019). Although the automatic feature learning capability of CNN enables more prominent integration of multimodal inputs, multi-channel 2D and 3D CNN have not been employed and evaluated for schizophrenia discrimination.

CNN models have demonstrated remarkable performance on image classification tasks but are often criticized as a "black box" as the learning process and predictions are not interpretable (Pinaya et al., 2019). Gradient-based methods and up-convolutional net have been proposed to visualize 2D CNN representations and provide visual explanations for decision making (Mahendran and Vedaldi, 2015; Selvaraju et al., 2017; Simonyan et al., 2014; Springenberg et al., 2015; Zeiler and Fergus, 2014; Zhang and Zhu, 2018; Zhou et al., 2016). As a state-of-the-art approach, gradient class activation map could localize the discriminative image regions from any CNN-based network without requiring architectural changes or re-training and could also be extended to 3D CNN (Selvaraju et al., 2017; Yang et al., 2018a). Identifying the critical regions for classification not only validates the underlying rationale of decision-making to enable clinical adoption but also facilitates biological inferences to improve our understanding of schizophrenia.

To fill these gaps, we developed naive 3D CNN models trained from scratch and a deep feature approach using pretrained 2D CNN networks to identify patients with schizophrenia using 3D sMRI and dMRI data. A multi-channel input approach was utilized to integrate representations from different feature maps and modalities. We implemented a state-of-the-art handcrafted feature-based machine learning approach with SVM as a benchmark. We hypothesized that both 2D and 3D CNN models would outperform handcrafted feature-based machine learning and multimodal neuroimaging-based models (i.e., integrating both structural and diffusion MRI) would have better performance than single modality-based models. Further, we aimed to identify the discriminative brain regions for classification of schizophrenia based on the best models using the gradient class activation map approach.

2. Methods

2.1. Participants

Two independent MRI datasets of schizophrenia and controls were used in this study which were comparable for age and gender for both groups (Table 1). The Northwestern University Schizophrenia Data and Software Tool (NUSDAST) is a repository of schizophrenia neuroimaging data collected from over 450 schizophrenia patients and healthy controls, which is publicly available on SchizConnect platform (Kogan et al., 2016; Wang et al., 2013). Overall, 141 schizophrenia patients and 134 healthy controls from this public dataset were included after quality control.

A similar dataset, with both structural MRI and diffusion MRI from the Institute of Mental Health (IMH), Singapore, was included as an independent dataset (Ho et al., 2017a, 2017b). In this dataset, 148 schizophrenia patients and 76 healthy controls were included after quality control.

2.2. Image acquisition

For the NUSDAST dataset, all MRI scans were collected using the same 1.5 T Vision scanner platform (Siemens Medical Systems).

Table 1
Subject demographics of the two datasets.

	NUSDAST		IMH	
	SZ	HC	SZ	HC
Subject number	141	134	148	76
Age (years), mean (SD)	35.06 (12.78)	32.88 (14.05)	32.72 (9.04)	31.33 (9.77)
Sex (male/female)	90/51	72/62	102/46	47/29
Modality	sMRI		sMRI & dMRI	

Abbreviations: SZ - schizophrenia patients, HC - healthy controls, SD - standard deviation.

Acquisition of all scans was performed at the Mallinckrodt Institute of Radiology at Washington University School of Medicine, where scanner stability (e.g., frequency, receiver gain, transmitter voltage, SNR) and artifacts were regularly monitored from 1998 to 2006 (Wang et al., 2013). 3D Turbo Flash images were acquired with following parameters: axial slice thickness = 1 mm, 180 slices, in-plane resolution = 1 × 1 mm^2, repetition time = 20 ms, echo time = 5.4 ms, flip angle = 7° and matrix size = 256 × 256 pixels.

For the IMH dataset, all MRI scans were performed using the same 3-Tesla whole-body scanner MRI (Philips Achieva, Best, The Netherlands) with an 8-channel SENSE (Sensitivity Encoding) head coil at the National Neuroscience Institute, Singapore, from 2006 to 2013 (Ho et al., 2017a, 2017b). T1-weighted magnetization-prepared rapid acquisition with gradient echo (MPRAGE) images were acquired with following parameters: axial slice thickness = 0.9 mm, 180 slices, in-plane resolution = 0.9 × 0.9 mm^2, repetition time = 7200 ms, echo time = 3.3 ms, flip angle = 8° and matrix size = 256 × 256 pixels. Diffusion MRI data were acquired with the following parameters: axial slice thickness = 3 mm, 42 slices, in-plane resolution = 0.9 × 0.9 mm^2, repetition time = 3.725 s, echo time = 56 ms, flip angle = 90°, acquisition matrix size = 112 × 109 pixels, reconstruction matrix = 256 × 256 pixels. 15 diffusion-weighted images (b = 800 s/mm^2) of non-parallel directions and 1 baseline image (b = 0 s/mm^2) were obtained. Three runs of such DW-MRI images were acquired in the same session and were concatenated for processing.

2.3. Image processing

Processing of the structural MRI data was performed using Computational Anatomy Toolbox in Statistical Parametric Mapping 12 (SPM12) for voxel-wise estimation of grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) compartment (Kurth et al., 2015) following our previous work (Ng et al., 2016). Images with motion artifacts were excluded after visual quality control. Subject-level probability maps were obtained from T1-weighted images with the following steps: (i) skull stripping; (ii) linear (FLIRT) and nonlinear (FNIRT) registration to the Montreal Neurological Institute (MNI) 152 standard space (Andersson et al., 2007); (iii) segmentation of the brain into GM, WM and CSF compartments with 1.5 mm isotropic resolution; (iv) modulation by multiplying voxel values with the linear and nonlinear component of the Jacobian determinant. The diffusion MRI data were preprocessed using FSL (http://www.fmrib.ox.ac.uk/fsl) with the following steps following our previous work (Ho et al., 2017b): (i) head movements and eddy current distortion correction with reference to the first b = 0 volume via affine registration of the diffusion-weighted images; (ii) diffusion gradients rotation to improve consistency with the motion parameters; (iii) visual inspection of signal dropout, artifacts and additional motion (subjects with >3 mm of motion displacement during the scan were excluded); (iv) tensor fitting to diffusion data at each voxel to create fractional anisotropy (FA) and mean diffusivity (MD) maps; (v) nonlinear registration to FMRIB58_FA standard space with FNIRT. For computational efficiency, all the resulting feature maps (GM, WM, CSF, FA and MD) in the standard space were downsampled from 1.5 × 1.5 × 1.5 mm^3 (121 × 145 × 121) to 3 × 3 × 3 mm^3 (61 × 121 × 61).

2.4. Study design

We employed three approaches to classify schizophrenia patients and healthy controls (see detailed study design in Fig. 1). To compare with the state-of-the-art handcrafted feature-based machine learning approach, we implemented linear and nonlinear SVM classifiers as the benchmark as SVM achieved the best performance in most previous studies for neuroimaging-based schizophrenia classification (Rozycki et al., 2018; Salvador et al., 2017; Winterburn et al., 2019). To utilize powerful pre-trained 2D CNN networks, we applied deep feature approach based on feature maps extracted from pre-trained networks in a 2D manner. To exploit the 3D contextual information and investigate the effect of model structure, we developed naive 3D CNN models which were trained from scratch with different architectures and depths.

GM, WM and CSF probability maps were used as inputs for structural MRI models trained on the NUSDAST dataset. Nested 5-fold cross-validation was used to select hyperparameters and obtain testing results for all the models. To test the generalizability of trained models, sMRI models that trained on the NUSDAST dataset were further tested

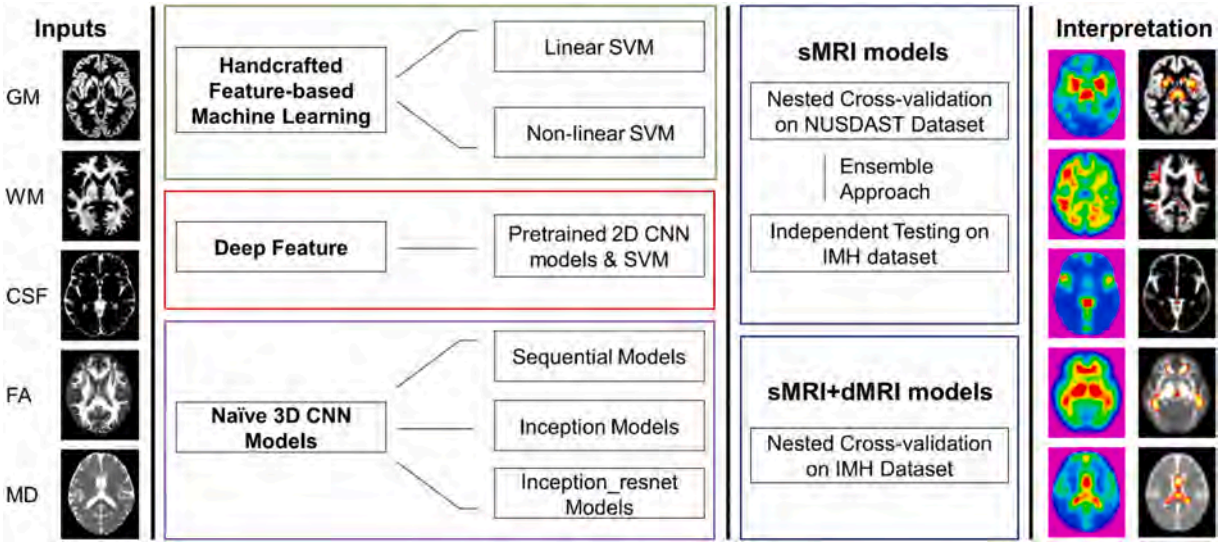

Fig. 1. Study design. We employed three approaches to classify schizophrenia patients and healthy controls. We implemented linear and nonlinear SVM classifiers as the benchmark methods. We applied deep feature approach which classified feature maps extracted from pretrained networks to utilize the powerful pretrained networks. Naive 3D CNN models trained from scratch with different architecture and depths were developed to exploit 3D contextual information. GM, WM and CSF probability maps were used as inputs for structural MRI models trained on the NUSDAST dataset. Nested 5-fold cross-validation was used to select hyperparameters and obtain testing results for all the models. To test the generalizability, sMRI models trained using the NUSDAST dataset were further tested on the independent IMH dataset with ensemble approach. We also developed multimodal models using GM, WM, CSF, FA and MD maps as inputs and evaluated on the IMH dataset using cross-validation. Lastly, we used gradient class activation map approach to interpret 3D CNN models with the best performance to identify brain regions whose GM, WM, CSF, or white matter microstructure contributed significantly to the classification of schizophrenia patients.

on the independent IMH dataset with an ensemble approach.

