



Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations

Chong-Yaw Wee^a, Chaoqiang Liu^a, Annie Lee^a, Joann S. Poh^a, Hui Ji^b, Anqi Qiu^{a,*}, the Alzheimers Disease Neuroimage Initiative¹

^a Department of Biomedical Engineering and Clinical Research Center, National University of Singapore, Singapore

^b Department of Mathematics, National University of Singapore, Singapore

ARTICLE INFO

Keywords:

Dementia classification
Cortical thickness
Graph
Convolutional neural networks
Transfer learning

ABSTRACT

Combining machine learning with neuroimaging data has a great potential for early diagnosis of mild cognitive impairment (MCI) and Alzheimer's disease (AD). However, it remains unclear how well the classifiers built on one population can predict MCI/AD diagnosis of other populations. This study aimed to employ a spectral graph convolutional neural network (graph-CNN), that incorporated cortical thickness and geometry, to identify MCI and AD based on 3089 T₁-weighted MRI data of the ADNI-2 cohort, and to evaluate its feasibility to predict AD in the ADNI-1 cohort (n = 3602) and an Asian cohort (n = 347). For the ADNI-2 cohort, the graph-CNN showed classification accuracy of controls (CN) vs. AD at 85.8% and early MCI (EMCI) vs. AD at 79.2%, followed by CN vs. late MCI (LMCI) (69.3%), LMCI vs. AD (65.2%), EMCI vs. LMCI (60.9%), and CN vs. EMCI (51.8%). We demonstrated the robustness of the graph-CNN among the existing deep learning approaches, such as Euclidean-domain-based multilayer network and 1D CNN on cortical thickness, and 2D and 3D CNNs on T₁-weighted MR images of the ADNI-2 cohort. The graph-CNN also achieved the prediction on the conversion of EMCI to AD at 75% and that of LMCI to AD at 92%. The fine-tuned graph-CNN further provided a promising CN vs. AD classification accuracy of 89.4% on the ADNI-1 cohort and > 90% on the Asian cohort. Our study demonstrated the feasibility to transfer AD/MCI classifiers learned from one population to the other. Notably, incorporating cortical geometry in CNN has the potential to improve classification performance.

1. Introduction

Alzheimers disease (AD) is clinically characterized by the appearance of a progressive decline in memory and cognition (Alzheimer's Association, 2015). It is the most common form of neurodegenerative dementia and has an astounding impact at individual and societal levels (Prince et al., 2015; Rizzi et al., 2014; Wimo et al., 2017). Despite a concerted effort to establish treatments for moderate and severe AD, clinical trial for alleviating such a degenerative process has yielded meager success (Lawlor et al., 2018; Yiannopoulou and Papageorgiou, 2013). Symptomatic treatments for AD could be efficacious only if the diagnostic and treatment envelop fall back to the early prodromal stage, such as mild cognitive impairment (MCI) (Karakaya et al., 2013; Robinson et al., 2015). Such conceptual appeal of early diagnosis and

prognosis of MCI and AD in mitigating disease impact has been tested recently, providing promising results and evidences that prodromal stages of AD are windows of opportunity in reducing the incidence and symptoms of AD (Ewers et al., 2011; Pellegrini et al., 2018b; Rathore et al., 2017).

Neuroimaging has provided relevant information on the diagnostic status and disease progression of AD and MCI. In quantifying patterns of structural change during early stages of AD, several neuroimaging initiatives² have discovered biological markers associated with AD and MCI based on multi-modal brain images and machine learning. In general, multi-modal neuroimaging data performs better than unimodal data in terms of the diagnostic accuracy of AD/MCI (e.g., (Dyrba et al., 2015; Dyrba et al., 2013; Liu et al., 2015; Liu et al., 2013; Möller et al., 2016; Nir et al., 2015; Pellegrini et al., 2018a; Yu et al., 2016; Zhu et al.,

* Corresponding author at: Department of Biomedical Engineering, National University of Singapore, 4 Engineering Drive 3, Block E4 #04-08, 117583, Singapore.
E-mail address: bieqa@nus.edu.sg (A. Qiu).

¹ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

² https://www.alz.org/research/for_researchers/partnerships/wwadni/about_wwadni

<https://doi.org/10.1016/j.nicl.2019.101929>

Received 15 April 2019; Received in revised form 2 June 2019; Accepted 2 July 2019

Available online 04 July 2019

2213-1582/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2014)). To our knowledge, the AD classification accuracies from existing literature ranged between 86 and 93% using the volumetric morphology of a subset of the T₁-weighted images from the Alzheimer's Disease Neuroimaging Initiative³ (ADNI) (n = 100–420) (Cuingnet et al., 2011; Liu et al., 2015; Liu et al., 2014; Liu et al., 2013; Zhang et al., 2011). In contrast, AD classification using regional cortical thickness features performed slightly worse ($\approx 85\%$) as compared to the regional volumetric features (n = 220–400) (Liu et al., 2013; Pellegrini et al., 2018a; Wee et al., 2012). Multi-modal features that include regional gray matter (GM) volumes, regional mean intensity of Positron Emission Tomography (PET) and amyloid decomposition, and cerebrospinal fluid (CSF) showed further improvement in diagnostic accuracy for AD to 93–96% and for MCI 80–82% (Yu et al., 2016; Zhu et al., 2014). However, despite the high classification accuracy, multi-modal imaging studies were often constrained by a small sample size as compared to unimodal imaging studies due to difficulties in acquiring multi-modal neuroimaging data from the same subject. This thus limited their generalizability to large datasets and other populations.

