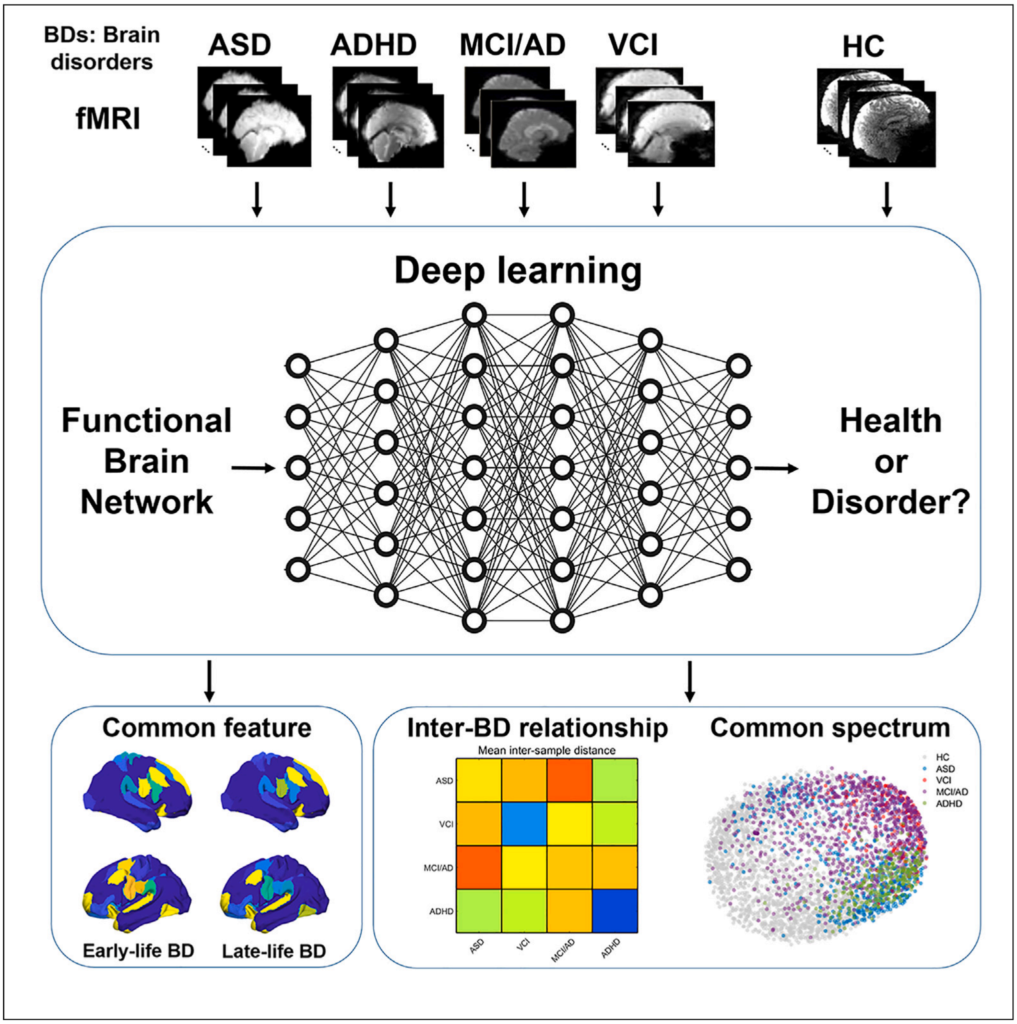
iScience

CellPress
OPEN ACCESS

Article

A common spectrum underlying brain disorders across lifespan revealed by deep learning on brain networks



Mianxin Liu, Jingyang Zhang, Yao Wang, ..., Han Zhang, Qian Wang, Dinggang Shen

16483073@life.hkbu.edu.hk (M.L.)
dinggang.shen@gmail.com (D.S.)

Highlights

A single-branch deep learning model identifies multiple brain disorders using fMRI

A set of commonly affected functional networks across brain disorders is decoded

Data representation of the model reveals a continuous spectrum of brain disorders

The spectrum informs inter-disorders relationships relating to their comorbidities

Liu et al., iScience 26, 108244
November 17, 2023 © 2023 The Author(s).
https://doi.org/10.1016/j.isci.2023.108244

iScience

CellPress
OPEN ACCESS

Article

A common spectrum underlying brain disorders across lifespan revealed by deep learning on brain networks

Mianxin Liu,[1,2,8,*] Jingyang Zhang,[1,8] Yao Wang,[3] Yan Zhou,[3] Fang Xie,[4] Qihao Guo,[5] Feng Shi,[6] Han Zhang,[1] Qian Wang,[1] and Dinggang Shen[1,6,7,9,*]

SUMMARY

Brain disorders in the early and late life of humans potentially share pathological alterations in brain functions. However, the key neuroimaging evidence remains unrevealed for elucidating such commonness and the relationships among these disorders. To explore this puzzle, we build a restricted single-branch deep learning model, using multi-site functional magnetic resonance imaging data (N = 4,410, 6 sites), for classifying 5 different early- and late-life brain disorders from healthy controls (cognitively unimpaired). Our model achieves 62.6 ± 1.9% overall classification accuracy and thus supports us in detecting a set of commonly affected functional subnetworks, including default mode, executive control, visual, and limbic networks. In the deep-layer representation of data, we observe young and aging patients with disorders are continuously distributed, which is in line with the clinical concept of the "spectrum of disorders." The relationships among brain disorders from the revealed spectrum promote the understanding of disorder comorbidities and time associations in the lifespan.

INTRODUCTION

The mental health of children and elders is frequently affected by a wide type of brain disorders (BDs), to which prevention, diagnosis, and treatment remain challenging. With the development of understanding of BDs, the concept of "spectrum" is utilized to integrate different BDs into a unified knowledge framework. Upon a "spectrum," a group of different disorders can share certain features, and their symptoms co-occur or occur on a continuum.[1] Such conceptual tool empowers clinical studies and applications to go beyond the apparent heterogeneity in symptoms and to focus on a set of core biological manifestations, which fosters the understanding of relationships among different disorders, the mutual learning of different fields, and the development of general treatment approaches.

Recently, researchers gradually realized that different BDs in early and late life may be located in their respective spectrums. Autism spectrum disorder (ASD)[2,3] and attention-deficit/hyperactivity disorder (ADHD)[4] are two representative BDs in the early life of humans, respectively, affecting social interaction and attention abilities in typical cases. Studies have started to explore a potential common spectrum underlying ADHD and ASD,[5,6] as they could have similar symptoms, often co-occur with each other,[7] and share certain genetic architectures.[8] Meanwhile, mild cognitive impairment (MCI)[9] and dementia frequently occur in elders, due to Alzheimer's disease (AD),[10] vascular diseases,[11,12] and other etiology. Akin to early-life developmental disorders, cognitive impairments in elders cover a broad range of heterogeneous behavior disabilities and are also associated temporally; for instance, vascular cognitive impairments (VCI) often promote the development of AD.[11,13] Thus, studies on the commonality among these late-life BDs have emerged, with a hypothesis on the existence of another spectrum underlying late-life BDs.[14,15]

Although conventional views regard early development and aging as two dichotomized processes in human life, there has been a debate in the past 20 years on a potentially common neurological process shared by them and the associated BDs.[16,17] First, early- and late-life BDs can show similar cognitive-behavioral symptoms. ASD and AD can both manifest memory deficits, language impairment, visuospatial ability decline, and executive function alteration.[18] Second, early- and late-life BDs exhibit strong temporal connections, even though the lapse may span several decades. Patients with ASD can develop into dementia at 2.6 times more likely when compared to the general population.[19]

[1] School of Biomedical Engineering, State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China
[3] Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200001, China
[4] PET Center, Huashan Hospital, Fudan University, Shanghai 200040, China
[5] Department of Gerontology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China
[6] Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd, Shanghai 200232, China
[7] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
[8] These authors contributed equally
[9] Lead contact
*Correspondence: 16483073@life.hkbu.edu.hk (M.L.), dinggang.shen@gmail.com (D.S.)
https://doi.org/10.1016/j.isci.2023.108244

Third, early- and late-life BDs can have common genetic factors involved in their progressions. A cohort study in Sweden suggested that ADHD and AD are associated across generations, implying common genetic risks shared by them.[20] The common transcriptomic alteration in both early- and late-life BDs has recently been reported.[21] Finally, separated large-scale neuroimaging studies suggested that ASD, ADHD, AD, and other BDs may have the same deficits in the neural functional subsystem in the brain functional networks (BFNs). The "triple networks" are the typically detected common subsystems in different studies,[22] including the default mode network, executive control network, and saliency network.[23–27]

Enlightened by this evidence, when covering ASD, ADHD by an "early-life BD spectrum" and different MCIs and dementias in elders by a "late-life BD spectrum", we are curious about what knowledge we can obtain if the two spectrums are further integrated into an approximate "lifespan BD spectrum" spanning a majority of the time in the human life? In principle, building such a lifespan BD spectrum could potentially offer a common target for treatment or interference of different BDs and a unified viewpoint to theoretically understand relationships among early- and late-life BDs as the basis of their comorbidities. It could also help to integrate and reform the separate fields in aging, development, and different BDs. However, few studies provide strong and direct neuroimaging evidence to explore this hypothesis, while the advance in deep learning (DL) technology provides promise. Advanced DL models have been applied to analyze different BDs with high sensitivity in feature extractions.[28] In addition, the DL model shows a high capacity to represent large-scale data from different sources within the same model architecture.[29] When hypothesizing a set of common neural features among early- and late-life BDs exists, an advanced DL model is promising to automatically identify the common features by learning from a large amount of neuroimaging data on different BDs, as different assessments of the common information. The data representation extracted from the model space will naturally inform about the relationships among early- and late-life BDs in the lifespan BD spectrum.

In this work, we aim to implement a validated DL method to investigate the common neurological factor among early- and late-life BDs in the BFN and explore the relationships of BDs depicted by the lifespan BD spectrum. We build a DL model based on multiscale BFNs from 4410 functional image data, including 2512 data from healthy controls (HCs), and 1898 data from patients suffering from ASD, ADHD, MCI, AD, or VCI. Specifically designed biclassification and transfer learning experiments are performed to demonstrate the existence of common features. Based on deep-layer features of the model, we further investigate the data representation space of multiple BDs for exploration on an integrated lifespan BD spectrum.