In addition, we developed multimodal models using GM, WM, CSF, FA and MD maps as inputs for all three approaches above. We performed cross-validation on the IMH dataset and compared the classification performance with single-modality models.

Lastly, we evaluated the interpretability of our derived 3D CNN models with the best performance with reference to previous literature. Gradient class activation map approach was employed to identify brain regions whose GM, WM, CSF or white matter microstructure contributed significantly to the classification of schizophrenia patients.

Further details of each step are given as follows.

2.4.1. Handcrafted feature-based machine learning

To make a comparison with the state-of-art approach using handcrafted feature-based machine learning, we employed voxel-based morphometry as handcrafted features and SVM as classifiers to implement a benchmark method. GM, WM and CSF probability maps extracted from sMRI as well as FA and MD maps extracted from dMRI were employed as features and flattened as feature vectors. Feature reduction was completed by principal components analysis (PCA) with 99% variance explained for each input map. Linear SVM and nonlinear SVM with the Radial Basis Function (RBF) kernel were trained and tested.

2.4.2. Deep feature based on pre-trained 2D CNN

To investigate the impact of model architecture, we employed six state-of-the-art CNN models pre-trained on ImageNet dataset: VGG16 (Simonyan and Zisserman, 2015), Xception (Chollet, 2017), Resnet101 (He et al., 2016), Densenet121 (Huang et al., 2017), Inception_v3 (Szegedy et al., 2015) and Inception_resnet_v2 (Szegedy et al., 2017). Each 2D slice of the 3D maps was used as input to pre-trained networks and the deep feature map taken from the last convolutional layer was used as the resulting extracted features. To fully utilize the contextual information of 3D images, the feature maps from each 2D slice of the three views (axial, coronal and sagittal) were combined and flattened as a large feature vector at the individual level. Feature reduction was performed with PCA with 99% variance explained for each input map and classification was completed with linear SVM.

2.4.3. Naive 3D CNN models

Typical CNN models consist of sequential convolutional layers, pooling layers and fully connected layers and they use backpropagation to learn multi-level features (Lee et al., 2017). The convolutional layer computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. The pooling layer performs a downsampling operation along the spatial dimensions. The fully connected layer connects each neuron to all the numbers in the previous volume and computes the class probabilities. Note that in 3D CNN networks, the convolutions and pooling operate in a cubic manner with 3D feature volumes instead of 2D feature maps thus fully utilizing the 3D contextual information in brain structural imaging. Furthermore, advanced architectures interconnect the layers and form modules with more complicated topologies, such as inception module and residual module (He et al., 2016; Szegedy et al., 2015). In this study, the following three types of 3D CNN model architectures with different depths were implemented and evaluated.

2.4.3.1. Sequential models. Sequential models followed the typical CNN sequential architecture with convolutional layers, pooling layers and fully connected layers. The convolutional kernel and pooling kernel were set with 3x3x3 dimensions using grid search. GM, WM and CSF probability maps extracted from sMRI along with FA and MD maps extracted from dMRI were employed as independent input maps connecting to different network branches. The resulting feature maps from each branch were flattened and connected to a fully connected layer with 128 neurons. Output was obtained by sigmoid function.

Three sequential models with different number of layers were trained and tested. As shown in Fig. 2A, Sequential_1 consists of one convolutional layer (Conv), one maxpooling layer (Maxpooling) and one fully connected layer (FC) thus giving rise to a Conv+Maxpooling+FC structure; Sequential_2 has a 2(Cov + Maxpooling) + FC structure as it consists of two levels of convolutional layer and maxpooling layer connecting to a fully connected layer; Sequential_3 has a 3(Conv+Maxpooling) + FC structure as it consists of 3 levels of convolutional layer and maxpooling layer connecting to a fully connected layer. The deeper the network, the smaller the feature volumes become.

2.4.3.2. Inception models. Inspired by the GoogLeNet (Szegedy et al., 2015), a 3D inception module was utilized in inception models (Fig. 2B and C). The inception module divides the network into multiple branches with different convolutional kernels thus allowing operating convolutions with different kernels on the same level. The inception module not only improves the performance of the network but also controls overfitting and reduces computational expenses.

2.4.3.3. Inception_resnet models. Inspired by the residual module (He et al., 2016), inception_resnet models combined inception architecture and residual module to utilize information from previous layers (Fig. 2D). Inception_resnet_1 model has the same arterial structure as Inception_1 model with an extra connection that adds up the output from the previous layer and output from the inception module. Similarly, Inception_resnet_2 model has two extra connections adding outputs from different layers together.

2.4.4. Nested cross-validation and ensemble approach

To avoid overfitting, we used nested cross-validation consisting of an inner loop and an outer loop. The outer loop split data into 5 folds and each round used 4 folds for training and the remaining for testing. The inner loop further split the training data from the outer loop into 5 folds and used cross-validation to select hyperparameters for 3D CNN models (optimizer, kernel size, kernel number) and nonlinear SVM (kernel type, kernel coefficient, regularization term) (Supplementary Table 1). Testing results were obtained through outer 5-fold cross-validation which averaged the test error over multiple train–test splits. The testing results for naive 3D CNN were reported as the average of 10 repeats to reduce randomness generated from the training process. Furthermore, a random seed was appointed for the data split of nested cross-validation to reduce randomness and ensure consistency of training and testing data among different approaches and different repeats. Training and testing were completed using a NVIDIA V100 TENSOR CORE GPU with batch size 5 and maximum epoch 150 (Supplementary Table 2).

After the nested cross-validation on the NUSDAST dataset, we took the ensemble approach to evaluate the classification performance on the independent testing IMH dataset. Specifically, 5-fold cross-validation trained 5 optimized models in each fold for handcrafted feature-based machine learning and deep feature based on pretrained 2D CNN networks. Independent testing results on the IMH dataset were determined by majority voting among 5 predictions from the 5 models trained for each method. For naive 3D CNN, we selected 5 models from one repeat of 5-fold cross-validation with the highest accuracy out of 10 repeats and the independent testing results on the IMH dataset were determined by majority voting among 5 selected models.

2.4.5. Interpretation

To interpret the 3D CNN models and classification decision process, we adopted the gradient class activation map approach on both single and multi-modality models with the best performance to localize the brain regions that contributed significantly to the classification of

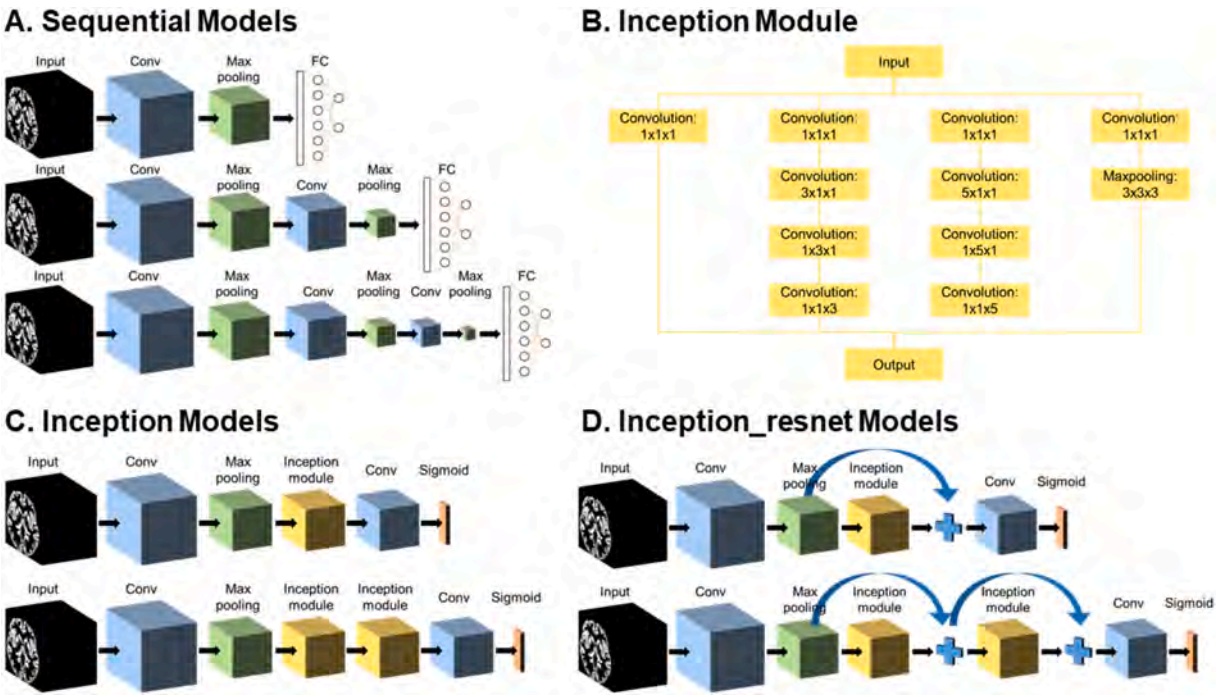

Fig. 2. Architectures of the naive 3D CNN models. A) Three sequential models with different numbers of layers were trained and tested. Sequential_1 model contains 1 convolutional layer (Conv), 1 maxpooling layer (Maxpooling) and 1 fully connected layer (FC); Sequential_2 model has two levels of convolutional layer and maxpooling layer connected to one final fully connected layer; Sequential_3 has structure 3(Conv+Maxpooling) + FC, representing 3 levels of convolutional layer and maxpooling layer connecting to fully connected layer. B) 3D inception module was utilized in inception models. The inception module divides the network into multiple branches with different convolutional kernels. C) Inception_1 model contains 1 convolutional layer and 1 maxpooling layer followed by an 3D inception module, the fully connected layer is replaced by a convolutional layer and sigmoid function. Inception_2 model contains 2 inception modules. D) Inception_resnent models employed the same stream structure as inception models. An extra connection adds the output from the previous layer and output from the inception module together. Inception_resnet_1 model has one inception module while inception_resnet_2 model has two inception modules.

schizophrenia patients. The discriminative regions were determined by calculating the backpropagated class signal at the last rectified convolutional layer of each input branch and thus heat maps were generated separately for GM, WM, CSF, FA and MD for each individual. Subsequently, one sample *t*-test (familywise error corrected at $p < 0.01$) identified the critical regions for classification at the group-level.