The accuracy of MCI and AD prediction has been further improved with the recent advent of deep learning. Most of the existing deep learning based implementation is performed on a coarse patch or brain region of T₁-weighted images (Gupta et al., 2013; Payan and Montana, 2015; Suk et al., 2017). The classification accuracies reached to 94% and 83% for AD and MCI classifications respectively when a sparse auto-encoder based on patches of T₁-weighted images and a convolutional neural network (CNN) based on 2D T₁-weighted images were employed (Gupta et al., 2013). Moreover, when a CNN was implemented on 3D T₁-weighted images, the analysis further improved classification accuracies to 95% and 92% for AD and MCI respectively (Payan and Montana, 2015). Despite high classification accuracies, there are several caveats of these neural network based methods that are worth further investigation. For example, the model architecture was limited to CNN with only 1 convolutional layer (Gupta et al., 2013). The sample in a majority of these existing studies was limited to the subsample of ADNI-1 (e.g., 755 MRI scans for each clinical class) (Gupta et al., 2013; Payan and Montana, 2015). Hence, a couple of questions remain unclear: 1) whether the accuracy of these classifiers on a large sample can achieve the same level as that using a small sample; 2) whether the classifiers trained on a specific dataset or population are generalizable or transferable to other datasets/populations. In this realm, whether an AD/MCI classifier built on a Caucasian population could predict well the diagnosis of an Asian population has yet been explored. The generalization of these classifiers to other datasets or populations, particularly those with a small sample size, is also crucial to mitigate the burden of building a reliable population-specific classifier from scratch.

In this study, we aimed to answer these questions using the cortical thickness data extracted from T₁-weighted MR scans, deep neural network classifier, and transfer learning. Cortical thinning has been reported in AD/MCI patients (Du et al., 2007; Lerch et al., 2008; Lerch et al., 2005; Singh et al., 2006), and has been identified as imaging biomarkers for the identification of AD/MCI, as well as the progression from MCI to AD (Querbes et al., 2009; Racine et al., 2018; Schaerer et al., 2016). However, a substantial body of existing studies (Eskildsen et al., 2013; Wee et al., 2012) employed regional features, e.g., the mean cortical thickness within a region-of-interest (ROI), for classification, but did not incorporate the cortical geometry. Two brain regions are close in terms of Euclidean distance but are far along the cortical surface, which influences the convolution operation in CNN. In addition, cortical sulci become wider as thickness decreases in MCI/AD patients. Hence, it is crucial to take into account of cortical geometry in CNN. In this study, we employed a spectral graph-CNN (Defferrard et al., 2016) to incorporate the cortical geometry that can be

represented as a graph. Unlike the traditional CNN in which the convolution and pooling operate on a regular Euclidean grid, the convolution and pooling operations in a graph-CNN were designed on an irregular grid. In our case, the convolution filters and pooling were applied along the cortical ribbon. We trained a spectral graph-CNN on cortical thickness of the ADNI-2 dataset, which consisted mainly of a Caucasian population, and then transferred this model to predict AD and MCI diagnosis of the ADNI-1 cohort and an Asian cohort. Unlike existing studies that used a subset of the ADNI dataset (e.g., (Korolev et al., 2017)), we used all available MRI scans from the ADNI-2 cohort (n = 3089) to train a robust spectral graph-CNN for dementia classification. We then used transfer learning to evaluate the generalizability of the spectral graph-CNN based AD/MCI classifier trained on a sizable Caucasian dataset to both the full dataset of the ADNI-1 cohort and a small Asian dataset. In the context of neural networks, transfer learning employs features/knowledge learned by a base network on a base dataset/task to a target network via fine-tuning on a target dataset/task (Yosinski et al., 2014). The target network tends to perform better than the same network that trained from scratch if the learned features share specific structures of both the base and target datasets/tasks. We thus expected that the classifiers obtained from the ADNI-2 dataset could achieve similar classification accuracy for the ADNI-1 and Asian populations.

2. Materials and methods

2.1. Datasets

2.1.1. ADNI cohorts⁴

This study included 1083 and 1012 subjects from the ADNI-2 and ADNI-1 cohorts, respectively. Subjects were primarily Caucasian and aged between 55 and 90 years old. The ADNI-1 cohort had 242 cognitive normal (CN), 415 MCI, and 355 AD. The ADNI-2 cohort had 300 CN, 314 early MCI (EMCI), 208 late MCI (LMCI) and 261 AD at the first visit, while the number of visits of each subject varied from 1 to 7 (i.e., baseline, 3-, 6-, 12-, 24-, 36-, and 48-month). At each visit, the subjects in the ADNI-1 cohort were diagnosed with one of the three clinical statuses and those in the ADNI-2 cohort were diagnosed with one of the four clinical statuses. The general diagnostic criteria for early and late MCI were the same except LMCI subjects had a lower cut-off point for logical memory II subscale from Wechsler Memory Scale. Table 1 provides demographic and clinical information of subjects from the ADNI-1 and ADNI-2 cohorts.

As each subject may have multiple MRI scans due to multiple visits, we included all available T₁-weighted images with good quality after processing. We used the clinical status at the time of the MRI acquisition as the classification ground truth, i.e., a subject with multiple scans may have different clinical labels if he/she converted from one clinical status to another at the following visits. Out of 3703 downloaded scans in the ADNI-1 cohort, we discarded 101 scans with no clinical label, thus leaving 3602 scans for further analyses. Out of 3234 downloaded scans in the ADNI-2 cohort, we discarded 47 scans with no clinical label thus leaving 3187 scans for further analysis.

2.1.2. Asian cohort

This study included 347 subjects (176 CN, 128 Moderate MCI and 43 AD) from an ongoing Asian aging study. The diagnosis of MCI and AD followed the same criteria as given in the ADNI (see details in Thong et al., 2014; Thong et al., 2013). The moderate MCI category in the Asian cohort is equivalent to the LMCI in the ADNI-2 cohort due to its more severe cognitive impairment (i.e., impairment in at least three

⁴ The number of subjects for each group was based on the clinical status during the MRI acquisition visit. Some subjects fell into 2 or more groups due to conversion from one clinical status to another.