RESULTS

A multiscale-BFN-based DL model learns to classify multiple BDs from HCs

We applied our previously established method, the "multiscale atlas-based hierarchical graph convolution network (MAHGCN)",[30] to perform a biclassification between HCs (i.e., cognitively unimpaired) and various BDs and also conduct a transfer learning experiment on functional neuroimaging data from six sites (Figure 1A, $N = 4,410$). Note that the DL method, like multi-head encoder or decoder models,[31,32] can utilize different pathways (and corresponding mappings) inside its architecture to generate the predictions, which potentially influences the extraction of common features and integrated representation space. We therefore restricted the model as a "single-branch architecture," to largely ensure an extraction of one set of features being diagnostic for all BDs in the same representation space. Upon the success of these experiments, we expect to identify common features shared by the classification tasks and provide a unified framework to study these BDs.

The MAHGCN analyzing pipeline is shown in Figure 1C. Briefly, after building the BFNs at different spatial scales based on predefined multiscale atlases and individual fMRI data, the MAHGCN extracts disease-related features from multiscale BFNs, based on stacked graph convolution networks (GCNs) and atlas-guided pooling (AP) operations. Specifically, we implement multiscale atlases from Schaefer et al.,[33] where brains are parcellated into coarse- and fine-scale regions of interest (ROIs), but a similar correspondence to the seven large-scale resting-state functional networks (RSN)[34] is preserved (Figure S1). The RSNs include visual network (VIS), somatomotor network, dorsal attention network, salience network (SAL), limbic network (LIM), executive control network (ECN), and default mode network (DMN). Therefore, the spatial relationships among ROIs in these multiscale atlases can thus be regarded as a biologically meaningful brain hierarchy. Using this prior of the hierarchical relationship between neighboring-scale atlases, the AP is designed to guide nodal feature integration between GCNs. Furthermore, the extracted features from each scale will join the individualized diagnosis decision via skip connections, feature concatenation, and the process of multiple fully connected layers. In our previous works, we demonstrated the capability of this method in optimally classifying AD, MCI, ASD, and VCI from HC, respectively.[30,35,36] From the methodological aspect, our proposed method based on GCN has advantages over MLP and convolutional neural network (CNN) frameworks, where graph topology is preserved and optimally utilized to identify disorder-related features in the brain functional network. In our previous methodology paper,[30] we comprehensively compared our proposed method with advanced GCN methods, multiscale atlas fusion methods, and state-of-the-art CNN methods on brain disorder diagnosis tasks and confirmed its superiority. Clues have been achieved by MAHGCN on shared BFN features among MCI, AD, and VCI.[30,35,36] Therefore, MAHGCN can be a promising choice for this work to effectively classify BD subjects from HC and explore the common BFN features among multiple BDs.

Biclassification experiment

Figure 2A and Table S1 report the quantitative comparison of site-averaged classification results among different methods using a 10-fold cross-validation (see STAR Methods). The competing methods include the single-scale-based GCN (with representative results of using 500 ROIs in Figure 2A and other results of using other scales in Table S1), three data-driven methods (called DIFFPOOL, gPOOL, and SAGPOOL) for building hierarchical GCN, the conventional method (namely MAPGCN) for fusing multiscale BFNs, and our previously

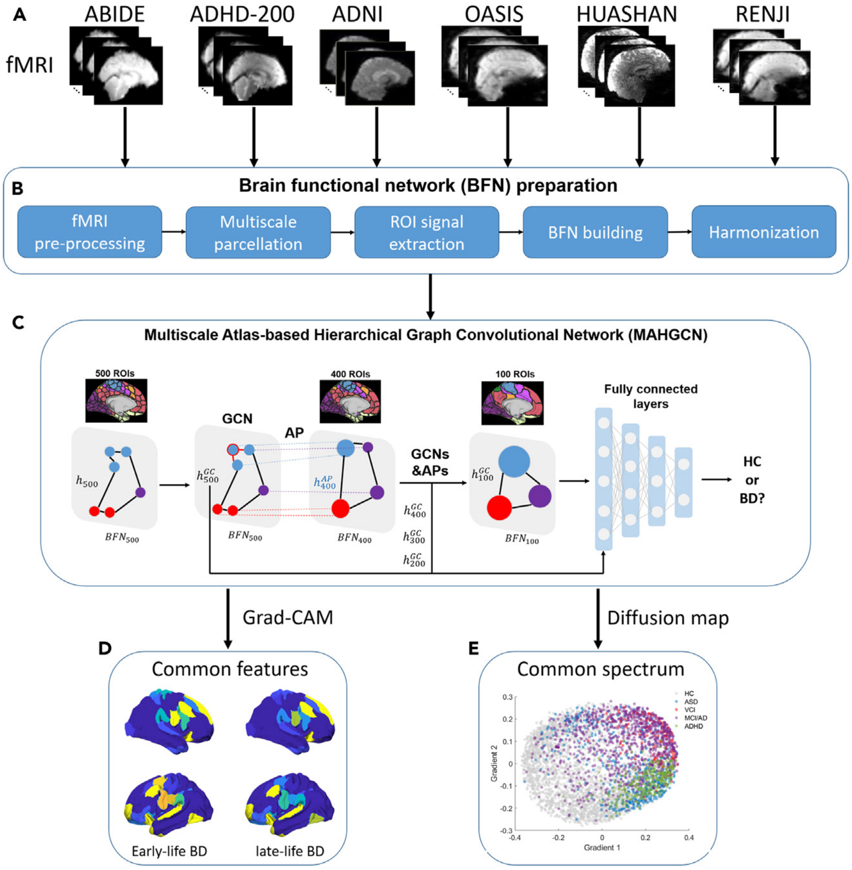


Figure 1. The workflow of the main analysis pipeline

(A) The fMRI data from 6 sites. ABIDE and ADHD-200 datasets are respectively for ASD and ADHD studies. ADNI, OASIS, and HUASHAN datasets are for MCI and AD studies. And RENJI dataset is for the VCI study.

(B) The BFN preparation steps. The construction of multiscale BFNs is based on the multiscale atlases shown in Figure S1.

(C) The neural network structure of the multiscale atlas-based hierarchical graph convolution network (MAHGCN), which hierarchically extracts and integrates features from multiscale BFNs using stacked graph convolutional networks (GCN) and atlas-based pooling (AP) to make the diagnostic decision. The "BD" group includes ASD, ADHD, MCI, AD, and VCI, and the model performs biclassification tasks.

(D) After the model building, the "Grad-CAM" method is used to explore the common features among BDs encoded inside the MAHGCN model.

(E) The "Diffusion map" method is performed to investigate the deep-layer representation of different data in a potentially common spectrum under different BDs.

proposed MAHGCN (see STAR Methods for details of the comparison). The site-averaged metrics define the average prediction performance on each site and the corresponding task. In general, all competing methods achieve 58%–60% accuracies and areas under the curve (AUCs) (Table S1). The three data-driven methods and the conventional multiscale BFN fusion method did not result in improved metrics than single-scale-based GCNs (Figure 2A). This might be reasonable that the noises and variations in BFNs increase the difficulty for a data-driven method to learn a generalizable hierarchical representation from 500-ROI BFN to capture the commonality among different BDs. The MAPGCN processes multiscale BFNs independently in the feature extraction stage and the late-stage fusion may not be capable of reducing the feature redundancy efficiently, introducing risks of overfitting. In contrast, our multiscale-based MAHGCN obtains an accuracy of 62.6 ± 3.4%, a sensitivity of 61.0 ± 8.6%, a specificity of 66.8 ± 3.5%, and an AUC of 63.9 ± 4.3%. The performance from MAHGCN is significantly higher than all the competing methods (Figure 2A; Table S1), underpinning that the MAHGCN could optimally detect generalizable common BFN features of different BDs among all the considered methods.

In addition, Figure 2B and Table S2 offer details of the site-specific diagnostic performance of MAHGCN. Since the HC-to-BD class ratio can fluctuate and be imbalanced in certain sites, we regard sensitivity and AUC as more informative metrics on predictability. For MCI, AD, and VCI, our method obtains AUCs of 68.2 ± 9.5%, 68.9 ± 7.8%, 61.2 ± 3.4%, and 69.4 ± 6.1% in RENJI, HUASHAN, Alzheimer's disease neuroimaging initiative (ADNI), and Open Access Series of Imaging Studies (OASIS) datasets, respectively. For ASD and ADHD, the model results in lower AUCs than MCI, AD, and VCI, with AUCs of 57.8 ± 2.4% in Autism Brain Imaging Data Exchange (ABIDE) and 58.6 ± 4.8% in ADHD-200. In terms of sensitivity, the model achieves 48.3 ± 14.1%, 62.9 ± 16.4%, 61.9 ± 16.1%, 57.2 ± 5.6% 74.9 ± 12.6%, and 58.9 ± 16.0% for

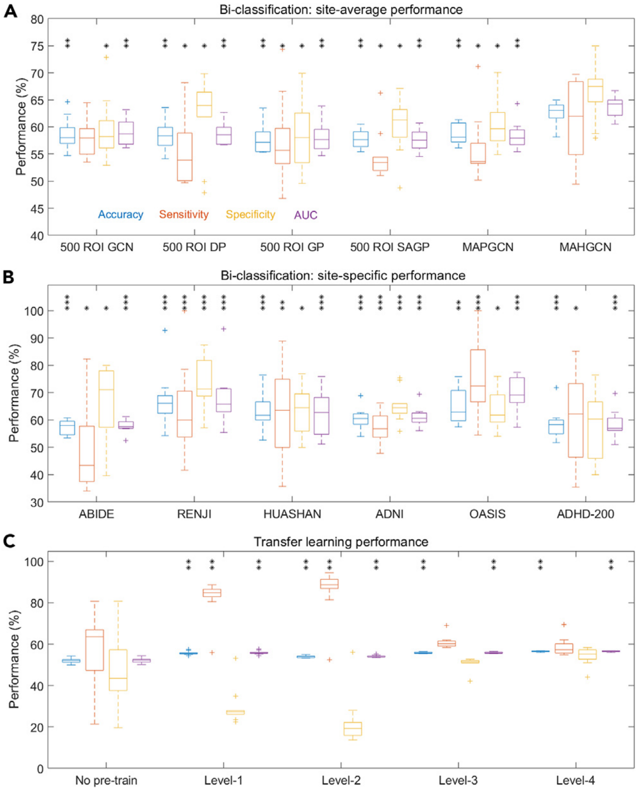


Figure 2. The detailed distributions of performances in different prediction experiments

(A) Boxplots for site-averaged performances of the biclassification experiment by different methods. The symbol * indicates a significantly higher performance of MAHGCN than the competing methods at a significance level of $p < 0.05$. **: $p < 0.01$ and ***: $p < 0.001$ after FDR correction. The one-sided Wilcoxon signed-rank test is used to assess the significance.