3. Results

3.1. Nested cross-validation results on the NUSDAST dataset with sMRI models

Accuracy, sensitivity, specificity and ROC-AUC from nested cross-validation (mean and standard deviation) were reported for sMRI models trained on the NUSDAST dataset (Table 2). In general, both 2D and 3D CNN models outperformed handcrafted feature-based machine learning approaches. Nonlinear SVM with the RBF kernel achieved an averaged accuracy of 70.22% and ROC-AUC of 0.78. Deep feature based on pretrained inception_v3 outperformed nonlinear SVM with an accuracy of 72.41% and ROC-AUC of 0.82. Naive 3D CNN models trained from scratch further exceeded the performance of deep feature based on pretrained 2D CNN networks with an accuracy of 79.27% and ROC-AUC of 0.81.

The classification performance of different models was further visualized in Figs. 3A and 4A. Fig. 3A directly compared the accuracy of different models in bar charts. Both 2D and 3D CNN models obtained higher accuracy than handcrafted feature-based machine learning. Naive 3D CNN model inception_resnet_1 achieved the highest accuracy. Fig. 4A further demonstrated model performance regarding sensitivity and specificity with prediction probability of patients plotted. Random predictions for both controls and patients had a distribution peak at around 0.5 while predictions generated by models had a distribution peak towards 0 for controls and 1 for patients. In concordance with accuracies and ROC-AUC reported in Table 2, handcrafted feature-based machine learning performed slightly better than random predictions. Deep feature based on pre-trained 2D CNN outperformed handcrafted feature-based machine learning. Naive 3D CNN models distinctly separated patients from controls with wider gaps between the prediction distribution peaks of the two groups.

Replication of the same cross-validation process was performed on the sMRI data of the IMH dataset and demonstrated similar observations (Supplementary Table 4 and Supplementary Fig. 1).

3.2. Independent testing results on the IMH dataset with sMRI models

To further test the generalizability of our models, accuracy, sensitivity, specificity and ROC-AUC on the independent testing IMH dataset were reported for sMRI models trained on the NUSDAST dataset (Table 3 and Fig. 3B). Naive 3D CNN inception_resnet_1 model achieved the highest accuracy of 70.98% and ROC-AUC of 0.75. Handcrafted feature-based machine learning had inferior performance with low accuracy. Deep feature based on pretrained 2D CNN and naive 3D CNN models with sequential architecture achieved higher accuracy but sensitivity and specificity were greatly imbalanced. Meanwhile, 3D CNN models with more complex architecture obtained higher accuracy as well as more balanced sensitivity and specificity.

Prediction distributions of the models corresponded well with the findings on accuracy, sensitivity, specificity, and ROC-AUC. Handcrafted feature-based machine learning approaches and deep feature based on pre-trained 2D CNN barely differed from random predictions. Sequential models categorized most subjects into one class and thus resulted in highly imbalanced specificity and sensitivity. On the other

Table 2
Single modal (sMRI) cross-validation results on the NUSDAST dataset.

		ACC	SP	SE	ROC-AUC
Handcrafted feature-based machine learning	Linear SVM	69.85% (±4.02%)	67.98% (±4.70%)	71.63% (±3.93%)	0.78 (±0.04)
	Nonlinear SVM (RBF kernel)	70.22% (±5.28%)	75.41% (±3.56%)	65.25% (±7.95%)	0.78 (±0.05)
Deep feature based on pretrained 2D CNN	VGG16	68.05% (±5.22%)	63.53% (±9.81%)	72.29% (±5.89%)	0.75 (±0.04)
	Xception	66.59% (±7.92%)	69.43% (±7.81%)	63.84% (±10.87%)	0.73 (±0.07)
	Resnet101	70.59% (±6.83%)	69.52% (±9.59%)	71.55% (±10.72%)	0.78 (±0.04)
	Densenet121	69.07% (±3.21%)	67.95% (±7.84%)	70.12% (±8.21%)	0.79 (±0.05)
	Inception_V3	72.41% (±4.70%)	70.94% (±8.45%)	73.74% (±5.86%)	0.82 (±0.05)
	Inception_resnet_V2	72.40% (±3.82%)	70.88% (±4.38%)	73.77% (±9.17%)	0.79 (±0.06)
Naive 3D CNN models	Sequential_1	77.78% (±3.56%)	80.32% (±6.98%)	75.35% (±8.08%)	0.82 (±0.04)
	Sequential_2	76.50% (±3.52%)	75.91% (±7.12%)	77.07% (±2.58%)	0.79 (±0.05)
	Sequential_3	73.91% (±2.17%)	73.13% (±7.13%)	74.57% (±4.48%)	0.77 (±0.04)
	Inception_1	77.71% (±4.05%)	76.90% (±6.62%)	78.42% (±5.36%)	0.81 (±0.04)
	Inception_2	76.24% (±3.52%)	78.08% (±4.47%)	74.42% (±7.40%)	0.79 (±0.05)
	Inception_resnet_1	**79.27% (±3.92%)**	80.44% (±5.96%)	78.15% (±4.12%)	0.81 (±0.05)
	Inception_resnet_2	78.76% (±3.70%)	81.54% (±5.12%)	76.10% (±4.55%)	0.81 (±0.05)

The results of outer cross-validation of all three approaches are listed with mean (+/−standard deviation). Both deep features based on pretrained 2D CNN models and naive 3D CNN models obtained higher accuracy than handcrafted feature-based machine learning. Naive 3D CNN models obtained higher accuracy than deep feature 2D CNN models. The highest accuracy obtained within each approach is highlighted in bold. The highest accuracy is obtained by Inception_resnet_1 across all approaches. Abbreviations: ACC – accuracy, SP – specificity, SE – sensitivity, ROC-AUC - area under the receiver operating characteristic curve.

hand, naive 3D CNN inception models and inception_resnet models satisfactorily separated patients from controls with prediction distribution peaks apart from each other, and thus achieved better classification results with more balanced sensitivity and specificity.

Subsequent testing on the NUSDAST dataset based on models trained with the IMH dataset yielded similar conclusions (Supplementary Table 5 and Supplementary Fig. 1).

3.3. Nested cross-validation results on the IMH dataset with sMRI+dMRI models

To evaluate whether the integration of multimodal images could further improve classification results, we built up models using all three approaches with a combination of sMRI and dMRI maps. Accuracy, sensitivity, specificity and ROC-AUC trained on the IMH dataset were reported for sMRI+dMRI models (Table 4 and Fig. 3C). Overall, multimodal sMRI+dMRI models outperformed single-modal sMRI models. Comparing with the cross-validation results on sMRI models (Supplementary Table 4), multimodal sMRI+dMRI models achieved superior performance for all three approaches, especially for specificity. Multimodal inception_resnet_1 model obtained the highest accuracy of 81.02%, ROC-AUC of 0.84 and 5.31% improvement on specificity when compared with the sMRI inception_resnet_1 model. The corresponding visualization of prediction distributions is included in Fig. 4C. Multimodal 3D CNN models showed superior ability to differentiate patients from controls with wider gaps between their prediction distribution peaks.

3.4. Interpretation of the trained CNN models

To examine the interpretability of our models, we applied the gradient class activation map approach on multi-channel and multimodal inception_resnet_1 models which had the highest classification accuracy on the NUSDAST dataset and IMH dataset to localize the discriminative regions for classification. Individual-level heatmap (Fig. 5, left) and group-level statistical maps (Fig. 5, right) were generated for each input channel (GM, WM, CSF, FA and MD).

Overall, the brain regions within GM, WM, CSF, or white matter microstructure that contributed significantly to the classification of schizophrenia patients resembled previous literature in schizophrenia (Fig. 5, Supplementary Tables 6 and 7). Specifically, grey matter volume in the bilateral insula, orbital prefrontal cortex, putamen, caudate, amygdala, thalamus and cerebellum lobule VI were found to be critical features for classification. In addition, the WM volume of widespread deep WM regions and corpus callosum as well as the CSF of the third ventricle and fourth ventricle were identified. Using dMRI, the FA in frontotemporal, interhemispheric and cortico-striatal-thalamic white matter tracts, including corpus callosum, fornix, corona radiata and thalamic radiation as well as the MD in the lateral ventricle, third ventricle and fourth ventricle were reported from group-level discrimination maps.