³ adni.loni.usc.edu

Table 1

Demographic and clinical information of the ADNI-1 and ADNI-2 cohorts at the time of MRI acquisition.

ADNI-1 cohort				
	CN	MCI	AD	
Number of subjects ^a	242	415	355	
Number of scans	1071	1515	1016	
Female/male	493/578	525/990	443/573	
Age (mean $\pm$ SD)	76.9 $\pm$ 5.3	75.9 $\pm$ 7.3	76.3 $\pm$ 7.2	
MMSE (mean $\pm$ SD)	29.1 $\pm$ 1.1	27.0 $\pm$ 2.4	22.0 $\pm$ 4.2	
CDR-SB (mean $\pm$ SD)	0.1 $\pm$ 0.4	1.8 $\pm$ 1.1	5.2 $\pm$ 2.5	
ADNI-2 cohort				
	CN	EMCI	LMCI	AD
Number of subjects ^a	300	314	208	261
Number of scans	960	899	638	592
Female/Male	512/448	408/491	283/355	256/336
Age (Mean $\pm$ SD)	75.6 $\pm$ 6.9	72.9 $\pm$ 7.6	73.7 $\pm$ 8.0	75.3 $\pm$ 7.7
MMSE (Mean $\pm$ SD)	29.0 $\pm$ 1.3	28.1 $\pm$ 1.8	27.3 $\pm$ 2.2	21.8 $\pm$ 4.4
CDR-SB (Mean $\pm$ SD)	0.1 $\pm$ 0.3	1.3 $\pm$ 0.9	1.8 $\pm$ 1.2	5.7 $\pm$ 2.8

Abbreviations. CN: Control normal; AD: Alzheimer's disease; MCI: Mild cognitive impairment; EMCI: Early MCI; LMCI: Late MCI; MMSE: Mini-Mental State Exam; CDR-SB: Clinical Dementia Rating Scale-Sum of Boxes.

^a The number of subjects for each group was based on the clinical status during the MRI acquisition visit. There are subjects who fall into 2 or more groups due to conversion from one clinical status to another.

Table 2

Demographic and clinical information of the Asian cohort.

	CN	Moderate MCI	AD
Number of subjects	176	128	43
Female/male	79/97	84/44	29/14
Age (mean $\pm$ SD)	67.4 $\pm$ 5.1	74.1 $\pm$ 6.4	76.5 $\pm$ 7.7
MMSE (mean $\pm$ SD)	26.5 $\pm$ 2.1	21.0 $\pm$ 3.9	15.3 $\pm$ 4.7
CDR-SB (mean $\pm$ SD)	0.1 $\pm$ 0.2	0.9 $\pm$ 1.4	6.7 $\pm$ 3.7

Abbreviations. CN: Control normal; MCI: Mild Cognitive Impairment; AD: Alzheimer's disease; MMSE: Mini-Mental State Exam; CDR-SB: Clinical Dementia Rating Scale-Sum of Boxes.

cognitive domains of a formal neuropsychological test battery). Note that the Mini-Mental State Exam (MMSE) cutoff used for this Asian cohort was relatively lower than that used in the ADNI-2 but was validated well in the Asian population (Chin, 2002). Table 2 lists demographic and clinical information of subjects from the Asian cohort.

2.2. MRI data acquisition and analysis

2.2.1. ADNI-1 and ADNI-2 cohorts

Structural T₁-weighted MRI scans were acquired using either 1.5 T or 3 T scanners. The typical 1.5 T acquisition parameters were repetition time (TR) = 2400 ms, minimum full echo time (TE), inversion time (TI) = 1000 ms, flip angle = 8°, field-of-view (FOV) = 240 $\times$ 240 mm², acquisition matrix = 256 $\times$ 256 $\times$ 170 in the x-, y-, and z-dimensions, yielding a voxel size of 1.25 $\times$ 1.25 $\times$ 1.2 mm³. For 3 T scans, the acquisition parameters were a TR = 2300 ms, minimum full TE, TI = 900 ms, flip angle = 8°, FOV = 260 $\times$ 260 mm², acquisition matrix = 256 $\times$ 256 $\times$ 170, yielding a voxel size of 1.0 $\times$ 1.0 $\times$ 1.2 mm³.

2.2.2. Asian cohort

All MRI scans were acquired on a 3 T Siemens Magnetom Trio Tim scanner using a 32-channel head coil at the Clinical Imaging Research Centre of the National University of Singapore. The T₁-weighted MR images were acquired using magnetization prepared rapid gradient

recalled echo with 192 slices, 1 mm thickness, in-plane resolution 1 mm, no inter-slice gap, sagittal acquisition, FOV = 256 $\times$ 256 mm², acquisition matrix = 256 $\times$ 256, TR = 2300 ms, TE = 1.9 ms, TI = 900 ms and flip angle = 9°.

2.2.3. MRI data analysis

All T₁-weighted images from the ADNI-1, ADNI-2, and Asian datasets were segmented using FreeSurfer (Fischl et al., 2002). The white and pial cortical surfaces were generated at the boundary between white and gray matter and the boundary of gray matter and CSF, respectively. Cortical thickness was computed as the distance between the white and pial cortical surfaces. We represented cortical thickness on the mean surface, the average between the white and pial cortical surfaces. We employed a large deformation diffeomorphic metric mapping (LDDMM) algorithm (Du et al., 2011; Tan and Qiu, 2016, 2018; Zhong et al., 2010) to align individual cortical surfaces to the atlas and transferred the thickness of each subject to the atlas. This study included 3602, 3089, and 347 scans in the ADNI-1, ADNI-2, and Asian cohorts. Any processed imaging data that included any of the errors listed on <https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData> was discarded.