(B) Boxplots for site-specific performances of MAHGCN on the biclassification experiment. *: The performance metrics are significantly higher than the chance level at a significance level of $p < 0.05$. **: $p < 0.01$ and ***: $p < 0.001$ after FDR correction. The one-sided Whitney-Mann's U test is used to assess the significance.

(C) Boxplots for the prediction performances in transfer learning experiments under no pre-training and different transfer learning schemes. *: The performance metrics are significantly higher than the baseline performance at a significance level of $p < 0.05$. **: $p < 0.01$ and ***: $p < 0.001$ after FDR correction. The one-sided Wilcoxon signed-rank test is used to assess the significance.

ABIDE, RENJI, HUASHAN, ADNI, OASIS, and ADHD-200. According to permutation tests, the predictabilities in terms of accuracy, sensitivity, and AUC are all significantly higher than the chance level (Table S2, along with Figure S2 for supporting the significance of sensitivity for ASD). Overall, these results validate a certain level of capability of MAHGCN in diagnosing multiple BDs using a single-branch neural network architecture, which suggests that a set of informative common features among multiple BDs in BFNs have been detected by the model.

Transfer learning experiment

We further collect evidence for the shared features among BDs using additional transfer learning experiments. A MAHGCN model is pre-trained using all datasets except ABIDE to learn features for MCI, AD, VCI, and ADHD. If the predictive features are shared, the model is expected to possess certain knowledge about ASD and can be quickly transferred to identify ASD by fine-tuning the model parameters with a small number of training samples (e.g., $N = 20$) from the ABIDE dataset. The pre-trained model should exhibit significantly higher performance than a model without pre-training. Different transfer learning schemes are designed to preserve different levels of learned features during the pre-training (see STAR Methods). A higher-level scheme allows less tuning of the parameters and keeps more learned information. In Figure 2C and Table S3, the predictability in terms of accuracy and AUC generally increases with the level of preservation of the learned features. All pre-trained model gives higher accuracies and AUCs than the non-trained model. The level-4 scheme provides significantly higher accuracy and AUC than the baseline, along with relatively balanced and stable sensitivity and specificity. The level-3 scheme also exhibits increased accuracy, sensitivity, and AUC by roughly 3% when compared to the baseline. In addition, it can be also noted that, in levels 1 and 2, tuning the parameter with less preservation of the pre-trained information will significantly degrade the specificity and case-imbalanced sensitivity and specificity. Results using 50 and 100 training samples from ABIDE are shown in Tables S4 and S5, which also indicate that high levels of preservation of learned information lead to increasing performances. All observations support diagnostic feature sharing of the MCI, AD, VCI, and ADHD with ASD.

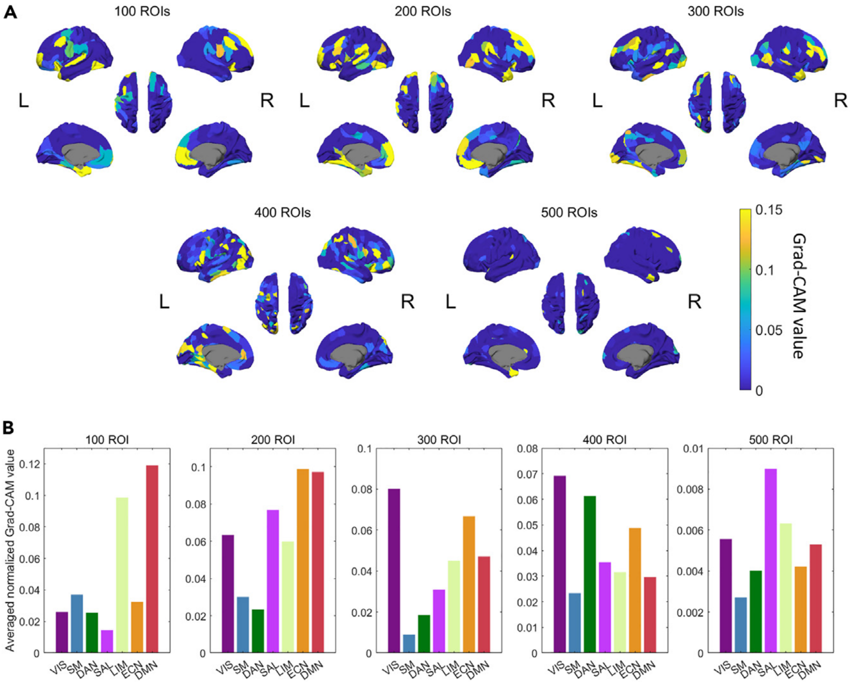


Figure 3. Shared diagnostic brain regions and RSNs for all involved brain disorders learned by the deep learning model

(A) The brain maps showing common diagnostic regions identified by high Grad-CAM values; (B) Bar plots for common diagnostic RSNs indicated by RSN-wise averaged Grad-CAM values. The data distributions behind B are offered in Figure S3A.

One would be interested in further exploring relationships among BDs using a similar transferring learning framework. We perform transfer learning between each pair of BDs, with results provided in Table S6. These results indicate that all configurations of pre-training lead to certain improvements in the mean values of AUCs. However, significant improvement can only be found when ASD or ADHD is the target BD. In addition, the improvement is not symmetric. For example, the significance can be identified during the transfer from VCI to ASD, but not in the reversed way. The confusion may be attributed to differences in levels of overfitting risks when using different pre-training datasets. It is also elusive to discuss inter-BD relationships when the source and target representation spaces may not be fully aligned. Therefore, a clear conclusion could be hard to establish from such analysis. This difficulty highlights the importance of constructing a unified representation space to more reliably explore relationships among BDs.

The common features encoded by DL models indicate shared neurological factors among early- and late-life BDs

We further explore the encoded features in the established models from biclassification experiments, aiming to capture the commonness of various BDs with different etiology in BFNs (Figure 1D). We apply the Grad-CAM method to the DL model and evaluate the features. A high Grad-CAM score indicates a high contribution to the prediction. Joint consideration of features learned in BD populations by models from cross-validations is achieved by using a series of normalization and weighted averaging. As the Grad-CAM values can vary significantly across different models and different datasets due to the multi-site nature, a normalization in value range is, respectively, applied to the Grad-CAM values from different models and different datasets before averaging (see STAR Methods).

First, the normalized Grad-CAM values are averaged over all data to estimate the all-BD common features. In Figure 3A, the detected diagnostic brain regions are not identical on different spatial scales. For 100- and 200-ROI scales, the predictive regions appear in the frontal cortex, while, at 300- and 400-ROI scales, the features are distributed but are largely located in the parietal cortex. Using the language of brain RSNs (Figure 3B, the RSN-wise averaged Grad-CAM value is used to evaluate the predictability of each RSN), the DMN and LIM are the most predictive features for multiple BDs at 100-ROI scales. At the 200-ROI scale, the model relies more on DMN, ECN, and SAL. For 300- and 400-ROI scales, the model regards VIS, VAN, and ECN as the common diagnostic features, while, in the 500-ROI scale, the SAL is highlighted. Note that though we introduced the skip connections in MAHGCN, the gradient value can still remarkably drop in the shallower layers, which influences the inter-scale comparison (Figure 3B). For instance, the 500-ROI BFN is processed by the shallowest GCN layer and could thus be weighted with the least gradient values, leading to the least Grad-CAM values.

In Figures 4, S3, and S4, we investigate the features identified by our model for early- and late-life BDs separately, as the common features in Figure 3 are a mixture of the contributions from different disorders and cannot directly suggest commonness of early- and late-life BDs. In Figures 4A and 4C, the brain maps suggest that, from 100-ROI to 400-ROI scales, the locations of diagnostic brain regions in early- and late-life BDs are quite consistent, despite certain variations in the amplitudes. There is little brain regional consistency on a 500-ROI scale.

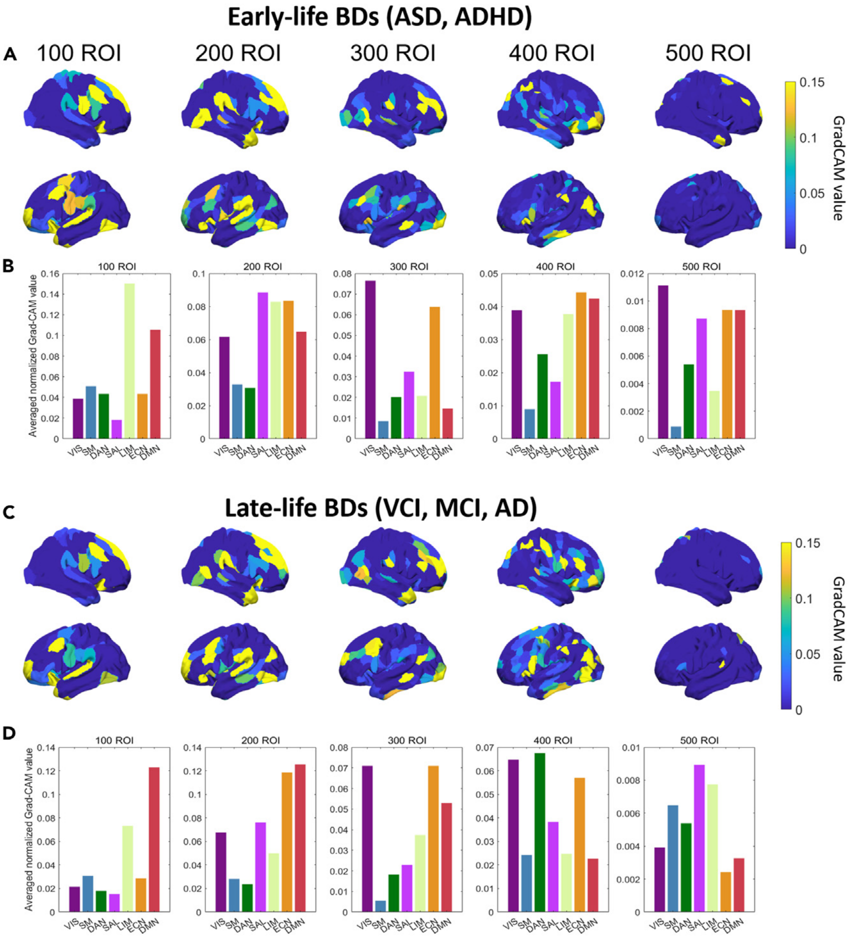


Figure 4. Respective diagnostic brain regions and RSNs for early-life and late-life BDs

(A and B) The brain maps show common diagnostic regions and bar plots for common diagnostic RSNs for early-life BDs (ASD and ADHD data from ABIDE and ADHD-200 datasets).