4. Discussion

This study investigated the performance of both 2D and 3D CNN on the classification of schizophrenia patients based on multimodal structural brain imaging. Deep feature based on pretrained 2D CNN and naive 3D CNN models trained from scratch were compared with handcrafted feature-based machine learning. Naïve 3D CNN models achieved superior performance in terms of both cross-validation and generalizability on an independent testing dataset. For the first time, we also demonstrated that multimodal neuroimaging-based CNN models which integrated both structural and diffusion MRI accomplished performance superior to single modality-based models. Furthermore, the 3D CNN

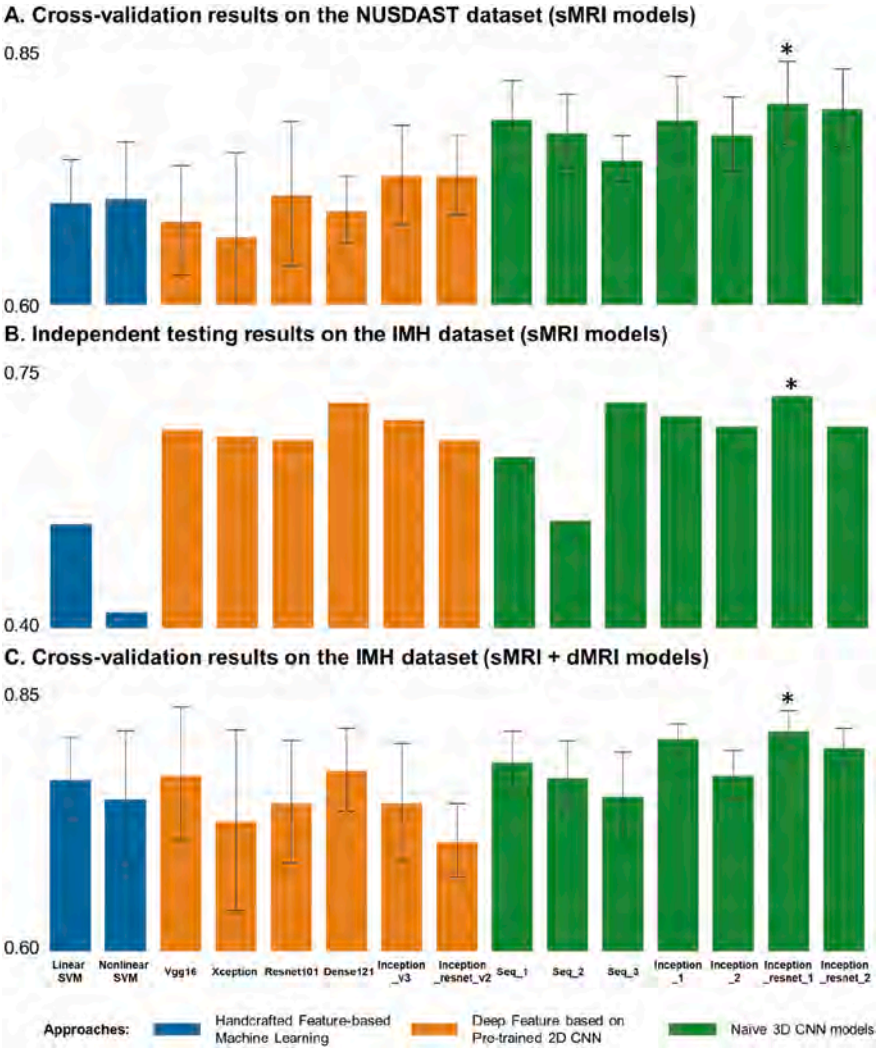

Fig. 3. Classification Accuracy in cross-validation and independent testing across all models. Row A represents cross-validation results on the NUSDAST dataset using sMRI models. Some deep feature models performed better than handcrafted feature-based machine learning. Naive 3D CNN models achieved higher accuracy than both handcrafted feature-based machine learning and deep feature based on pretrained 2D CNN. Row B represents independent testing results on the IMH dataset using sMRI models trained with the NUSDAST dataset. Most naive 3D CNN models achieved relatively high accuracy except sequential_2 model. Row C represents cross-validation results on the IMH dataset using sMRI+dMRI models. Naive 3D CNN models achieved higher accuracy than both handcrafted feature-based machine learning and deep feature based on pretrained 2D CNN. In overall, 3D Inception_resnet_2 model achieved the highest accuracy across all tasks.

models were interpreted with gradient class activation maps to identify brain regions critical for classification. We localized insula, orbital prefrontal cortex, striatum, thalamus, cerebellum lobule VI in grey matter volume, cortico-subcortical fiber tracts in FA and enlarged ventricles in CSF and MD. These areas were consistent with previous literature and supported the salience network/striatum dysfunction hypothesis in schizophrenia.

4.1. CNN performed better than handcrafted feature-based machine learning

The classification performance in previous studies varied due to different subject selection criteria, preprocessing protocols and model implementation settings (Arbabshirani et al., 2017; Winterburn et al., 2019). We implemented the benchmark handcrafted feature-based machine learning and CNN models on the same data for a fair comparison among different approaches. In our cross-validation results, deep feature approach based on pre-trained 2D inception_v3 outperformed nonlinear SVM with a gap of 9.05% in accuracy. Naive 3D inception_resnet_1 model trained from scratch outperformed deep feature approach with 6.87% improvement in accuracy. Compared to other diseases such as Alzheimer's disease which is characterized by distinct and specific atrophy patterns, the neuroanatomical alterations in schizophrenia tend to be subtle, variable and widely distributed (Kelly et al., 2018; Van Erp et al., 2018). The superior performance of CNN models reported here further demonstrated its powerful capabilities of anatomical contextual information extraction and feature learning, which are critical for schizophrenia classification. A recent study on predicting clinical improvement in psychosis with functional MRI compared handcrafted feature-based machine learning and deep learning, which yielded consistent findings with our work (Smucny et al., 2021). Further, the independent testing results indicated higher generalizability of 3D CNN models compared to handcrafted feature-based machine learning, which is critical to real-world applications. Feature extraction in handcrafted feature-based machine learning might be affected by heterogeneity in clinical and demographic factors such as disease duration, medications, age, and hence pose restrictions on applying trained models to new data (Arbabshirani et al., 2017). Our approach demonstrated the potential of 3D CNN as an efficient deep learning model for classifying schizophrenia with unseen data.

4.2. 3D naive CNN performed better than deep feature based on pretrained 2D CNN

Our results showed that naive 3D CNN trained from scratch outperformed deep feature approach based on pretrained 2D CNN in terms of both cross-validation and independent testing results. Previous studies compared 2D and 3D CNN models as well as models trained from scratch and transfer learning in other classification tasks (Kermany et al., 2018; Litjens et al., 2017; Yang et al., 2018b; Yu et al., 2019; Zhu et al., 2019). Transfer learning that included both fine-tune and deep feature approaches have been shown to be superior in many 2D image

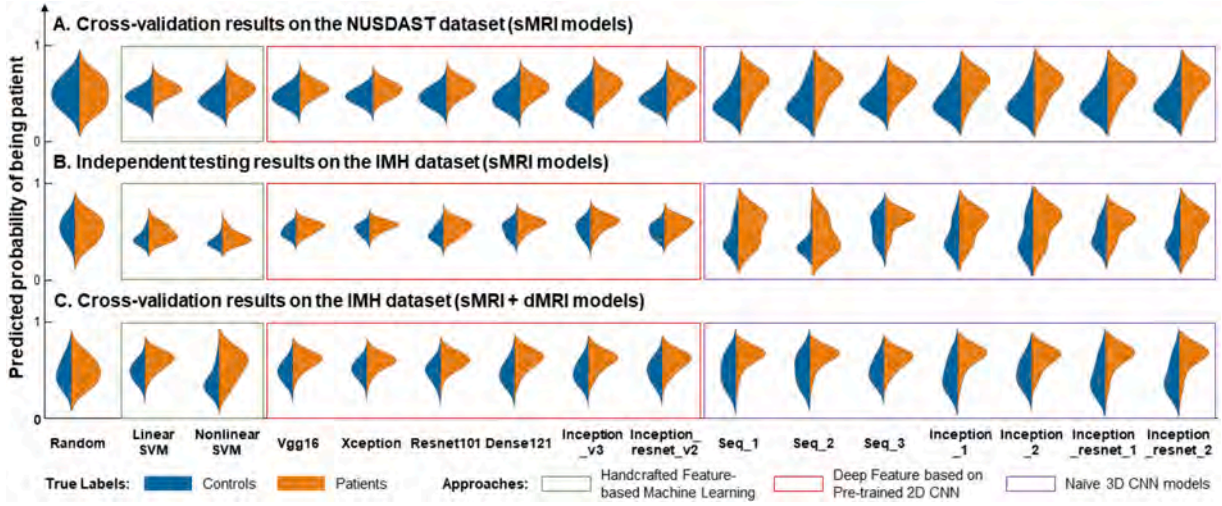

Fig. 4. Prediction probability distribution of schizophrenia and controls in cross-validation and independent testing across all approaches. Random predictions for both controls (blue) and patients (orange) have a distribution peak at around 0.5 (first column). Y-axis represents the predicted probability of being patient from 0 to 1. Prediction probability distributions are presented in the following order: handcrafted feature-based machine learning are in the green box, deep feature based on pretrained 2D CNN are in the red box and naive 3D CNN models are in the purple box. Row A represents cross-validation results on the NUSDAST dataset using sMRI models. Predictions generated by the proposed models demonstrated higher than randomness accuracy, which had wider gaps between the predicted probability distributions of controls and patients. Naive 3D CNN models performed the best with clear separation of patients and controls. Row B represents independent testing results on the IMH dataset using sMRI models trained with the NUSDAST dataset. Handcrafted feature-based machine learning approaches and deep feature based on pretrained 2D CNN barely differ from random predictions. Most naive 3D CNN models obtained relatively higher performance, but sequential models categorize all subjects into one class. Inception models and Inception_resnet models obtained better classification results with more balanced specificity and sensitivity. Row C represents cross-validation results on the IMH dataset using sMRI+dMRI models. Handcrafted feature-based machine learning and deep feature based on pretrained 2D CNN performed better than random predictions. Naive 3D CNN models demonstrated better performance with wider gaps between the prediction of patients and controls. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Single modal (sMRI) independent testing results on the IMH dataset.