2.3. Cortical graph neural network

Mathematically, cortical thickness is a function defined on the cortical surface, where the cortical surface is a triangular mesh, represented as a graph, $\mathbf{M}^h = \{\mathbf{V}^h, \mathbf{E}^h\}$. $\mathbf{V}^h = \{\mathbf{v}_i^h | i = 1, \dots, n_h\}$ and $\mathbf{E}_{ij}^h = \{e_{ij}^h = (v_i^h, v_j^h) | 1 \leq i \leq n_h, 1 \leq j \leq n_h, i \neq j\}$ are the vertex and edge sets, respectively, and n_h is the number of vertices on the graph (brain cortical surface) of hemisphere h . The cortical thickness of one hemisphere can be represented as $\mathbf{c}^h = \{c_i^h | i = 1, \dots, n_h\}$, with c_i^h is the cortical thickness value at vertex i of hemisphere h . Thus, the cortical surface graph $\mathbf{M}^h$ and the corresponding cortical thickness vector ($\mathbf{c}$) of an individual in the atlas space can be represented respectively as

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}^l & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^r \end{pmatrix} \quad (1)$$

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}^l & \mathbf{0} \\ \mathbf{0} & \mathbf{c}^r \end{pmatrix} = \begin{pmatrix} c_1^l & c_2^l & \dots & c_{n_l}^l & \mathbf{0} \\ \mathbf{0} & c_1^r & c_2^r & \dots & c_{n_r}^r \end{pmatrix} \quad (2)$$

where $\mathbf{M}^l$ and $\mathbf{M}^r$ denote the cortical graphs for the left and right hemispheres, respectively, and $\mathbf{c}^l$ and $\mathbf{c}^r$ denote the cortical thickness vectors of the left and right hemispheres, respectively. Note that $\mathbf{M}$ was the same for all subjects and was used with the individual cortical thickness vectors $\mathbf{c}$ as the input to train a graph-CNN model for disease classification as described below. In our case, the numbers of vertices on $\mathbf{M}^l$ and $\mathbf{M}^r$ were 152,461 and 152,671 respectively.

Fig. 1A shows the architecture of the spectral graph-CNN used in this study, which was based on the formulation in (Defferrard et al., 2016). This graph-CNN concatenates an input layer, followed by three graph convolutional layers, a fully connected layer, and an output layer. The input layer took individual cortical thickness ($\mathbf{c}$) and its underlying graph ($\mathbf{M}$) as the input and then fed them to the first graph convolutional layer. The convolution on the graph is defined as

$$\mathbf{y} = \sum_{k=0}^{K-1} \theta_k T_k(L) \mathbf{x} \quad (3)$$

where $\mathbf{y}$ is the output of the convolution, $\mathbf{x}$ is the input signal (e.g., cortical thickness or output of the previous convolutional layer). The vector $\theta = (\theta_1, \theta_2, \dots, \theta_{K-1})^T$ is a vector of Chebyshev coefficients, and $T_k(L)$ is the Chebyshev polynomials of order k evaluated at the Laplacian of $\mathbf{M}$, $L = D^{-1/2} \mathbf{W} D^{1/2}$, where $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$ such that its eigenvalues lie within $[-1, 1]$. With recurrence relation of Chebyshev polynomials and let $\bar{x}_k = T_k(L) \mathbf{x} = 2L\bar{x}_{k-1} - \bar{x}_{k-2}$ with $\bar{x}_0 = \mathbf{x}$ and $\bar{x}_1 = L\mathbf{x}$, the convolutional