(C and D) The corresponding plots for late-life BD data from ADNI, RENJI, HUASHAN, and OASIS datasets. Detailed brain patterns are offered in Figures S4 and S5. The detailed data distributions behind B and C are offered in Figures S3B and S3C.

Furthermore, in Figures 4B and 4D, it can be observed that the RSN-level features for early- and late-life BDs at 100-ROI scales agree with the all-BD estimation, which regards the DMN and LIM as the most informative RSNs. For the 200-ROI scale, the common RSNs are DMN and ECN, but note that the identification of early-life BDs relies more on the SAL and LIM. The feature distributions at 300- and 400-ROI scales are relatively stable and consistent with the all-BD estimation, and VIS and ECN can be regarded as commonness. At 500-ROI scales, estimations from different disorders exhibit large variations and identify different crucial RSNs (also different from the all-BD estimation).

The deep-layer presentation of the DL model for BDs suggests inter-BD relationships in a lifespan spectrum

The model has learned to recruit a set of robust common features as characterizing dimensions to perform the classification between HC and various BDs. Moreover, deep-layer representations are capable of representing various disorders in a common space with meaningful structure. Then, we explore this deep-layer representation for an integrated "lifespan BD spectrum" and investigate the inter-BD relationships upon this spectrum (Figure 1E).

We extract inter-subject relationships as the averaged inter-sample correlational distance matrix (Figure S6A) between features from the models' deep layers in the cross-validation. The high-dimensional relationship matrix is decomposed for obtaining the data representation space using diffusion map analysis (see STAR Methods). The distribution of the explained variance of the gradients can be found in Figure S7. According to the principle of the diffusion map algorithm, the closeness among the individual data in the space, spanned by the first two decomposing dimensions (called gradient 1 and gradient 2, respectively), largely informs the similarity among the data in terms of the abstracted RSN features (Figure 5A). It can be first observed that HC and BD data are roughly separately distributed at two ends of the two gradients. We use the ADNI data (covering multiple stages of HC-to-dementia progression) to verify these HC-to-BD gradients encoded

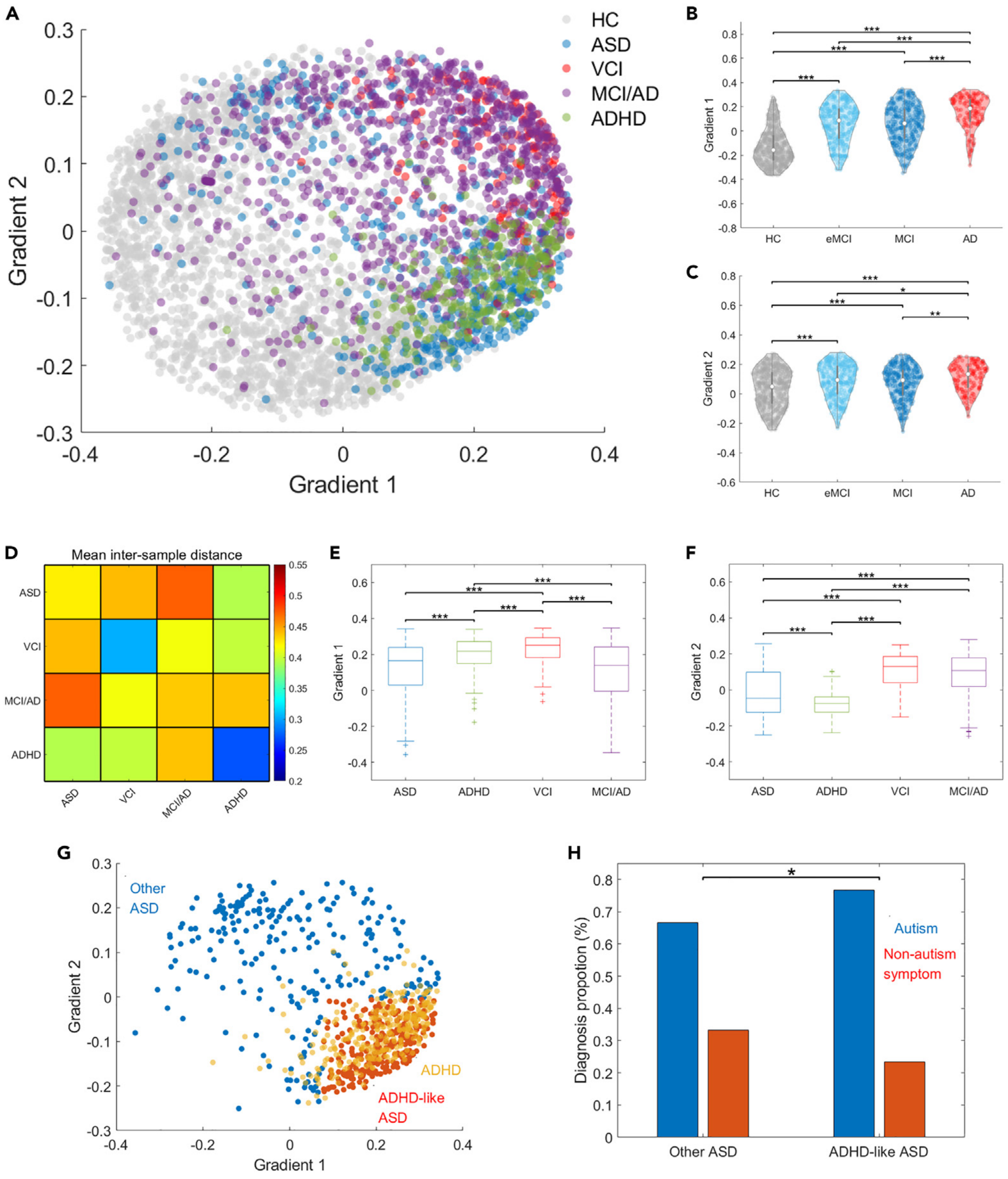


Figure 5. A spectrum-like low-dimensional representation underlying multiple disorders emerges from deep learning and informs the inter-BD relationships

(A) Two-dimensional space representation for the HC and BD data from the diffusion map analysis on the averaged deep-layer feature relationships from DL models.

(B and C) Violin plots for distributions of subjects under AD progression in gradients 1 and 2, respectively.

(D) Averaged inter-sample distance matrix among different BD populations.

(E and F) Boxplots for ASD, ADHD, VCI, and MCI/AD distributions in gradients 1 and 2, respectively. “+” indicates the outliers of the distributions.

(G) The locations of ASD and ADHD data in the two-dimensional space. The ADHD-like ASD (being close to ADHD data) and the other ASD (not belonging to ADHD-like ASD) can be identified.

(H) Bar plots for diagnosis proportions of autism and non-autism diagnosis in ADHD-like and other ASD populations, respectively. In B–G, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Two-sided Whitney-Mann’s U test is used in B–F, and the chi-squared test is used in H to generate the p values.

in our model. In Figures 5B and 5C, it can be found that, along with the increase of values in the gradients, the diagnosis decisions for subjects gradually change from HC to early-stage MCI (eMCI), late-stage MCI (lMCI), and AD, with statistical significance identified. eMCI and lMCI are not significantly differentiable. These observations support the capability of these gradients to correctly encode the HC-to-BD variation trends. Further, as we did not control age and gender when training the model, one would suspect whether the gradients could be associated with these factors besides the HC-to-BD variations. In Figure S8, it can be observed that, when considering all data, the variations along gradient 1 and gradient 2 are associated with age and gender changes, respectively. Gradient 1 is negatively associated with age with a correlation of $r = -0.19$ and $p = 2.7e\text{-}36$, while gradient 2 is positively correlated with age with $r = 0.46$ and $p = 3.8e\text{-}230$. Females exhibit a significantly higher value than males in gradient 1, while there is no significant difference in gradient 2. However, these macroscopic correlations among age, gender, and gradient do not reliably hold when using part of the BD datasets. Therefore, there may not exist generally strong correlations among age, gender, and gradient. Other potential confounding factors, such as head motion, are not considered in this correlational analysis, due to data limitations.

We then quantitatively explore relationships among BDs. First, using the original inter-sample distance matrix, we select the BD samples out of all data (Figure S6B) and compute the averaged inter-sample distance within and between different BD populations (Figure 5D). It can be observed that the within-population distances in VCI and ADHD are relatively small while those in ASD and MCI/AD are large. This is consistent with the compact distribution of VCI and ADHD data, as well as the scattered distribution of ASD and MCI/AD. Both ASD and VCI exhibit the lowest average distance to ADHD, whose values are relatively close. For MCI/AD, the closest BD appears to be VCI, for which the value is even lower than the within-population distance in MCI/AD. Further, we inspect the data distributions of different BD populations in the representation space (Figures 5E and 5F). When ranking BDs based on the population median of the gradient values, gradient 1 depicts a spectrum with the order of MCI/AD, ASD, ADHD, and then VCI, which does not clearly separate early- and late-life BDs. Also, while the mean inter-sample distance between ASD and MCI/AD exhibits the largest value, there is not enough evidence to reject the overlapping (i.e., non-significant separation) between the data distributions of ASD and MCI/AD in gradient 1 (two-sided Whitney-Mann's U test, $p = 1.00$, FDR-corrected). Overall, observations in gradient 1 suggest early- and late-life BDs are connected. Gradient 2 puts ASD and ADHD on one end and MCI/AD and VCI on the other end. There is not enough evidence to reject the overlapping between VCI and MCI/AD in gradient 2 (two-sided Whitney-Mann's U test, $p = 0.0668$, FDR-corrected).

Notably, the ASD data concentrate on two centers in the space, and overlap *not only* the ADHD data *but also* other late-life BD data (VCI, MCI/AD) (Figure 5G). To explore the potential different traits of these ASD subjects, we first define the ASD data falling within the major distribution of ADHD data (i.e., the 10%–100% percentile in gradient 1, and 0%–90% percentile in gradient 2) as "ADHD-like ASD", and all other remaining ASD data as "other ASD". The defined ADHD-like ASDs strongly overlap with ADHD data in gradient 1 (two-sided Whitney-Mann's U test, $p = 0.9848$). Based on the diagnosis results from ABIDE datasets, the proportions of autism and other non-autism diagnoses ("Asperger's syndrome" or "pervasive developmental disorder not specified") within "ADHD-like ASD" and "other ASD" are computed and compared (Figure 5H). The "ADHD-like ASD" population shows a significantly higher proportion of autism as a diagnosis than the "other ASD" population (chi-squared test, $p = 0.0208$). This indicates that the "ADHD-like ASD" population has a higher probability of exhibiting autism as the key symptom, while the "other ASD" population overlapping with other BDs tends to show other manifestations in the ASD.