		ACC	SP	SE	ROC-AUC
Handcrafted feature-based machine learning	Linear SVM	54.02%	85.53%	37.84%	0.72
	Nonlinear SVM (RBF kernel)	42.41%	97.37%	14.19%	0.73
Deep feature based on pretrained 2D CNN	VGG16	66.52%	59.21%	70.27%	0.71
	Xception	65.62%	36.84%	80.41%	0.70
	Resnet101	64.73%	69.74%	62.16%	0.73
	Densenet121	70.09%	42.11%	84.46%	0.73
	Inception_V3	67.86%	36.84%	83.78%	0.71
	Inception_resnet_V2	65.18%	48.68%	73.65%	0.70
Naive 3D CNN models	Sequential_1	62.95%	75.00%	56.76%	0.73
	Sequential_2	54.46%	93.42%	34.46%	0.72
	Sequential_3	70.09%	46.05%	82.43%	0.72
	Inception_1	68.30%	68.42%	68.24%	0.74
	Inception_2	66.96%	65.79%	67.57%	0.71
	Inception_resnet_1	**70.98%**	63.16%	75.00%	0.75
	Inception_resnet_2	66.96%	72.37%	64.19%	0.61

The testing results on the independent IMH dataset using ensemble approach with models trained on the NUSDAST dataset are listed. The best testing accuracy is achieved by the Naive 3D CNN models - Inception_resnet_1, indicating better generalizability.

classification tasks, especially when compared to new models trained from scratch with limited sample sizes (Kermany et al., 2018; Yang et al., 2018b; Yu et al., 2019). Deep feature approach outperformed the fine-tune approach in a task of tumor classification based on MRI patches (Zhu et al., 2019). Moreover, 3D CNN models trained from scratch performed better than 2D models trained from scratch in the classification of pulmonary nodules (Dou et al., 2017; Liu and Kang, 2017).

To date, there is no consensus regarding the performance comparison between transfer learning approaches based on 2D CNN and 3D CNN models trained from scratch. Our results suggest that 3D networks are more suitable for 3D MRI data-based classification as compared to 2D networks. With convolution and pooling in 3D CNN models operating in a cubic manner using 3D kernels, 3D CNN models are naturally more suitable for volumetric medical image processing with higher proficiency in processing 3D spatial and contextual information than 2D networks. This in turn results in more efficient feature representations across levels, which is critical for schizophrenia identification.

4.3. Multimodal inputs and complex topologies improved classification accuracy

Multimodal imaging-based models integrating sMRI and dMRI data outperformed single-modal models with only sMRI data for all three approaches. Notably, the integration of multimodal information greatly improved classification specificity, which is essential for classification with imbalanced classes. Multimodal features provide information from different perspectives thus allowing models to understand the neural substrates associated with schizophrenia with complementary information from various modalities. Previous studies illustrated the benefits of combining sMRI and dMRI as well as fMRI for the classification of schizophrenia using handcrafted feature-based machine learning (Isobe et al., 2016; Lei et al., 2020; Saarinen et al., 2020; Salvador et al., 2019), suggesting the possibility of future research involving the incorporation of additional neuroimaging modalities.

In addition to using multimodal inputs, 3D CNN models with complex topologies such as inception module and residual module improved the classification accuracy further. The inception module divides the network into multiple branches with different convolutional kernels thus allowing operating convolutions with different kernels on the same level (Szegedy et al., 2015). The residual module allows incorporation of information from previous layers (He et al., 2016). Together, the inception module and residual module enhanced the feature learning process of CNN and upgraded the performance of developed models. A

Table 4
Multimodal (sMRI and dMRI) cross-validation results on the IMH dataset.

		ACC	SP	SE	ROC-AUC
Handcrafted feature-based machine learning	**Linear SVM**	76.37% (±3.96%)	52.50% (±13.27%)	88.51% (±5.88%)	0.82 (±0.02)
	Nonlinear SVM (RBF kernel)	74.58% (±6.36%)	74.92% (±5.27%)	74.32% (±11.74%)	0.81 (±0.04)
Deep feature based on pretrained 2D CNN	**VGG16**	76.85% (±6.36%)	55.25% (±12.92%)	87.86% (±5.87%)	0.82 (±0.05)
	Xception	72.45% (±8.56%)	46.08% (±11.22%)	85.91% (±9.75%)	0.77 (±0.07)
	Resnet101	74.18% (±5.80%)	52.50% (±5.82%)	85.20% (±11.23%)	0.81 (±0.03)
	Densenet121	77.26% (±3.93%)	55.08% (±11.93%)	88.51% (±7.54%)	0.84 (±0.06)
	Inception_V3	74.19% (±5.57%)	48.75% (±9.29%)	87.22% (±4.32%)	0.81 (±0.04)
	Inception_resnet_V2	70.54% (±3.51%)	46.00% (±5.33%)	83.10% (±5.70%)	0.77 (±0.03)
Naive 3D CNN models	Sequential_1	78.07% (±3.64%)	53.87% (±19.14%)	90.46% (±7.74%)	0.81 (±0.06)
	Sequential_2	76.61% (±3.72%)	49.85% (±19.84%)	90.20% (±8.13%)	0.79 (±0.06)
	Sequential_3	74.81% (±4.94%)	50.91% (±22.99%)	87.10% (±9.61%)	0.75 (±0.08)
	Inception_1	80.28% (±2.24%)	64.39% (±13.52%)	88.41% (±6.91%)	0.84 (±0.02)
	Inception_2	76.78% (±3.06%)	57.67% (±20.43%)	86.64% (±11.31%)	0.79 (±0.05)
	Inception_resnet_1	**81.02% (±2.52%)**	70.42% (±12.00%)	86.44% (±6.45%)	0.84 (±0.03)
	Inception_resnet_2	79.43% (±2.46%)	63.73% (±18.18%)	87.39% (±8.74%)	0.84 (±0.03)

The cross-valuation results using both sMRI and dMRI of all three approaches are listed with mean (+/−standard deviation). Naive 3D CNN models achieved higher accuracy than both handcrafted feature-based machine learning and deep feature approach. Multimodal models generally outperformed the single-modal models based on sMRI only. The highest accuracy obtained by each approach is highlighted in bold. The overall highest accuracy was obtained by Inception_resnet_1.

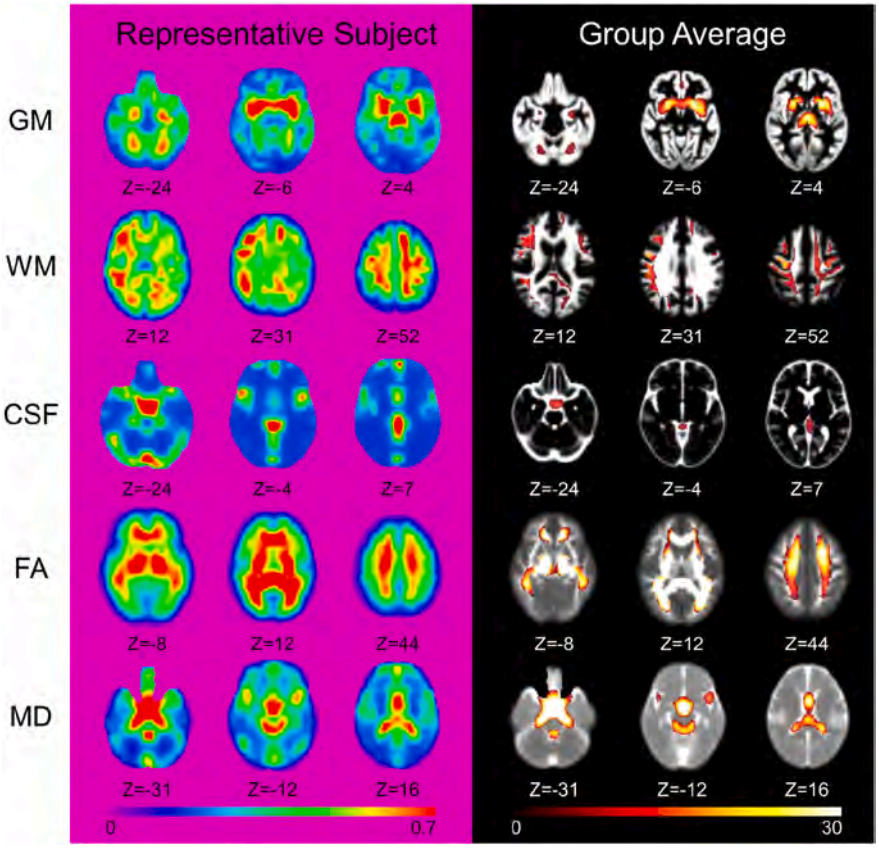

Fig. 5. Gradient class activation mapping for schizophrenia classification. The left panel shows heat maps generated by gradient class activation map of one representative subject at the convolutional layer in each network branch of different modalities respectively. Hot colors represent more significant contribution to classification results. The right panel represents T maps (Bonferroni correction at $p < 0.01$) generated from group-level statistical analysis with heat maps of all subjects in each modality. Bright colors represent a higher chance that these regions contribute more to classification results at a group-level. The region names in grey matter and FA are listed in Supplementary Tables 6 and 7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previous study examining the classification of lung nodules using CT also found that the inception module outperformed sequential models (Liu and Kang, 2017) but the residual module did not enhance the CNN models' performance. The small scaling factor used in that study might have attributed but another plausible factor is the training sample size as a previous study with a rather large sample size achieved good cross-validation performance with sequential models (Oh et al., 2020). Future work is needed to validate the hypothesis that training the proposed multimodal 3D CNN models with complex topologies with a larger sample yields higher accuracy and generalizability.