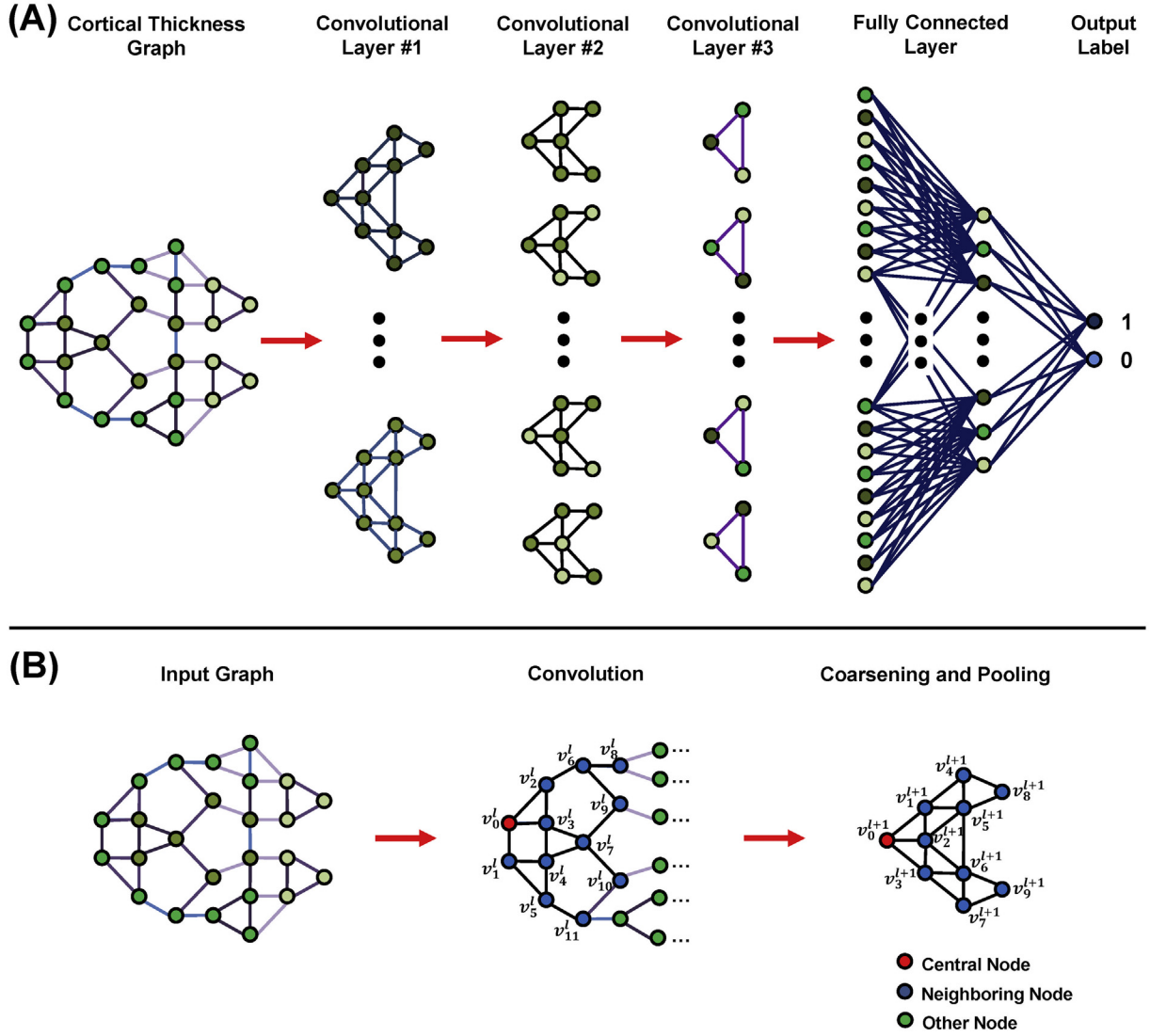


Fig. 1. Graph-CNN Architecture. (A) The graph-CNN model used in this study. (B) The coarsening and pooling operations for an input graph in a convolutional layer.

operation can be simplified as $y = g_\theta(L)x = [\tilde{x}_0, \dots, \tilde{x}_{K-1}]\theta$ with $g_\theta(L) = \sum_{k=0}^{K-1} \theta_k T_k(L)$ are the learnable filters (Defferrard et al., 2016). The polynomial order K ensures that the filters are strictly localized in a ball of radius K , i.e., K hops from the central node. In the graph convolutional layer, the input was convolved with learnable filters and went through a nonlinear activation function called Rectified Linear Unit (ReLU), defined as $f(x) = \max(0, x)$ (Nair and Hinton, 2010), to form the output of the current layer, i.e., feature map.

After the convolution, an average-pooling process with stride > 1 was interspersed to reduce dimensionality and to produce a more compact hierarchical representation of the input data. The pooling process was performed by applying a simple regular 1D pooling operator on the rearranged meaningful neighboring nodes as shown in Fig. 1B. The meaningful neighboring nodes were determined via a two-step approach: 1) coarsening the graph by a factor of two at each level, and 2) creating a balanced binary tree such that each node in the coarser graph has either one or two child nodes. In the coarsening step, two neighboring nodes that with maximum local normalized cut (Shi and Malik, 2000) were merged until the number of nodes at the coarser level was approximately half of the previous level. The coarsening process was repeated until the coarsest level was achieved. At the coarsest level, the merged nodes were *arbitrarily arranged* and this ordering was propagated stepwise to the finest level (i.e., the input

graph). Finally, a simple regular 1D average-pooling was performed on the rearranged nodes. The average-pooled feature maps were then fed as the input to the next graph convolutional layer. In the last graph convolutional layer, the filtered data were flattened and fed to the fully connected layer. The fully connected layer, which was the same as the conventional multilayer neural network, integrated all information from the last convolutional layer to make a clinical decision at the output (logits) layer via a softmax function.

The choice of the number of layers, the number of filters in each layer and the order of Chebyshev polynomials are highly application-specific. In this study, the graph-CNN was constructed with three graph convolutional layers, one fully connected layer and one output layer as shown in Fig. 1A. The number of filters in each graph convolutional layer was set as [8, 16, 32] respectively, the order of Chebyshev polynomial as 3 in every graph convolutional layer, and the number of the hidden nodes in the fully connected layer as 128 (Please see Figs. S1, S2, S3 and S4 in Supplementary Materials for the effects of network parameters). The network parameters were trained using a back propagation algorithm with a mini-batch size of 32, an initial learning rate of $1e^{-3}$, a learning rate decay of 0.05 for every 20 epochs, and a momentum of 0.9 (Cotter et al., 2011). The model was implemented using the TensorFlow⁵ library. During the training process, an l_2 -norm regularization function of $5e^{-4}$ was applied on all trainable filter weights

to prevent over-fitting to the training data.

In our experiments, we evaluated the performance of the spectral graph-CNN based on 10-fold cross-validation. We randomly selected a predefined percentage (10%) of subjects from each class as the testing subjects, and the remaining subjects were used as the training subjects for each cohort data. This ensures that the ratio of the number of subjects in each class in the testing dataset was similar to that in the training dataset. All MRI scans of the testing subjects were used to form the testing set, and the remaining MRI scans were used to form the training set. The model that performed the best on the validation subset according to adjusted geometric mean was used to predict the clinical status of the testing set. We opted to employ an adjusted geometric mean ($AGMean = \sqrt{SEN \times SPE} - \frac{1}{2}(SEN - SPE)^2$, see the definitions of sensitivity (SEN) and specificity (SPE) below) as the optimization criterion to identify the most effective and balanced graph-CNN model during training. It not only maximizes the accuracy on each of the two classes but also minimizes the difference between the sensitivity and specificity, i.e., balanced performance for both the positive and negative classes. The experiment was repeated for five times, and the average performance was reported in the result section. We performed the same training and evaluation procedures for six classifiers, including CN vs. AD, CN vs. LMCI, CN vs. EMCI, EMCI vs. LMCI, EMCI vs. AD, and LMCI vs. AD.