DISCUSSION

In this work, we build a DL model, based on a multiscale brain functional network, using 4,410 functional magnetic resonance data, to classify healthy (i.e., cognitively unimpaired) populations from ASD, ADHD, MCI, AD, and VCI with 62.6% accuracy and 63.9% AUC, being significantly higher than the chance levels. The results provide a foundation for this study to directly quantify the level of commonality and the relationships among a wide range of BDs in a unified space. Previous studies consider identifying different BDs (or different facets of cognitive abilities) using multiple models working on specific BD or multiple-head architecture under the multi-task framework[31,32] and then checking the common and unique features from different (parts of) models. The classification performances on different sets of BDs are used as indications of the effectiveness of feature extractions, but a nice performance for identifying specific BD could rely more on the extraction of specific features. In addition, the conclusions about the commonality of different BDs could also be unreliable as these frameworks utilize different embedding spaces and did not provide a reasonable basis to discuss relationships among BDs. On the contrary, our work using a single-branch architecture is different from previous studies and prevents limitations of using different or biased embedding spaces. This constraint essentially provides a set of common features and a unified representation space to investigate relationships among BDs with different etiology. By this paper, we hope to draw attention from the field to the study of BD commonality using a similar framework. We expect future studies using advanced methods, such as building large pre-trained models, may obtain higher performances and could validate our presented preliminary findings.

Despite the limited predictability, our model identifies a set of common features in the BFN, such as connectivity abnormalities with DMN, LIM, ECN, and VIS on different spatial scales, which is consistent with previous findings in several independent studies. For example, ASD, ADHD, MCI, AD, and VCI are all found to be associated with abnormal connectivity in the DMN and ECN.[22,37] And, MCI, AD, and VCI are related to damages within LIM.[9,11,23] This consistency supports the effectiveness of our model learning. According to the neuroscience evidence, the DMN and ECN are related to executive ability (sustained attention and working memory),[22,38] the VIS is associated with the processing of visual information, and the LIM is related to memory storage and retrieval.[39–41] Our RSN-level finding also explains the overlapped behavioral symptoms among BDs, as ASD, ADHD, MCI, AD, and VCI could all exhibit executive ability, visuospatial ability, attention, and memory alterations.[18] These observed neurological factors may also suggest a potential common target for drug delivery and other ways of interferences and treatments in future studies. However, regarding the fact that the diagnostic network features emerge on different scales, we admit that one should be careful when interpreting the results. On the one hand, this could be in line with previous findings suggesting that certain BD could affect the functional interactions in specific scales.[42–44] On the other hand, we should also consider that methodological reasons, such as feature redundancy and hierarchical processing of our neural network design, could induce inconsistency across the results from different scales.

In addition, our model learns to represent multiple early- and late-life BDs within a unified space. The inter-sample feature distance analysis and the gradient analysis among the data help to understand the inter-BD relationships in a ''lifespan'' spectrum, connecting disorders with different etiologies. First, we observed the MCI/AD and ASD exhibit high within-population sample distance and occupy a large space without obvious concentration in the spectrum space. Such variation suggests heterogeneity in BFN deficits underlying the MCI/AD and ASD, which is consistent with clinical observations that these brain disorders also show remarkable variations in cognitive manifestations.[9,45] The data distributions of ADHD and VCI are relatively compact and tend to be distributed close to ASD. The distance from MCI/AD to other BDs is relatively even but slightly biased to VCI. These observations, together with the overlap among ASD, VCI, and MCI/AD data distributions in gradients, could be in line with observations that VCI and ASD patients have a higher likelihood of developing AD-type dementia.[13] Also, it may explain the comorbidities of ASD, ADHD, and general cognitive impairments in elders.[6,19] Along the same direction, we explore two sub-populations of ASD exhibiting similarities, respectively, to ADHD and other late-life BDs. The analysis of ''ADHD-like ASD'' and ''other ASD'' finds differences in the frequency of symptoms (diagnoses) between two subpopulations. In theory, this may imply that the heterogeneity in ASD symptoms (diagnoses) could indicate a different tendency to co-occur or develop into different BDs. The lifespan spectrum is thus a new and informative perspective to review ASD and potentially other BDs.

From methodological aspects, the ability of our study to detect the common pathology in brain dynamics also paves the way for building a general brain-disease diagnosis model. As demonstrated in our results, a model transfer is feasible among BDs. In the field of natural language and image processing, pre-training a DL model based on multiple source tasks has become a widely accepted and powerful framework to build generalizable models for multiple downstream tasks, with or without fine-tuning model parameters.[46–48] However, this pre-training framework has not been widely adopted in the brain image analysis field, and a generalizable pre-trained model is still lacking to fit clinical usages. We hope our study can facilitate the exploration in this direction, toward the development of generalizable artificial intelligence tools for brain imaging applications.

Limitations of the study

Firstly, the model predictions suffer from inter-individual variations in large datasets, such as ADNI, ABIDE, and ADHD. The model could have risks of bias toward late-life BDs than early-life BDs. In addition, due to the limitations in data collection protocol, the VCI group could potentially contain subjects with mixed etiology of AD and vascular diseases. What's more, to deal with the multi-site effects, we performed harmonization over all the data before training the model. This study is thus built on a simplified condition. More advanced methods, such as contrastive learning, can be investigated to ease multi-site effects and to enable real-world applications in unseen data. Finally, the currently presented lifespan spectrum could be incomplete as our analysis did not sufficiently include BDs from middle-aged subjects. Further study should be conducted by including the psychiatric disorders frequently observed in the middle ages.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Early-life brain disorders
  - Late-life brain disorders
- METHOD DETAILS
  - fMRI preprocessing
  - Multiscale functional network construction
  - Multi-site data harmonization
  - Deep learning architecture
  - Biclassification experiments
  - Transfer learning experiments
  - Diagnostic feature identification
  - Estimation of the spectrum representation under various BDs and quantification of the inter-BD relationship
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108244.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (grant number 62131015), the Science and Technology Commission of Shanghai Municipality (STCSM) (grant number 21010502600), Shanghai Leading Talent Program (grant number 2022LJ023), the National Natural Science Foundation of China (grant number 82171885), Shanghai Science and Technology Committee Project (Explorer Program: grant number 21TS1400700), and Shanghai Artificial Intelligence Laboratory.

AUTHOR CONTRIBUTIONS

M.L., Q.W., H.Z., F.S., and D.S. designed the research; Y.W., Y.Z., F.X., and Q.G. collected the data; M.L. and J.Z. wrote the code, analyzed the data, and drafted the initial manuscript; all authors revised the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 12, 2023
Revised: September 26, 2023
Accepted: October 16, 2023
Published: October 18, 2023

REFERENCES

1. Maser, J.D., and Akiskal, H.S. (2002). Spectrum concepts in major mental disorders. Psychiatr. Clin. *25*. xi–xiii.
2. Hodges, H., Fealko, C., and Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. Transl. Pediatr. *9*, S55–S65.
3. Lord, C., Brugha, T.S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E.J.H., Jones, R.M., Pickles, A., State, M.W., et al. (2020). Autism spectrum disorder. Nat. Rev. Dis. Prim. *6*, 5–23.
4. Faraone, S.V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J.K., Ramos-Quiroga, J.A., Rohde, L.A., Sonuga-Barke, E.J.S., Tannock, R., and Franke, B. (2015). Attention-deficit/hyperactivity disorder. Nat. Rev. Dis. Prim. *1*, 15020–15023.
5. Hattori, J., Ogino, T., Abiru, K., Nakano, K., Oka, M., and Ohtsuka, Y. (2006). Are pervasive developmental disorders and attention-deficit/hyperactivity disorder distinct disorders? Brain Dev. *28*, 371–374.
6. Frazier, J.A., Biederman, J., Bellordre, C.A., Garfield, S.B., Geller, D.A., Coffey, B.J., and Faraone, S.V. (2001). Should the diagnosis of Attention-Deficit/Hyperactivity Disorder be considered in children with Pervasive Developmental Disorder? J. Atten. Disord. *4*, 203–211.
7. Reiersen, A.M., and Todd, R.D. (2008). Co-occurrence of ADHD and autism spectrum disorders: Phenomenology and treatment. Expert Rev. Neurother. *8*, 657–669.
8. Mattheisen, M., Grove, J., Als, T.D., Martin, J., Voloudakis, G., Meier, S., Demontis, D., Bendl, J., Walters, R., Carey, C.E., et al. (2022). Identification of shared and differentiating genetic architecture for autism spectrum disorder, attention-deficit hyperactivity disorder and case subgroups. Nat. Genet. *54*, 1470–1478.
9. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al. (2006). Mild cognitive impairment. Lancet *367*, 1262–1270.
10. Ter Telgte, A., Van Leijsen, E.M.C., Wiegertjes, K., Klijn, C.J.M., Tuladhar, A.M., and De Leeuw, F.E. (2018). Cerebral small vessel disease: From a focal to a global perspective. Nat. Rev. Neurol. *14*, 387–398.
11. Beishon, L., Haunton, V.J., Panerai, R.B., and Robinson, T.G. (2017). Cerebral Hemodynamics in Mild Cognitive Impairment: A Systematic Review. J. Alzheimers Dis. *59*, 369–385.
12. Román, G.C. (2002). Vascular dementia may be the most common form of dementia in the elderly. J. Neurol. Sci. *203–204*, 7–10.
13. Gorelick, P.B., Scuteri, A., Black, S.E., Decarli, C., Greenberg, S.M., Iadecola, C., Launer, L.J., Laurent, S., Lopez, O.L., Nyenhuis, D., et al. (2011). Vascular contributions to cognitive impairment and dementia: A statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke *42*, 2672–2713.
14. Wilson, H., Pagano, G., and Politis, M. (2019). Dementia spectrum disorders: lessons learnt from decades with PET research. J. Neural. Transm. *126*, 233–251.
15. Emrani, S., Lamar, M., Price, C.C., Wasserman, V., Matusz, E., Au, R., Swenson, R., Nagele, R., Heilman, K.M., and Libon, D.J. (2020). Alzheimer's/Vascular Spectrum Dementia: Classification in Addition to Diagnosis. J. Alzheimers Dis. *73*, 63–71.
16. Schor, N.F., and Bianchi, D.W. (2021). Neurodevelopmental Clues to Neurodegeneration. Pediatr. Neurol. *123*, 67–76.
17. Lesch, K.P., and Mössner, R. (1998). Genetically driven variation in serotonin uptake: is there a link to affective spectrum, neurodevelopmental, and neurodegenerative disorders? Biol. Psychiatr. *44*, 179–192.
18. Khan, S.A., Khan, S.A., Narendra, A.R., Mushtaq, G., Zahran, S.A., Khan, S., and Kamal, M.A. (2016). Alzheimer's Disease and Autistic Spectrum Disorder: Is there any Association? CNS Neurol. Disord.: Drug Targets *15*, 390–402.
19. Vivanti, G., Tao, S., Lyall, K., Robins, D.L., and Shea, L.L. (2021). The prevalence and incidence of early-onset dementia among adults with autism spectrum disorder. Autism Res. *14*, 2189–2199.
20. Zhang, L., Du Rietz, E., Kuja-Halkola, R., Dobrosavljevic, M., Johnell, K., Pedersen, N.L., Larsson, H., and Chang, Z. (2022). Attention-deficit/hyperactivity disorder and Alzheimer's disease and any dementia: A multi-generation cohort study in Sweden. Alzheimers Dement. *18*, 1155–1163.
21. Sadeghi, I., Gispert, J.D., Palumbo, E., Muñoz-Aguirre, M., Wucher, V., D'Argenio, V., Santpere, G., Navarro, A., Guigo, R., and Vilor-Tejedor, N. (2022). Brain transcriptomic profiling reveals common alterations across neurodegenerative and psychiatric disorders. Comput. Struct. Biotechnol. J. *20*, 4549–4561.
22. Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. Trends Cogn. Sci. *15*, 483–506.
23. Veitch, D.P., Weiner, M.W., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Morris, J.C., et al. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement. *15*, 106–152.
24. Padmanabhan, A., Lynch, C.J., Schaer, M., and Menon, V. (2017). The Default Mode Network in Autism. Biol. Psychiatry. Cogn. Neurosci. Neuroimaging *2*, 476–486.
25. Uddin, L.Q., Kelly, A.M.C., Biswal, B.B., Margulies, D.S., Shehzad, Z., Shaw, D., Ghaffari, M., Rotrosen, J., Adler, L.A., Castellanos, F.X., and Milham, M.P. (2008). Network homogeneity reveals decreased integrity of default-mode network in ADHD. J. Neurosci. Methods *169*, 249–254.
26. Liddle, E.B., Hollis, C., Batty, M.J., Groom, M.J., Totman, J.J., Liotti, M., Scerif, G., and Liddle, P.F. (2011). Task-related default mode network modulation and inhibitory control in ADHD: effects of motivation and methylphenidate. JCPP (J. Child Psychol. Psychiatry) *52*, 761–771.
27. Mohan, A., Roberto, A.J., Mohan, A., Lorenzo, A., Jones, K., Carney, M.J., Liogier-Weyback, L., Hwang, S., and Lapidus, K.A.B. (2016). Focus: The Aging Brain: The Significance of the Default Mode Network (DMN) in Neurological and Neuropsychiatric