An interesting observation is that models with deeper architecture did not improve accuracy. Sequential_1 model, which had the least layers, achieved the highest accuracy among the three sequential models. Inception_1 and inception_resnet_1 model outperformed inception_2 and inception_resnet_2 model, both of which had deeper architectures. This is in line with a previous study, which had the same observation and concluded that using 3D CNN to classify lung nodules using CT did not need a very deep network (Liu and Kang, 2017). Our speculation is that preprocessing steps like downsampling and registration to standard template caused spatial or finer information loss, which in turn affected the efficient feature representations at very deep layers and subsequently led to lower performance of models with deeper architecture. Future investigations using high-resolution 3D neuroimaging data in native space is needed to evaluate models with deeper architectures.

4.4. Gradient class activation map identified critical regions for classification

Explanation of the classification process verifies the accountability of the model and also leads to biological inference (Guidotti et al., 2018). Using the gradient class activation map approach, we identified grey matter volume in the insula, striatum, thalamus, FA in cortico-subcortical fibers and enlarged ventricles as crucial brain structural features for differentiating schizophrenia patients from controls. These findings coincide with earlier schizophrenia literature.

Converging evidence have shown widespread grey matter reduction and cortical thickness thinning in frontal, temporal lobes and subcortical regions including striatum and hippocampus (Bora et al., 2011; Haukvik et al., 2018; Koelkebeck et al., 2019; Mitelman, 2019; Pergola et al., 2015; Van Erp et al., 2018; Zheng et al., 2019). Specifically, our results resembled previous meta-analysis findings that schizophrenia is associated with grey matter reduction at bilateral insula, medial frontal cortex and the thalamus (Bora et al., 2011). Insula is the critical component of the salience network which is responsible for switching between default mode network and central execution network (Sridharan et al., 2008). Abnormalities in the frontal lobe (especially insula, orbital prefrontal cortex and anterior cingulate cortex), striatum and thalamus suggest impairment in the cortico-striatal-thalamic loop circuits, which serve as a discrete regulatory loop circuit for salience network. Cerebellum lobule VI found in our results corresponds to the cerebellar contribution to the salience network (Habas et al., 2009). Along with our findings here, it collectively supports the salience network dysfunction hypothesis which impairs cognitive control, behavior and emotion thereby leading to symptoms of schizophrenia (Fornito et al., 2012; Miyata, 2019; Palaniyappan et al., 2013; Palaniyappan and Liddle, 2012; Peters et al., 2016; Van Den Heuvel and Fornito, 2014).

In parallel, widespread FA was reported in dMRI meta-analysis studies with regional specificity at frontotemporal, interhemispheric and corticothalamic regions forming the cortico-striatal-thalamic loop circuits. This is in line with our findings from the FA maps and salience network dysfunction hypothesis (Cookey et al., 2014; Di Biase et al., 2020; Kelly et al., 2018). We also identified ventricles as significant regions from CSF and MD maps, which corroborate with previous reports of ventricle enlargement (Gaser et al., 2004; Kempton et al., 2010; Kuo and Pogue-Geile, 2019; Wright et al., 2000). The findings of our study suggest that the 3D CNN model we developed has the potential to identify crucial neuroanatomical features for classification of psychiatric illnesses such as schizophrenia from healthy controls. This in turn allows for a better understanding of the neural basis of schizophrenia.

4.5. Limitations and considerations

There are some considerations related to the adoption of 3D CNN models in neuroimaging-based classification. Firstly the high computational cost during training caused by the high dimensionality of input data and the large number of parameters may constrain the development of 3D CNN models. In this study, we downsampled the input maps to reduce the GPU memory requirement. The depths of the models are also affected by the GPU memory constraints. Secondly, the increasing number of parameters in 3D CNN compared to 2D CNN makes it difficult to train the 3D models with limited training samples. It limits both the model architecture and the model performance. The major limitation of this study is that the sample size is relatively modest, especially for 3D CNN network training. This might result in less efficient feature extraction and lower generalizability. Registration error and downsampling may also ignore some subtle anatomical differences and low-level contextual features that could be important for classification. Furthermore, the 2D CNN transfer learning approach adopted here did not contain any manipulation of the produced feature maps. Further research on deep feature fusion across dimensions and scales might improve the performance of pretrained 2D CNN models. Future work in native brain space with data augmentation is needed to improve generalizability by accounting for inter-subject anatomical variability and data quality variation across sites.

5. Conclusion

To classify patients with schizophrenia and healthy controls using 3D brain MR images, we developed multimodal 3D CNN models with different architectures and utilized deep feature approach based on 2D pretrained CNN. Our study demonstrated the superiority of using 3D CNN models and multimodal deep learning features extracted from sMRI and dMRI data over approaches using handcrafted feature-based method and single neuroimaging modality. Based on the learned 3D CNN models, we further localized the crucial regions for classification which consisted of grey matter volume in insula, orbital prefrontal cortex, striatum, thalamus, FA in cortico-subcortical fiber tracts and enlarged ventricles. These findings were in agreement with previous literature and supported the salience network/striatum dysfunction hypothesis in schizophrenia. This study highlighted the potential of CNN for automatic and efficient feature extraction from 3D brain structural imaging data and multimodal imaging information integration, thereby providing an interpretable framework for objective imaging-based assays for individual-level classification in psychiatric disorders.

Acknowledgements

This work was supported by 1) NTU Institute for Health Technologies, Interdisciplinary Graduate Programme, Nanyang Technological University, 2) Yong Loo Lin School of Medicine, National University of Singapore, and 3) Duke-NUS Medical School Signature Program supported by Ministry of Health -Singapore. This study was also supported by the National Medical Research Council under the Centre Grant Programme (Institute of Mental Health, Singapore) (NMRC/CG/004/2013) (NFH), National Healthcare Group, Singapore (SIG/05004; SIG/05028), and the Singapore Bioimaging Consortium (RP C-009/2006) research grants awarded to KS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2021.06.011.

References

Andersson, J.L.R., Jenkinson, M., Smith, S., 2007. Non-linear Registration aka Spatial Normalisation FMRIB Technial Report TR07JA2. FMRIB Anal. Gr. Univ, Oxford.

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. https://doi.org/10.1016/j.neuroimage.2016.02.079.

Ardekani, B.A., Tabesh, A., Sevy, S., Robinson, D.G., Bilder, R.M., Szeszko, P.R., 2011. Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. Hum. Brain Mapp. 32, 1–9. https://doi.org/10.1002/hbm.20995.

Bora, E., Fornito, A., Radua, J., Walterfang, M., Seal, M., Wood, S.J., Yücel, M., Velakoulis, D., Pantelis, C., 2011. Neuroanatomical abnormalities in schizophrenia: a multimodal voxelwise meta-analysis and meta-regression analysis. Schizophr. Res. 127, 46–57. https://doi.org/10.1016/j.schres.2010.12.020.

Chin, R., You, A.X., Meng, F., Zhou, J., Sim, K., 2018. Recognition of schizophrenia with regularized support vector machine and sequential region of interest selection using structural magnetic resonance imaging. Sci. Rep. 8, 1–10. https://doi.org/10.1038/s41598-018-32290-9.

Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, pp. 1800–1807. https://doi.org/10.1109/CVPR.2017.195.

Cookey, J., Bernier, D., Tibbo, P.G., 2014. White matter changes in early phase schizophrenia and cannabis use: an update and systematic review of diffusion tensor imaging studies. Schizophr. Res. 156, 137–142. https://doi.org/10.1016/j.schres.2014.04.026.

Del Barrio, V., 2016. Diagnostic and statistical manual of mental disorders. In: The Curated Reference Collection in Neuroscience and Biobehavioral Psychology. American Psychiatric Pub. https://doi.org/10.1016/B978-0-12-809324-5.05530-9.

Di Biase, M.A., Pantelis, C., Zalesky, A., 2020. White matter pathology in schizophrenia. In: Kubicki, M., Shenton, M.E. (Eds.), Neuroimaging in Schizophrenia. Springer International Publishing, Cham, pp. 71–91. https://doi.org/10.1007/978-3-030-35206-6_4.

Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A., 2017. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng. 64, 1558–1567. https://doi.org/10.1109/TBME.2016.2613502.

Fanous, A.H., Amdur, R.L., O'Neill, F.A., Walsh, D., Kendler, K.S., 2012. Concordance between chart review and structured interview assessments of schizophrenic symptoms. Compr. Psychiatry 53, 275–279. https://doi.org/10.1016/j.comppsych.2011.04.006.

De Filippis, R., Carbone, E.A., Bruni, A., Pugliese, V., Segura-garcia, C., De Fazio, P., 2019. Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. Neuropsychiatr. Dis. Treat. 15, 1605–1627.

Fornito, A., Yücel, M., Patti, J., Wood, S.J., Pantelis, C., 2009. Mapping grey matter reductions in schizophrenia: an anatomical likelihood estimation analysis of voxel-based morphometry studies. Schizophr. Res. 108, 104–113. https://doi.org/10.1016/j.schres.2008.12.011.

Fornito, A., Zalesky, A., Pantelis, C., Bullmore, E.T., 2012. Schizophrenia, neuroimaging and connectomics. Neuroimage 62, 2296–2314. https://doi.org/10.1016/j.neuroimage.2011.12.090.

Gaser, C., Nenadic, I., Buchsbaum, B.R., Hazlett, E.A., Buchsbaum, M.S., 2004. Ventricular enlargement in schizophrenia related to volume reduction of the thalamus, striatum, and superior temporal cortex. Am. J. Psychiatry 161, 154–156. https://doi.org/10.1176/appi.ajp.161.1.154.