2.4. Generalization from ADNI-2 to ADNI-1 cohort

To evaluate the generalizability of the graph-CNN models to other datasets of similar Caucasian populations, we fine-tuned the models that were pre-trained based on the ADNI-2 cohort using the training set of the ADNI-1 cohort, and then evaluated their performance on the testing set of the ADNI-1 cohort. We hypothesized that a robust dementia classifier should be able to perform well on the MR scans that were acquired using different scanners with different field strengths and scanning protocols; i.e., *generalizability across datasets*. As the samples in the ADNI-1 cohort were diagnosed with 3 clinical statuses (i.e., CN, MCI and AD), we fine-tuned the CN vs. AD, CN vs. LMCI, and LMCI vs. AD models that performed the best on the testing set of the ADNI-2 cohort to the ADNI-1 cohort for CN vs. AD, CN vs. MCI, and MCI vs. AD classifications. As some of the samples in the ADNI-1 cohort were followed up in the following ADNI-2 study, we excluded those subjects (as well as their scans) that were used for training the spectral graph-CNN models based on the ADNI-2 cohort to avoid performance bias. Specifically, we utilized 654 CN scans and 965 AD scans of the ADNI-1 cohort for CN vs. AD classification, 661 CN scans and 1210 MCI scans of the ADNI-1 cohort for CN vs. MCI classification, and 1071 MCI scans and 944 AD scans of the ADNI-1 cohort for MCI vs. AD classification.⁶

For the models that were learned directly based on the ADNI-1 training set, all the parameters of the spectral graph-CNN models were initialized with a truncated normal distribution of zero mean and standard deviation of 0.1. For fine-tuning the trained models, we froze the convolutional layers, and updated only the parameters in the fully connected and the logits layers. We believed that the graph convolutional layers in the trained model are capable of capturing some basic yet essential dementia-associated patterns from the ADNI-2 cohort, and these patterns should also be essential to the ADNI-1 cohort, as both cohort were acquired from the similar populations. It is important to note that, for a fair comparison, the trained models were fine-tuned using the same experimental settings as the ‘learn-from-scratch models’, except with a smaller initial learning rate ($1e^{-4}$ vs. $1e^{-3}$).

2.5. Transfer learning from the ADNI-2 to the Asian cohort

We transferred the spectral graph-CNN models that were trained on the ADNI-2 dataset to the Asian dataset. Precisely, the models that performed the best on the testing set of the ADNI-2 cohort were fine-tuned on the training set of the Asian population. The performance of the fine-tuned model was then evaluated on the testing set of the Asian cohort. We hypothesized that the trained model should perform either *better than* or, at least, *comparable* to the learn-from-scratch model but with *less training effort*. We tested our hypothesis by comparing the performance of the trained models with the same models that learned from scratch for CN vs. Moderate MCI task. For our purpose, the model that performed the best for CN vs. LMCI task on the ADNI-2 cohort was fine-tuned for CN vs. Moderate MCI task on the Asian cohort.

To evaluate the transfer performance from the ADNI-2 to the Asian cohorts, we used the same experimental settings as for the transfer learning from the ADNI-2 to the ADNI-1 cohorts, as described in the previous section.

It is well-known that learning or fine-tuning a deep neural network requires datasets with a relatively large sample size. With a limited number of AD subjects ($n = 43$) in the Asian cohort, direct training a spectral graph-CNN or fine-tuning a trained model on the Asian cohort for AD prediction is unreliable. We thus directly used the previously fine-tuned CN vs. LMCI model to predict the labels of all 43 AD subjects in the Asian cohort. We anticipated that the trained model fine-tuned on the less severe condition (i.e., Moderate MCI) of the Asian cohort might be able to identify the more severe condition (i.e., AD) from the same cohort.

2.6. Evaluation measures

Classification accuracy (ACC), sensitivity (SEN), and specificity (SPE) are often used to quantify the classification performance. They are defined as follows,

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

$$SEN = \frac{TP}{TP + FN} \quad (5)$$

$$SPE = \frac{TN}{TN + FP} \quad (6)$$

where TP, TN, FN, and FP denote the true positive, true negative, false negative and false positive, respectively. The sensitivity and specificity provide the proportion of correctly identified samples for positive and negative classes, respectively. However, these measures, including the predictive accuracy, are sensitive to the ratio of the number of subjects in the positive and negative classes, and hence may provide inaccurate and misleading information on the performance of a classifier on an imbalanced dataset (López et al., 2013). To overcome this issue and to take into consideration the ratio of the number of subjects in the positive and negative classes, we used the geometric mean (GMean) and F1 score (F1), defined as

$$GMean = \sqrt{SEN \times SPE} = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \quad (7)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (8)$$

Both metrics attempt to maximize the accuracy of each of the two classes when the number of subjects in the positive and negative classes is imbalanced (Barandela et al., 2003).