Disorders: A Review. Yale J. Biol. Med. *89*, 49–57.

28. Zhang, L., Wang, M., Liu, M., and Zhang, D. (2020). A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis. Front. Neurosci. *14*, 779.
29. Zhang, L., Xie, Y., Xidao, L., and Zhang, X. (2018). Multi-source heterogeneous data fusion. In 2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018 47–51 (Institute of Electrical and Electronics Engineers Inc.). https://doi.org/10.1109/ICAIBD.2018.8396165.
30. Liu, M., Zhang, H., Shi, F., and Shen, D. (2023). Hierarchical Graph Convolutional Network Built by Multiscale Atlases for Brain Disorder Diagnosis Using Functional Connectivity. IEEE Transact. Neural Networks Learn. Syst. 1–13. https://doi.org/10.1109/TNNLS.2023.3282961.
31. He, T., An, L., Chen, P., Chen, J., Feng, J., Bzdok, D., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2022). Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. Nat. Neurosci. *25*, 795–804.
32. Liu, M., Zhang, J., Adeli, E., and Shen, D. (2019). Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis. IEEE Trans. Biomed. Eng. *66*, 1195–1206.
33. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebr. Cortex *28*, 3095–3114.
34. Thomas Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.
35. Liu, M., Zhang, H., Shi, F., and Shen, D. (2021). Building Dynamic Hierarchical Brain Networks and Capturing Transient Meta-states for Early Mild Cognitive Impairment Diagnosis. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021, *vol. Part VII* (Springer), pp. 574–583.
36. Liu, M., Wang, Y., Zhang, H., Yang, Q., Shi, F., Zhou, Y., and Shen, D. (2022). Multiscale functional connectome abnormality predicts cognitive outcomes in subcortical ischemic vascular disease. Cerebr. Cortex *32*, 4641–4656.
37. Sridharan, D., Levitin, D.J., and 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. USA *105*, 12569–12574.
38. Buckner, R.L., Andrews-Hanna, J.R., and Schacter, D.L. (2008). The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. *1124*, 1–38.
39. Frey, S., and Petrides, M. (2002). Orbitofrontal cortex and memory formation. Neuron *36*, 171–176.
40. Barbey, A.K., Koenigs, M., and Grafman, J. (2011). Orbitofrontal contributions to human working memory. Cerebr. Cortex *21*, 789–795.
41. Farovik, A., Place, R.J., McKenzie, S., Porter, B., Munro, C.E., and Eichenbaum, H. (2015). Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. J. Neurosci. *35*, 8333–8344.
42. Engel, T.A., Schölvinck, M.L., and Lewis, C.M. (2021). The diversity and specificity of functional connectivity across spatial and temporal scales. Neuroimage *245*, 118692.
43. Betzel, R.F., and Bassett, D.S. (2017). Multi-scale brain networks. Neuroimage *160*, 73–83.
44. Betzel, R.F., Bertolero, M.A., Gordon, E.M., Gratton, C., Dosenbach, N.U.F., and Bassett, D.S. (2019). The community structure of functional brain networks exhibits scale-specific patterns of inter- and intra-subject variability. Neuroimage *202*, 115990.
45. Díaz-Mardomingo, M., García-Herranz, S., Rodríguez-Fernández, R., Venero, C., and Peraita, H. (2017). Problems in classifying mild cognitive impairment (MCI): One or multiple syndromes? Brain Sci. *7*, 111.
46. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., and Gao, W. (2020). Pre-Trained Image Processing Transformer. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE Computer Society), pp. 12294–12305.
47. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems (Neural information processing systems foundation), pp. 1877–1901.
48. Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference (Association for Computational Linguistics (ACL)), pp. 4171–4186.
49. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatr. *19*, 659–667.
50. Milham, P.M., Damien, F., Maarten, M., and Stewart, H.M. (2012). The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. Front. Syst. Neurosci. *6*, 1–5.
51. Jack, C.R., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., et al. (2015). Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2. Alzheimers Dement. *11*, 740–756.
52. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., and Buckner, R.L. (2007). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cognit. Neurosci. *19*, 1498–1507.
53. Ding, D., Zhao, Q., Guo, Q., Liang, X., Luo, J., Yu, L., Zheng, L., and Hong, Z.; Shanghai Aging Study SAS (2016). Progression and predictors of mild cognitive impairment in Chinese elderly: A prospective follow-up in the Shanghai Aging Study. Alzheimers Dement. *4*, 28–36.
54. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.
55. Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.J., Langs, G., Valk, S., et al. (2020). BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Commun. Biol. *3*, 103–110.
56. Wang, Y., Tu, D., Du, J., Han, X., Sun, Y., Xu, Q., Zhai, G., and Zhou, Y. (2019). Classification of subcortical vascular cognitive impairment using single MRI sequence and deep learning convolutional neural networks. Front. Neurosci. *13*, 627.
57. Cox, R.W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. *29*, 162–173.
58. Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: A MATLAB toolbox for 'pipeline' data analysis of resting-state fMRI. Front. Syst. Neurosci. *4*, 13.
59. Orlhac, F., Eertink, J.J., Cottereau, A.S., Zijlstra, J.M., Thieblemont, C., Meignan, M., Boellaard, R., and Buvat, I. (2022). A guide to ComBat harmonization of imaging biomarkers in multicenter studies. J. Nucl. Med. *63*, 172–179.
60. Kipf, T.N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1609.02907.
61. Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.
62. Ying, R., Morris, C., Hamilton, W.L., You, J., Ren, X., and Leskovec, J. (2018). Hierarchical graph representation learning with differentiable pooling. In Advances in Neural Information Processing Systems (Neural information processing systems foundation), pp. 4805–4815.
63. Gao, H., and Ji, S. (2022). Graph U-Nets. IEEE Trans. Pattern Anal. Mach. Intell. *44*, 4948–4960. https://doi.org/10.1109/TPAMI.2021.3081010.
64. Lee, J., Lee, I., and Kang, J. (2019). Self-attention graph pooling. In 36th International Conference on Machine Learning, ICML 2019 (PMLR), pp. 6661–6670.
65. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (IEEE), pp. 618–626.
66. Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. USA *113*, 12574–12579.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
ABIDE dataset	Di Martino et al.[49]	https://fcon_1000.projects.nitrc.org/indi/abide/
ADHD-200 dataset	Milham et al.[50]	http://fcon_1000.projects.nitrc.org/indi/adhd200/
ADNI dataset	Jack et al.[51]	http://adni.loni.usc.edu/
OASIS dataset	Marcus et al.[52]	https://www.oasis-brains.org/
HUASHAN dataset	Ding et al.[53]	https://doi.org/10.1016/j.dadm.2016.03.004
RENJI dataset	Liu et al.[36]	https://doi.org/10.1093/cercor/bhab507
Software and algorithms		
MATLAB R2020b	Mathworks	https://www.mathworks.com/
Python V3.6	Python Software Foundation	https://www.python.org
Deep learning algorithms	This paper	https://github.com/MianxinLiu/MAHGCN-code/tree/main/multisite
Combat	Johnson et al.[54]	https://github.com/Jfortin1/ComBatHarmonization
Diffusion map	Vos de Wael.[55]	http://github.com/MICA-MNI/BrainSpace

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Dinggang Shen (Dinggang.Shen@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data from ADNI, OASIS, ABIDE, and ADHD-200 datasets are publicly available. Data from RENJI and HUASHAN datasets are available from the lead contact upon reasonable request due to privacy restrictions. DOIs/URLs are listed in the key resources table.
- All original code has been deposited at GitHub and is publicly available as of the date of publication; URLs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

From four public datasets and two private datasets, we include 4,410 data for training the model, which contains 2512 healthy controls (HC. More specifically we refer to "cognitively unimpaired" subjects) and 1898 brain disorder (BD) subjects. The corresponding demographic information is provided in Table S7.