Gong, B., Naveed, S., Hafeez, D.M., Afzal, K.I., Majeed, S., Abele, J., Nicolaou, S., Khosa, F., 2019. Neuroimaging in psychiatric disorders: a bibliometric analysis of the 100 most highly cited articles. J. Neuroimaging 29, 14–33. https://doi.org/10.1111/jon.12570.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM Comput. Surv. 51 https://doi.org/10.1145/3236009.

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C.F., Menon, V., Greicius, M. D., 2009. Distinct cerebellar contributions to intrinsic connectivity networks. J. Neurosci. 29, 8586–8594. https://doi.org/10.1523/JNEUROSCI.1868-09.2009.

Haijma, S.V., Van Haren, N., Cahn, W., Koolschijn, P.C.M.P., Hulshoff Pol, H.E., Kahn, R. S., 2013. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. Schizophr. Bull. 39, 1129–1138. https://doi.org/10.1093/schbul/sbs118.

Haukvik, U.K., Tamnes, C.K., Söderman, E., Agartz, I., 2018. Neuroimaging hippocampal subfields in schizophrenia and bipolar disorder: a systematic review and meta-analysis. J. Psychiatr. Res. 104, 217–226. https://doi.org/10.1016/j.jpsychires.2018.08.012.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Ho, N.F., Iglesias, J.E., Sum, M.Y., Kuswanto, C.N., Sitoh, Y.Y., De Souza, J., Hong, Z., Fischl, B., Roffman, J.L., Zhou, J., 2017a. Progression from selective to general involvement of hippocampal subfields in schizophrenia. Mol. Psychiatry 22, 142–152. https://doi.org/10.1038/mp.2016.4.

Ho, N.F., Li, Z., Ji, F., Wang, M., Kuswanto, C.N., Sum, M.Y., Tng, H.Y., Sitoh, Y.Y., Sim, K., Zhou, J., 2017b. Hemispheric lateralization abnormalities of the white matter microstructure in patients with schizophrenia and bipolar disorder. J. Psychiatry Neurosci. 42, 242–251. https://doi.org/10.1503/jpn.160090.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.

Iizuka, T., Fukasawa, M., Kameyama, M., 2019. Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. Sci. Rep. 9, 1–9. https://doi.org/10.1038/s41598-019-45415-5.

Isobe, M., Miyata, J., Hazama, M., Fukuyama, H., Murai, T., Takahashi, H., 2016. Multimodal neuroimaging as a window into the pathological physiology of schizophrenia: current trends and issues. Neurosci. Res. 102, 29–38. https://doi.org/10.1016/j.neures.2015.07.009.

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., Andreassen, O. A., Arango, C., Banaj, N., Bouix, S., Bousman, C.A., Brouwer, R.M., Bruggemann, J., Bustillo, J., Cahn, W., Calhoun, V., Cannon, D., Carr, V., Catts, S., Chen, J., Chen, J. X., Chen, X., Chiapponi, C., Cho, K.K., Ciullo, V., Corvin, A.S., Crespo-Facorro, B., Cropley, V., De Rossi, P., Diaz-Caneja, C.M., Dickie, E.W., Ehrlich, S., Fan, F.M., Faskowitz, J., Fatouros-Bergman, H., Flyckt, L., Ford, J.M., Fouche, J.P., Fukunaga, M., Gill, M., Glahn, D.C., Gollub, R., Goudzwaard, E.D., Guo, H., Gur, R. E., Gur, R.C., Gurholt, T.P., Hashimoto, R., Hatton, S.N., Henskens, F.A., Hibar, D.P., Hickie, I.B., Hong, L.E., Horacek, J., Howells, F.M., Hulshoff Pol, H.E., Hyde, C.L., Isaev, D., Jablensky, A., Jansen, P.R., Janssen, J., Jönsson, E.G., Jung, L.A., Kahn, R. S., Kikinis, Z., Liu, K., Klauser, P., Knöchel, C., Kubicki, M., Lagopoulos, J., Langen, C., Lawrie, S., Lenroot, R.K., Lim, K.O., Lopez-Jaramillo, C., Lyall, A., Magnotta, V., Mandl, R.C.W., Mathalon, D.H., McCarley, R.W., McCarthy-Jones, S., McDonald, C., McEwen, S., McIntosh, A., Melicher, T., Mesholam-Gately, R.I., Michie, P.T., Mowry, B., Mueller, B.A., Newell, D.T., O'Donnell, P., Oertel-Knöchel, V., Oestreich, L., Paciga, S.A., Pantelis, C., Pasternak, O., Pearlson, G., Pellicano, G.R., Pereira, A., Pineda Zapata, J., Piras, F., Potkin, S.G., Preda, A., Rasser, P.E., Roalf, D.R., Roiz, R., Roos, A., Rotenberg, D., Satterthwaite, T.D., Savadjiev, P., Schall, U., Scott, R.J., Seal, M.L., Seidman, L.J., Shannon Weickert, C., Whelan, C.D., Shenton, M.E., Kwon, J.S., Spalletta, G., Spaniel, F., Sprooten, E., Stäblein, M., Stein, D.J., Sundram, S., Tan, Y., Tan, S., Tang, S., Temmingh, H.S., Westlye, L.T., Tønnesen, S., Tordesillas-Gutierrez, D., Doan, N.T., Vaidya, J., Van Haren, N.E.M., Vargas, C.D., Vecchio, D., Velakoulis, D., Voineskos, A., Voyvodic, J. Q., Wang, Z., Wan, P., Wei, D., Weickert, T.W., Whalley, H., White, T., Whitford, T. J., Wojcik, J.D., Xiang, H., Xie, Z., Yamamori, H., Yang, F., Yao, N., Zhang, G., Zhao, J., Van Erp, T.G.M., Turner, J., Thompson, P.M., Donohoe, G., 2018. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. Mol. Psychiatry 23, 1261–1269. https://doi.org/10.1038/mp.2017.170.

Kempton, M.J., Stahl, D., Williams, S.C.R., DeLisi, L.E., 2010. Progressive lateral ventricular enlargement in schizophrenia: a meta-analysis of longitudinal MRI studies. Schizophr. Res. 120, 54–62. https://doi.org/10.1016/j.schres.2010.03.036.

Kennedy, J.L., Altar, C.A., Taylor, D.L., Degtiar, I., Hornberger, J.C., 2014. The social and economic burden of treatment-resistant schizophrenia: a systematic literature review. Int. Clin. Psychopharmacol. 29, 63–76. https://doi.org/10.1097/YIC.0b013e32836508e6.

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, Justin, Prasadha, M.K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, Jason, Ziyar, I., Shi, A., Zhang, R., Zheng, L., Hou, R., Shi, W., Fu, X., Duan, Y., Huu, V.A.N., Wen, C., Zhang, E.D., Zhang, C.L., Li, O., Wang, X., Singer, M.A., Sun, X., Xu, J., Tafreshi, A., Lewis, M.A., Xia, H., Zhang, K., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172, 1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.

Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A., Catheline, G., 2018. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies.

Koelkebeck, K., Dannlowski, U., Ohrmann, P., Suslow, T., Murai, T., Bauer, J., Pedersen, A., Matsukawa, N., Son, S., Haidl, T., Miyata, J., 2019. Gray matter volume reductions in patients with schizophrenia: a replication study across two cultural backgrounds. Psychiatry Res. Neuroimaging 292, 32–40. https://doi.org/10.1016/j.pscychresns.2019.08.008.

Kogan, A., Alpert, K., Ambite, J.L., Marcus, D.S., Wang, L., 2016. Northwestern University schizophrenia data sharing for SchizConnect: a longitudinal dataset for large-scale integration. Neuroimage 124, 1196–1201. https://doi.org/10.1016/j.neuroimage.2015.06.030.

Kuo, S.S., Pogue-Geile, M.F., 2019. Variation in fourteen brain structure volumes in schizophrenia: a comprehensive meta-analysis of 246 studies. Neurosci. Biobehav. Rev. 98, 85–94. https://doi.org/10.1016/j.neubiorev.2018.12.030.

Kurth, F., Gaser, C., Luders, E., 2015. A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (SPM). Nat. Protoc. 10, 293–304. https://doi.org/10.1038/nprot.2015.014.

Lee, J.G., Jun, S., Cho, Y.W., Lee, H., Kim, G.B., Seo, J.B., Kim, N., 2017. Deep learning in medical imaging: general overview. Korean J. Radiol. 18, 570–584. https://doi.org/10.3348/kjr.2017.18.4.570.

Lei, D., Pinaya, W.H.L., Young, J., van Amelsvoort, T., Marcelis, M., Donohoe, G., Mothersill, D.O., Corvin, A., Vieira, S., Huang, X., Lui, S., Scarpazza, C., Arango, C., Bullmore, E., Gong, Q., McGuire, P., Mechelli, A., 2020. Integrating machining learning and multimodal neuroimaging to detect schizophrenia at the level of the individual. Hum. Brain Mapp. 41, 1119–1135. https://doi.org/10.1002/hbm.24863.

Lerman-Sinkoff, D.B., Kandala, S., Calhoun, V.D., Barch, D.M., Mamah, D.T., 2019. Transdiagnostic multimodal neuroimaging in psychosis: structural, resting-state, and

task magnetic resonance imaging correlates of cognitive control. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 4, 870–880. https://doi.org/10.1016/j.bpsc.2019.05.004.

Liang, S., Li, Y., Zhang, Z., Kong, X., Wang, Q., Deng, W., Li, X., Zhao, L., Li, M., Meng, Y., Huang, F., Ma, X., Li, X.M., Greenshaw, A.J., Shao, J., Li, T., 2019. Classification of first-episode schizophrenia using multimodal brain features: a combined structural and diffusion imaging study. Schizophr. Bull. 45, 591–599. https://doi.org/10.1093/schbul/sby091.

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X., 2018. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front. Neurosci. 12, 1–13. https://doi.org/10.3389/fnins.2018.00777.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

Liu, K., Kang, G., 2017. Multiview convolutional neural networks for lung nodule classification. Int. J. Imaging Syst. Technol. 27, 12–22. https://doi.org/10.1002/ima.22206.