2.7. Most discriminating brain regions based on cortical thickness

Identifying the most discriminating brain regions based on cortical

⁵ <https://www.tensorflow.org/>

⁶ Different numbers of CN, MCI and AD scans of the ADNI-1 cohort were used in the classification tasks since subjects were randomly selected from the ADNI-2 cohort for training the graph-CNN models from scratch.

thickness could potentially provide insight into anomalies of the brain morphological geometry in individuals with dementia. To accomplish this goal, we first grouped vertices on the cortical surface based on their shortest distance to one of the 76 cortical regions defined in the well-known Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). We then utilized a “leave-one-region-out” strategy to evaluate the contribution of cortical thickness within a brain region to dementia classification. In this strategy, cortical thickness of vertices within a brain region was first set to zero to remove its morphological information but preserving the topology of the cortical thickness graph. This “trimmed” cortical thickness graph was then fed to the spectral graph-CNN fine-tuned on the ADNI-2 cohort to predict its clinical label. The same process was repeated for all 76 brain regions to quantify the contribution of every brain region to dementia classification. The discriminative power or the contribution of a brain region is proportional to the drop of classification accuracy when cortical thickness of vertices within that region was excluded during classification. A brain region is considered the most discriminative if its removal causes the most significant drop in classification accuracy. We provided the top ten brain regions that with the most significant discriminative power (i.e., the most significant drop of classification accuracy) for the CN vs. AD and CN vs. LMCI classification tasks based on all available CN, AD, and LMCI scans from the ADNI-2 cohort.

3. Results

We validated the effectiveness of the proposed framework for the dementia classification based on the cortical thickness data. To prevent potential leak of information in 10-fold across-validation, we constructed non-overlapping training and testing sets. Specifically, we randomly selected 10% of subjects from each class as the testing subjects, and the remaining subjects were used as the training subjects. We then used all MRI scans of the testing subjects to form our testing set and all MRI scans of the training subjects to form our training set. This ensured no overlapping in subjects and scans in the training and testing sets and thus guaranteed unbiased evaluation performance.

In this section, we first provided the classification performance of the spectral graph-CNN on the ADNI-2 cohort and compared its performance with that of neural networks in Euclidian domain, including a conventional multilayer network (MLN) and a conventional 1D CNN on the cortical thickness data and a conventional 2D CNN on T₁-weighted MRI images. We then provided the prediction accuracy for the conversion of MCI to AD. Then, we provided the transfer classification performance of the spectral graph-CNN that trained on the ADNI-2 cohort, but fine-tuned and tested on the ADNI-1 and Asian cohorts.

3.1. Classification performance on the ADNI-2 cohort

Table 3 lists the mean classification performance of the spectral graph-CNN on the ADNI-2 cohort over five repetitions. The classification accuracies for CN vs. AD (85.8%) and EMCI vs. AD (79.2%) were high, given the large sample size, followed with CN vs. LMCI (69.3%), LMCI vs. AD (65.2%), and EMCI vs. LMCI (60.9%). The classification accuracy for CN vs. EMCI (51.8%) was the lowest among the six classifiers and had the largest variability of all measures, suggesting similar brain morphology between the normal controls and early MCI subjects. The same pattern was observed when F1 and GMean were employed.

3.1.1. Comparison with multilayer network (MLN)

To assess the potential benefits of incorporating cortical geometry into the spectral graph-CNN, we compared the spectral graph-CNN with a conventional multilayer network (MLN) defined in Euclidean domain that took the cortical thickness vector as input but discarded the underlying geometric information of the cortical graph. We designed the MLN with four hidden fully connected layers that respectively had [1024, 512, 256, 128] hidden nodes and a bias node for each layer. The

network parameters were trained using a stochastic gradient descent approach with a mini-batch size of 32, an initial learning rate of $1e^{-3}$, a learning rate decay of 0.05 after every 20 epochs, and a momentum of 0.9, the same as those used in training the graph-CNN. However, a larger maximum training epoch of 128 was used to obtain a good validation performance.

The pair-wise *t*-test revealed that the spectral graph-CNN outperformed the MLN in three classifiers, including CN vs. AD, CN vs. LMCI, and EMCI vs. LMCI, in terms of classification accuracy and GMean (Table 3; all *p*-values < 0.05). The spectral graph-CNN significantly improved the classification accuracies by 4.0% (CN vs. AD), 4.7% (CN vs. LMCI), and 6.1% (EMCI vs. LMCI). The spectral graph-CNN also performed marginally better than the MLN for EMCI vs. AD (*p*-value < 0.06 for classification accuracy and *p*-value $\approx$ 0.08 for F1 and GMean). Comparable results were shown for CN vs. EMCI and LMCI vs. AD. Due to bias to the ratio of the number of subjects in each class, the sensitivity and specificity showed complementary findings to each other.

3.1.2. Comparison with 1D CNN

To illustrate the importance of pooling meaningful neighboring nodes on the cortical graph in pooling operation, we compared the spectral graph-CNN with a conventional 1D CNN that took the cortical thickness vector as input and discarded the spatial correlation of nodes in the cortical graph. The architecture of the 1D CNN was the same as the spectral graph-CNN and with the same number of network parameters except replacing the graph-based pooling process that considers the node-node spatial correlation with a simple grid-based pooling process. The 1D CNN was trained using the same setting as that used for training the spectral graph-CNN but with a larger maximum training epoch of 128.

Similar to the MLN, the spectral graph-CNN outperformed the 1D CNN in three classifiers, including CN vs. AD, CN vs. LMCI, and EMCI vs. LMCI, in terms of classification accuracy and GMean (Table 3; all *p*-values < 0.05). The spectral graph-CNN significantly improved the classification accuracies by 4.1% (CN vs. AD), 6.2% (CN vs. LMCI), and 1.8% (EMCI vs. LMCI). The spectral graph-CNN also performed marginally better than the 1D CNN for EMCI vs. AD (*p*-values $\approx$ 0.08 for classification accuracy, F1 and GMean). The 1D CNN achieved comparable results compared to the spectral graph-CNN for CN vs. EMCI and LMCI vs. AD.