Early-life brain disorders

The Autism Brain Imaging Data Exchange (ABIDE) and ADHD-200 datasets, containing neuroimaging from ASD and ADHD subjects, are used for early-life BD identifications.

ABIDE

From the Autism Brain Imaging Data Exchange (ABIDE-I),[49] we select scans with a duration longer than 300s, yielding 512 HC and 499 ASD subjects. As a multi-site dataset, the acquisition protocols and diagnostic criteria in ABIDE vary according to data collection sites (16 scan protocols among the sites). Overall, the fMRI scanning parameters are: TR = 1.5-3 s, TE = 15-33 ms, in-plane resolution $3 \times 3\text{-}3.438 \times 3.438$ mm^2, slice thickness 3-4.5 mm, 28-40 axial slices, and 304-486 s in duration (120-300 volumes). A detailed protocol can be found at https://fcon_1000.projects.nitrc.org/indi/abide/.

ADHD-200

From ADHD-200,[50] we use fMRI scans from 488 HC and 280 ADHD subjects, sampled from 8 sampling sites. Again, the acquisition protocols significantly vary (9 scan protocols): TR = 1.5-2.5 s, TE = 15-40 ms, in-plane resolution 3 × 3-3.8× 3.8 mm^2, slice thickness 3.0-4.0 mm, and 29-47 axial slices. The specific scanning parameters can be found at http://fcon_1000.projects.nitrc.org/indi/adhd200/.

Late-life brain disorders

The Alzheimer's disease neuroimaging initiative (ADNI), Open Access Series of Imaging Studies (OASIS), and an in-house HUASHAN dataset contain neuroimaging data from MCI (referring to the prodromal state of AD) or AD elderly subjects. And the in-house RENJI dataset contains data from subjects with VCI. The four datasets are used for investigating late-life BDs.

ADNI

For the Alzheimer's disease neuroimaging initiative (ADNI) dataset,[51] a total of 1350 fMRI data are selected, which contains 565 HC and 785 MCI or AD subjects. Each fMRI data is acquired with TR = 3 s, TE = 30 ms, resolution = 3.3× 3.3× 3.3 mm^3, 48 axial slices, and 420 s in duration (140 volumes). A detailed protocol can be found at http://adni.loni.usc.edu/.

OASIS

In the Open Access Series of Imaging Studies (OASIS) dataset,[52] we include 634 HC and 83 MCI or AD subjects in the study. The fMRI acquisition protocols are TR = 2.2 s, TE = 27 ms, resolution = 4× 4 × 4 mm^3, 36 axial slices, and 372 s in duration (169 volumes). More information can be found at https://www.oasis-brains.org/.

HUASHAN

fMRI data from 167 HC and 100 MCI or AD subjects were obtained from Huashan Hospital in Shanghai.[53] fMRI scans are obtained by using a multi-slice single-shot gradient echo-planar imaging sequence: TR = 0.8 s, TE = 37 ms, resolution = 2 × 2 × 2mm^3, 72 axial slices, and 390.4 s in duration (488 volumes). The participants are instructed to close their eyes but remain awake during the scanning.

RENJI

fMRI data from 146 HC and 151 VCI subjects were obtained from Renji Hospital in Shanghai. MRI scan is performed using a SignaHDxt 3T MRI scanner (GE Healthcare, United States), with an eight-channel standard head coil with foam paddings to restrict head motions. The parameters of the echo-planar imaging sequence for the resting-state fMRI data collection are as follows: TR = 2 s, TE = 24 ms, resolution = 2 × 2 × 2 mm^3, 34 axial slices, and 440 s in duration (220 volumes). The diagnostic criteria are reported in our previous publications.[36,56] In brief, firstly, two experienced radiologists identified the subcortical ischemic vascular disease (SIVD) by detecting white matter lesions as at least one lacunar infarct on the T2-FLAIR image. Further, within SIVD population, a battery of neuropsychological tests is used to test the cognitive abilities of participants, which covers attention, executive function, memory, language, and visuospatial function. The population distribution of the scores was constructed based on the scores for each measure of normal-aged people in Shanghai, China. The SIVD participants with scores falling beyond $\pm$1.5 standard deviations from the mean are regarded as VCI (MCI or dementia). Note that we utilize all participants without cognitive impairments as HC (cognitively unimpaired) even if some of them exhibit SIVD, which could be reasonable as the brain functional network could be normal (and thus the cognition is normal) under brain lesions. The VCI subjects did not receive positron emission tomography (PET) scans to exclude the AD-related pathology, and thus the VCI group could potentially contain subjects with mixed etiology of both vascular disease and AD.

METHOD DETAILS

fMRI preprocessing

We apply well-accepted toolboxes, AFNI[57] (for ADNI) and DPARSF[58] (for ABIDE, OASIS, HUASHAN, and RENJI datasets), to perform a standardized preprocessing procedure for fMRI data. In particular, the first several volumes (5–10, the exact number varies across datasets) of each image are discarded due to potential non-equilibrium magnetization. The slice timing correction is done except for the HUASHAN dataset as the data was sampled with high temporal resolution. The rigid-body transformation is performed to correct the subject's head motion. Subjects with large head motions are excluded. We do not further perform scrubbing/censoring of data as it may introduce additional artifacts. The signals of white matter, cerebrospinal fluid, and head motion are regarded as nuisance covariates and are regressed out from individual data. The fMRI images are then normalized to the Montreal Neurological Institute (MNI) space and spatially smoothed with a Gaussian kernel with full width at half maximum (FWHM) of 4 × 4 × 4 mm^3. The BOLD signals are further band-pass filtered ($0.01 \leq f \leq 0.1$ Hz) to remove the neural-irrelevant high-frequency noises and low-frequency drift from the MRI machine. For the ABIDE dataset, since volumes of scans are different among the collecting sites, we use its minimum common length, i.e., 115 volumes, around the middle volume of the preprocessed fMRI sequence for further processing. We use preprocessed data for ADHD-200 provided in http://preprocessed-connectomes-project.org/adhd200/, using the Athena pipeline.

Multiscale functional network construction

Schaefer et al. provided a set of atlases for multiscale brain parcellation,[33] which are used in this paper for generating multiscale BFNs and guiding the node pooling across scales. The atlases are generated by FC-pattern-based clustering on voxel (or vertices) by considering both global similarity and spatial proximity. Clustering in different resolutions results in brain functional parcellations at multiple scales, ranging from 100 to 1000 regions of interest (ROIs). It can be observed that the seven RSN structures[34] are largely preserved after parcellation at all scales (Figure 1A). Therefore, the atlases at different scales can be viewed as coarse-to-fine parcellation of the seven RSNs. The spatial relationship among the ROIs in these atlases at different scales thus characterizes a biologically meaningful functional hierarchy.

Given the atlas at a specific scale, the ROI-level signals can be obtained by averaging voxel-level BOLD signals within each ROI. The BFN at the given scale S is then computed by Pearson correlation among all pairs of ROI-level signals and is denoted as BFN_S. Consistent with our previous study,[30] we use the first five scales, i.e., from 100 to 500 ROIs.

Multi-site data harmonization

As we aim to explore the commonness and relationships among different BDs, the data distribution shifting caused by the multi-site effect is a nuisance factor to be removed A statistical regression-based harmonization method, called "Combat",[54,59] is thus applied to calibrate the BFN data. The codes are publicly available at https://github.com/Jfortin1/ComBatHarmonization. In Combat, with a linear regression model, the variation of each functional connectivity across individuals is modeled as the sum of essential mean, effects of biological co-variates (i.e., age, gender, and brain BDs), site-related bias in mean, and site-related noise level. Therefore, functional connectivity without site effect can be calculated by estimating the parameters of the regression model from data and removing site-related bias.

In this paper, we use a scanner-based harmonization since ABIDE and ADHD-200 contain data from multiple sites (ABIDE: 16 scanners; ADHD: 9 scanners; In total, 28 scan protocols for all data). We preserve the effect of age, gender, and type of BDs.

The Combat is performed over all data. This may violate the conventional setting for an "out-of-sample" test as the testing data has been first integrated with the training data during applications. However, as the effects of different brain disorders are theoretically preserved, the analysis still ensures a fair test on whether one model could classify different brain disorders in the biclassification task and could perform an "out-of-disorder test" in the transfer learning task. The primate scientific question can still be sufficiently explored.

Deep learning architecture

The Multiscale-Atlas-based Hierarchical Graph Convolutional Neural Network (MAHGCN) is proposed and systematically tested in our previous study.[30] Here we briefly review two crucial building blocks of MAHGCN, i.e., graph convolutional network (GCN) and the atlas-guided pooling (AP). The MAHGCN is then built by hierarchically stacking GCNs and APs (Figure 1C), together with the skip connections and fully-connected layers (FLs).

Graph convolutional network

The graph convolutional network (GCN)[60] is an effective deep-learning method to abstract features from graph data (e.g., the BFN data). It completes the convolutional operations via two steps, i) propagating nodal features via graph Laplacian, and ii) selecting features by applying a learned kernel on the features. Formally, for a given adjacency matrix A and nodal features h, one graph convolution layer updates the nodal feature by following the equation below:

$$h^{GC} = GC\,(A,h) = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}hW\right), \qquad \text{(Equation 1)}$$

where $\tilde{A} = A + I$, I is the identity matrix, $\tilde{D}$ is the corresponding degree matrix of $\tilde{A}$, W is the estimated kernel weight matrix, and $\sigma(\cdot)$ is a nonlinear activation function. Empirically, we skip the computation of graph Laplacian and directly use the adjacency matrix BFN_S from scale S to obtain optimal diagnosis performance.

Atlas-guided pooling

The AP operation is defined according to spatial overlapping among ROIs informed by atlases at different scales. The AP benefits information integration and introduces inter-scale dependency during feature extraction. It aims to convert the nodal features defined by the atlas at scale P into the nodal features for the atlas at scale Q ($P > Q$), based on the mapping matrix $M_{P\to Q}$:

$$M_{\mathcal{R}\to\wp}(i,j) = \begin{cases} 1, \rho > Th \\ 0, \text{Otherwise} \end{cases}, \qquad \text{(Equation 2)}$$

where the overlapping ratio ρ is computed by size (i.e., the number of voxels) of spatially overlapping between ROI i in the atlas at scale P and ROI j in the atlas at scale Q divided by the size of ROI i. And Th is a threshold applied to ρ for defining elements in $M_{P\to Q}$. We use $Th = 0$ according to the results in our previous methodological paper.[30] Through a matrix multiplication with $M_{P\to Q}$, a feature map h_P^{GC} defined in the atlas at scale P from GCN is converted into a new feature map h_Q^{AP} for the atlas at scale Q.