Mahendran, A., Vedaldi, A., 2015. Understanding deep image representations by inverting them. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5188–5196.

Mikolas, P., Hlinka, J., Skoch, A., Pitra, Z., Frodl, T., Spaniel, F., Hajek, T., 2018. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. BMC Psychiatry 18, 1–7. https://doi.org/10.1186/s12888-018-1678-y.

Mitelman, S.A., 2019. Transdiagnostic neuroimaging in psychiatry: a review. Psychiatry Res. 277, 23–38. https://doi.org/10.1016/j.psychres.2019.01.026.

Miyata, J., 2019. Toward integrated understanding of salience in psychosis. Neurobiol. Dis. 131, 104414. https://doi.org/10.1016/j.nbd.2019.03.002.

Ng, K.K., Lo, J.C., Lim, J.K.W., Chee, M.W.L., Zhou, J., 2016. Reduced functional segregation between the default mode network and the executive control network in healthy older adults: a longitudinal study. Neuroimage 133, 321–330. https://doi.org/10.1016/j.neuroimage.2016.03.029.

Nieuwenhuis, M., van Haren, N.E.M., Hulshoff Pol, H.E., Cahn, W., Kahn, R.S., Schnack, H.G., 2012. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. Neuroimage 61, 606–612. https://doi.org/10.1016/j.neuroimage.2012.03.079.

Oh, J., Oh, B.L., Lee, K.U., Chae, J.H., Yun, K., 2020. Identifying schizophrenia using structural MRI with a deep learning algorithm. Front. Psych. 11, 16. https://doi.org/10.3389/fpsyt.2020.00016.

Ott, C.V., Johnson, C.B., Macoveanu, J., Miskowiak, K., 2019. Structural changes in the hippocampus as a biomarker for cognitive improvements in neuropsychiatric disorders: a systematic review. Eur. Neuropsychopharmacol. 29, 319–329. https://doi.org/10.1016/j.euroneuro.2019.01.105.

Palaniyappan, L., Liddle, P.F., 2012. Does the salience network play a cardinal role in psychosis? An emerging hypothesis of insular dysfunction. J, Psychiatry Neurosci.

Palaniyappan, L., Simmonite, M., White, T.P., Liddle, E.B., Liddle, P.F., 2013. Neural primacy of the salience processing system in schizophrenia. Neuron 79, 814–828.

Pergola, G., Selvaggi, P., Trizio, S., Bertolino, A., Blasi, G., 2015. The role of the thalamus in schizophrenia from a neuroimaging perspective. Neurosci. Biobehav. Rev. 54, 57–75. https://doi.org/10.1016/j.neubiorev.2015.01.013.

Peters, S.K., Dunlop, K., Downar, J., 2016. Cortico-striatal-thalamic loop circuits of the salience network: a central pathway in psychiatric disease and treatment. Front. Syst. Neurosci. 10, 1–23. https://doi.org/10.3389/fnsys.2016.00104.

Pinaya, W.H.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. Hum. Brain Mapp. 40, 944–954. https://doi.org/10.1002/hbm.24423.

Power, B.D., Nguyen, T., Hayhow, B., Looi, J.C.L., 2016. Neuroimaging in psychiatry: an update on neuroimaging in the clinical setting. Australas. Psychiatry 24, 157–163. https://doi.org/10.1177/1039856215618525.

Rashid, B., Calhoun, V., 2020. Towards a brain-based predictome of mental illness. Hum. Brain Mapp. 1–68. https://doi.org/10.1002/hbm.25013.

Rozycki, M., Satterthwaite, T.D., Koutsouleris, N., Erus, G., Doshi, J., Wolf, D.H., Fan, Y., Gur, R.E., Gur, R.C., Meisenzahl, E.M., 2018. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. Schizophr. Bull. 44, 1035–1044.

Saarinen, A.I.L., Huhtaniska, S., Pudas, J., Björnholm, L., Jukuri, T., Tohka, J., Granö, N., Barnett, J.H., Kiviniemi, V., Veijola, J., Hintsanen, M., Lieslehto, J., 2020. Structural and functional alterations in the brain gray matter among first-degree relatives of schizophrenia patients: a multimodal meta-analysis of fMRI and VBM studies. Schizophr. Res. 216, 14–23. https://doi.org/10.1016/j.schres.2019.12.023.

Salvador, R., Radua, J., Canales-Rodrõâguez, E.J., Solanes, A., Sarroa, S., Goikolea, J.M., Valiente, A., Montea, G.C., Natividad, M.D.C., Guerrero-Pedraza, A., Moro, N., Fernaândez-Corcuera, P., Amann, B.L., Maristany, T., Vieta, E., Mckenna, P.J., Pomarol-Clote, E., 2017. Evaluation of machine learning algorithms and structural features for optimal MRI-based diagnostic prediction in psychosis. PLoS One 12, 1–24. https://doi.org/10.1371/journal.pone.0175683.

Salvador, R., Canales-Rodríguez, E., Guerrero-Pedraza, A., Sarró, S., Tordesillas-Gutiérrez, D., Maristany, T., Crespo-Facorro, B., McKenna, P., Pomarol-Clotet, E., 2019. Multimodal integration of brain images for MRI-based diagnosis in schizophrenia. Front. Neurosci. 13, 1–9. https://doi.org/10.3389/fnins.2019.01203.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.

Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W., 2001. A review of MRI findings in schizophrenia. Schizophr. Res. 49, 1–52. https://doi.org/10.1016/S0920-9964(01)00163-3.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc, pp. 1–14.

Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps. 2nd Int. Conf. Learn. Represent. ICLR 2014 - Work. (Track Proc).

Smucny, J., Davidson, I., Carter, C.S., 2021. Comparing machine and deep learning-based algorithms for prediction of clinical improvement in psychosis with functional magnetic resonance imaging. Hum. Brain Mapp. 42, 1197–1205. https://doi.org/10.1002/hbm.25286.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: The all convolutional net. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Work. (Track Proc).

Sridharan, D., Levitin, D.J., Menon, V., 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Natl. Acad. Sci. U. S. A. 105, 12569–12574. https://doi.org/10.1073/pnas.0800005105.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conf. Artif. Intell. AAAI, 2017, pp. 4278–4284.

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. Schizophr. Bull. 46, 387–394. https://doi.org/10.1093/schbul/sbz051.

Van Den Heuvel, M.P., Fornito, A., 2014. Brain networks in schizophrenia. Neuropsychol. Rev. 24, 32–48. https://doi.org/10.1007/s11065-014-9248-7.

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., 2018. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. Biol. Psychiatry 84, 644–654.

Wang, L., Kogan, A., Cobia, D., Alpert, K., Kolasny, A., Miller, M.I., Marcus, D., 2013. Northwestern University Schizophrenia Data and Software Tool (NUSDAST). Front. Neuroinform. 7, 1–13. https://doi.org/10.3389/fninf.2013.00025.

Wang, C., Ji, F., Hong, Z., Poh, J.S., Krishnan, R., Lee, J., Rekhi, G., Keefe, R.S.E., Adcock, R.A., Wood, S.J., Fornito, A., Pasternak, O., Chee, M.W.L., Zhou, J., 2016. Disrupted salience network functional connectivity and white-matter microstructure in persons at risk for psychosis: findings from the LYRIKS study. Psychol. Med. 46, 2771–2783. https://doi.org/10.1017/S0033291716001410.

Wang, X., Zhao, N., Shi, J., Wu, Y., Liu, J., Xiao, Q., Hu, J., 2019. Discussion on the application of multi-modal magnetic resonance imaging fusion in schizophrenia. J. Med. Syst. 43 https://doi.org/10.1007/s10916-019-1215-7.

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. Schizophr. Res. 214, 3–10. https://doi.org/10.1016/j.schres.2017.11.038.

Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W.R., David, A.S., Murray, R.M., Bullmore, E. T., 2000. Meta-analysis of regional brain volumes in schizophrenia. Am. J. Psychiatry 157, 16–25. https://doi.org/10.1176/ajp.157.1.16.

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. Neuropsychiatr. Dis. Treat. 15, 935–946. https://doi.org/10.2147/NDT.S195134.

Yang, C., Rangarajan, A., Ranka, S., 2018a. Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. In: AMIA… Annu. Symp. Proceedings. AMIA Symp, 2018, pp. 1571–1580.

Yang, Y., Yan, L.F., Zhang, X., Han, Y., Nan, H.Y., Hu, Y.C., Hu, B., Yan, S.L., Zhang, J., Cheng, D.L., Ge, X.W., Cui, G. Bin, Zhao, D., Wang, W., 2018b. Glioma grading on conventional MR images: a deep learning study with transfer learning. Front. Neurosci. 12, 1–10. https://doi.org/10.3389/fnins.2018.00804.

Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T., 2019. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Trans. Biomed. Eng. 66, 1006–1016. https://doi.org/10.1109/TBME.2018.2866166.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. Springer, pp. 818–833.

Zhang, Q., Zhu, S.-C., 2018. Visual interpretability for deep learning: a survey. Front. Inf. Technol. Electron. Eng. 19, 27–39.

Zheng, F., Li, C., Zhang, D., Cui, D., Wang, Z., Qiu, J., 2019. Study on the sub-regions volume of hippocampus and amygdala in schizophrenia. Quant. Imaging Med. Surg. 9, 1025–1036. https://doi.org/10.21037/qims.2019.05.21.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, pp. 2921–2929. https://doi.org/10.1109/CVPR.2016.319.

Zhu, Z., Albadawy, E., Saha, A., Zhang, J., Harowicz, M.R., Mazurowski, M.A., 2019. Deep learning for identifying radiogenomic associations in breast cancer. Comput. Biol. Med. 109, 85–90. https://doi.org/10.1016/j.compbiomed.2019.04.018.