Implementation

All models are implemented using the open-source framework "Pytorch" in Python. We choose the ReLU function as the non-linear activation function for both GCN and FLs. An identity matrix (node×feature dimension, $S \times S$) is used as the initial nodal feature to make the MAHGCN model focus on the topology of the BFN_S. Therefore, for each GCN, the input channel size is set as the node number of the graph (i.e. scale S) and the output channel size as 1. For building a stacked network, we use the outputted feature from the previous GCN and AP layer as the diagonal elements in a new diagonal matrix to resume the feature dimension from 1 to the matched scale. The GCN layers in MAHGCN are attached with dropout functions (rate = 0.2), and the last GCN layer is followed by four FLs, whose output channel sizes are decreasing (i.e. 512, 256, 128, 2). Each FL is associated with a batch normalization and a ReLU activation function. The outputs from the last (the 4th) FL are normalized by a Softmax function to generate the diagnostic probabilities for two classes. These configurations for GCNs and FLs are kept consistent in the single-scale-based GCN methods. The detailed implementations of all models can be accessed in our open codes (https://github.com/MianxinLiu/MAHGCN-code/tree/main/multisite).

Biclassification experiments

The MAHGCN model first performs a conventional biclassification experiment using all BDs and HCs. The model is restricted to be a single-branch architecture and thus forced to extract one set of features being diagnostic for all BDs. Thus, a successful classification demonstrates common features among all BDs.

Training scheme

Since sample size and class ratio (i.e., HC-vs-BD ratio) are different in each dataset, a site-specific weight and a cross-entropy loss function are used to supervise the training process. The weighted cross-entropy loss is based on the inverse of the HC-vs-disorder ratio for each site, estimated in the training samples. For each update iteration, we randomly sample (equally, 100 samples) from each set, which are inputted to the model to calculate their site-specific losses, respectively. The yielded site-specific cross-entropy loss is further multiplied with a penalty designed by the square root of the inverse of the site sample size for re-weighting. All re-weighted site-specific losses are accumulated with the linear summation, based on which the model parameters are finally updated.

The training parameters for neural network models are identically set as training epoch = 150, and learning rate = 0.01 for the first 50 epochs and then 0.001 for the remaining epochs. Adam[61] with a weight decay of 0.01 is used as an optimizer. Other parameters of the neural network models are initialized with random weights with the default setting of Pytorch.

Comparison methods

First, the single-scale GCN method is compared as baselines. The results from 500-ROI BFN are shown in the main text while results from other scales are offered in Table S1. Secondly, three prevalent methods,[62–64] namely DIFFPOOL (DP), gPOOL (GP), and SAGPOOL (SAGP), building stacked GCN by learning the hierarchical representation from the graph are compared. In contrast, our MAHGCN builds stacked GCNs based on priors from atlases without learning. To obtain reasonable comparison results, we apply DP, GP, and SAGP to work on the 500-ROI BFN and generate the hierarchical representations in 400-, 300-, 200-, and 100-ROI scales, in the same manner as the structure of MAHGCN with five scales. Besides similar processing of stacked GCN and pooling, skip connections and four FLs with the same configurations are implemented. Finally, we compared our MAHGCN with a conventional method to integrate multiscale BFNs. We train multiple GCNs to parallel process the inputs of multiscale BFNs, whose outputs are concatenated and fused via the following FLs. As this network design implemented parallel processing rather than hierarchical processing on multiscale BFNs, we call this method a "multi-scale-atlas-based parallel GCN (MAPGCN)".

Validation scheme

A classical ten-fold cross-validation is performed. The data is randomly shuffled and equally split into ten folds. In each round of cross-validation, nine folds of data will be used as training samples and the remaining one as testing samples. Ten rounds of cross-validation are performed until all folds play as testing samples once. Four metrics are adopted to evaluate performance in the testing samples, i.e., accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC). Since sample sizes in different disorders from different sites are significantly varying, computing "global" statistics simply as the ratio of correct predictions against all samples will assign larger weights to the sites with larger sample sizes. We thus compute four performance metrics for each site ("site-specific" statistics) and then average over all sites ("site-averaged" statistics) for each cross-validation. The mean and standard deviation of "site-specific" statistics and "site-averaged" statistics from cross-validation are reported.

Transfer learning experiments

We tested whether a model pre-trained using all data except the ABIDE dataset (N=3399) can be transferred to perform ASD identification in ABIDE data (N=1011) with restricted samples. This experiment aims to provide additional evidence for the common features under different BDs. We also explore the transfer learning between each pair of BDs using a similar framework.

Training scheme

The model is trained on the five datasets until it converges with 250 epochs. This pre-trained model is used as an initial model and further fine-tuned using training samples from ABIDE with 50 epochs. For both pre-training and fine-tuning, other configurations are the same as the settings in biclassification experiments. In addition, four levels of fine-tuning schemes are designed to test the model with different amounts of preservation of the learned information during the pre-training. "Level 1" refers to fine-tuning all model parameters. "Level 2" refers to fine-tuning all FLs and batch normalization layers (BN). "Level 3" refers to fine-tuning the last FL and all the BNs in the model. "Level 4" refers to fine-tuning only the last FL and the last BN. Intuitively, higher level fine-tuning preserves more learned information during the pre-training.

Validation scheme

A ten-fold "K-shot" cross-validation is performed. For each round of cross-validation, the data are shuffled and split into training (N=100) and testing sets (N=911). For the K-shot condition, training samples are the first K samples in the training set. In the main text, results using a 20-shot condition are depicted. In Tables S4 and S5, we offer the results under 50 shots and 100 shots, which are consistent with the results under the 20-shot condition. The mean and standard deviation of ACC, SEN, SPE, and AUC in the testing set are used to assess the performance.

Diagnostic feature identification

To reveal the predictive features of deep learning methods, we utilize a Gradient-guided Class Activation Map (Grad-CAM) algorithm[65] and analyze the established biclassification models. In short, the Grad-CAM regards the gradient between prediction outputs and the feature maps at intermediate hidden layers (in this work, we used features from intermedia GCN layers for each scale) of the deep neural network as the importance of features. This thus applies gradient values to weight elements in the feature maps, i.e., the product between the gradient map and feature map, which offers a visual map for spotting predictive features.

To investigate the common features of BDs, the Grad-CAM from correctly predicted BD subjects is extracted using different models from cross-validations. However, the Grad-CAM values can vary significantly across different models and different datasets due to the nature of using multi-sites. Therefore, we designed a double normalization procedure to relieve the Grad-CAM value heterogeneity across the models and datasets to spot the common features more properly. First, all Grad-CAM values from a given model are normalized into a range from zero to one according to the minimum and maximum values. Then, all normalized Grad-CAM from subjects belonging to different datasets is averaged respectively. To obtain a joint estimation of the models from cross-validations, we utilize the prediction AUCs to perform a weighted average on the normalized Grad-CAM. In this way, the normalized Grad-CAMs for every specific dataset are established. We again normalize these dataset-specific Grad-CAMs into a range from zero to one to address amplitude differences in Grad-CAMs across datasets. For the all-BD common features (Figure 3), we average all six double-normalized Grad-CAMs. For common features of early-life BDs (Figure 4A), the double-normalized Grad-CAMs from ABIDE and ADHD-200 are averaged. For common features of late-life BDs (Figure 4B), the double-normalized Grad-CAMs from ADNI, OASIS, HUASHAN, and RENJI are averaged.

Estimation of the spectrum representation under various BDs and quantification of the inter-BD relationship

To explore the potential common spectrum under different BDs and the inter-BD relationships on it, we investigate deep-layer data representations from the established biclassification models during cross-validations. We regard each model as one expert of the inter-sample relationship and integrate the inter-sample relationship, rather than the feature values, from each model's latent space based on their prediction performances. Based on the integrated inter-sample relationship matrix, individual data are embedded into a low-dimensional Euclidean space with relationship preservation for visualization and quantitative analyses. By this approach, the inter-BD relationships can be optimally preserved and explored.

Operationally, the encoded features of individuals are extracted from third-layer FL, and the sample relationships are computed by the correlation distance based on these features. The inter-sample relationships from models under different rounds of validation are then weighted-averaged using AUCs to provide an integrated estimation. We then use the diffusion map method to embed the averaged relationship matrix into an Euclidean space, which is in alignment with the brain gradient analysis.[66] The diffusion map method estimates a non-linear mapping of the data into a new low-dimensional Euclidean space to ensure a distance-preserved mapping, so that the Euclidean distances among individuals in the mapped space roughly keep the original distances reflected in the sample relationship matrix. The implementation of the diffusion map is based on the open-source "BrainSpace" toolbox (http://github.com/MICA-MNI/BrainSpace),[55] with default settings. The diffusion map is usually regarded as the advanced non-linear version of the conventional multi-dimensional scaling (cMDS) method. We have tested cMDS, whose results are depicted in Figure S9 and are similar to those from the diffusion map.

QUANTIFICATION AND STATISTICAL ANALYSIS

The differences in performance from different methods are tested by the Wilcoxon signed-rank test using a built-in function "signrank" in Matlab. Other comparisons are performed by Whitney-Mann's U test with "ranksum" in Matlab. Both the Wilcoxon signed-rank test and Whitney-Mann's U test are non-parametric. Under multiple comparisons, the raw p-values are corrected by the false discovery rate (FDR) correction.

To assess the significance of predictability during the biclassification, we conduct permutation to randomize the ground-truth labels and re-calculate the performance metrics to estimate the corresponding distribution under chance level (null model). As we use the ten-fold cross-validation scheme, 100 times permutations are separately conducted on the prediction results from each round of cross-validation. The results from these 1000 permutations are then pooled to generate the estimation of the chance-level distribution. The significance (*p*-value) is then obtained by statistically comparing the empirical distribution from trained models and the chance-level distribution, using a one-sided Whitney-Mann's U test.

ADDITIONAL RESOURCES

The data collection of the HUASHAN dataset has been registered as a clinical trial "ChiCTR2000036842" (URL of registry "http://www.chictr.org.cn/showproj.aspx?proj=59802").