3.1.3. Comparison with 2D CNN

To compare the performance of the spectral graph-CNN with the popular 2D CNN defined in Euclidean domain, we implemented an architecture widely used in computer vision, i.e., the ResNet (He et al., 2016) of 50 layers pre-trained⁷ using ImageNet dataset.⁸ It took the average intensity values across all 2D slides in the axial, coronal, and sagittal views and fed them to the three channels of the ResNet-50 model. We replaced the output layer of the ResNet-50 model, which initially contained 1000 nodes, with a layer of 2 nodes for binary classification (e.g., CN vs. AD). We fine-tuned all layers with a relatively small initial learning rate of $1e^{-4}$ and a mini-batch size of 64 for 100 epochs because the representation of the ResNet-50 model learned using natural image dataset might not characterize well the morphological patterns of the brain.

The spectral graph-CNN outperformed the 2D CNN in four classifiers, including CN vs. AD, CN vs. LMCI, EMCI vs. AD and LMCI vs. AD, in terms of classification accuracy and GMean (Table 3; all *p*-values < .05). The spectral graph-CNN significantly improved the classification accuracies by 7.4% (CN vs. AD), 9.9% (CN vs. LMCI), 12.8% (EMCI vs. AD) and 2.7% (LMCI vs. AD). The classification accuracy of

⁷ <https://github.com/tensorflow/models>

⁸ <http://www.image-net.org/>

Table 3

Classification performance of the graph CNN, multilayer network (MLN), 1D and 2D convolutional neural networks (CNN) on the ADNI-2 dataset.

Model	Task	Sample	ACC (%)	SEN (%)	SPE (%)	F1 (%)	GMean (%)
Graph CNN	CN vs. AD	960/592	85.8 ± 0.8	83.5 ± 3.2	87.5 ± 2.8	82.9 ± 0.9	85.4 ± 0.8
	CN vs. LMCI	960/638	69.3 ± 2.2	65.6 ± 7.6	72.0 ± 5.4	64.2 ± 4.3	68.5 ± 3.0
	CN vs. EMCI	960/899	51.8 ± 1.2	55.3 ± 5.1	48.6 ± 6.4	52.6 ± 2.1	53.5 ± 4.2
	EMCI vs. LMCI	899/638	60.9 ± 2.2	52.5 ± 8.8	67.8 ± 9.8	53.5 ± 3.5	59.1 ± 1.4
	EMCI vs. AD	899/592	79.2 ± 2.6	70.4 ± 4.7	85.8 ± 4.7	74.4 ± 3.3	77.6 ± 2.7
MLN	LMCI vs. AD	638/592	65.2 ± 1.6	62.6 ± 5.2	68.0 ± 6.6	64.1 ± 2.2	65.3 ± 1.4
	CN vs. AD	960/592	81.8 ± 1.0 ^a	78.7 ± 4.0 ^a	84.0 ± 3.8	78.3 ± 1.1 ^a	81.6 ± 1.6 ^a
	CN vs. LMCI	960/638	64.6 ± 3.2 ^a	55.7 ± 3.2 ^a	71.0 ± 5.6	56.9 ± 3.0 ^a	62.7 ± 2.7 ^a
	CN vs. EMCI	960/899	55.3 ± 3.1	58.7 ± 5.9	52.2 ± 7.0	56.2 ± 3.3	55.0 ± 3.2
	EMCI vs. LMCI	899/638	54.8 ± 6.9 ^a	54.5 ± 10.9	55.0 ± 5.8 ^a	50.7 ± 8.5	54.6 ± 7.3 ^a
1D CNN	EMCI vs. AD	899/592	76.4 ± 1.2	68.9 ± 3.5	82.2 ± 0.9	71.3 ± 2.0	75.1 ± 1.6
	LMCI vs. AD	638/592	61.4 ± 5.6	63.1 ± 4.7	59.8 ± 8.2	61.9 ± 4.7	61.3 ± 5.7
	CN vs. AD	960/592	81.7 ± 1.6 ^a	80.0 ± 5.3	83.0 ± 3.5	78.3 ± 2.3 ^a	81.4 ± 1.9 ^a
	CN vs. LMCI	960/638	63.1 ± 4.3 ^a	53.6 ± 7.8 ^a	69.9 ± 4.8	54.6 ± 6.4 ^a	61.0 ± 5.1 ^a
	CN vs. EMCI	960/899	51.4 ± 2.5	52.6 ± 6.3	50.3 ± 5.7	51.3 ± 3.6	51.3 ± 2.2
2D CNN	EMCI vs. LMCI	899/638	59.1 ± 4.0 ^a	55.0 ± 6.4	62.2 ± 8.9	53.8 ± 3.3	58.2 ± 3.2 ^a
	EMCI vs. AD	899/592	74.1 ± 3.5	64.9 ± 7.3	81.3 ± 6.1	68.0 ± 4.6	72.1 ± 3.6
	LMCI vs. AD	638/592	63.9 ± 3.6	56.7 ± 9.9 ^a	72.5 ± 7.7	59.5 ± 4.5	63.7 ± 4.2
	CN vs. AD	960/592	78.4 ± 2.8 ^a	57.4 ± 7.3 ^a	86.9 ± 3.0	60.4 ± 5.5 ^a	70.5 ± 4.5 ^a
	CN vs. LMCI	960/638	59.4 ± 2.5 ^a	51.8 ± 4.4 ^a	64.9 ± 2.5 ^a	51.6 ± 3.6 ^a	57.9 ± 2.9 ^a
2D CNN	CN vs. EMCI	960/899	52.3 ± 3.1	49.4 ± 5.4	54.6 ± 7.9	50.2 ± 3.2	51.9 ± 3.1
	EMCI vs. LMCI	899/638	60.5 ± 2.8	45.9 ± 6.2 ^a	71.7 ± 3.6	50.0 ± 4.8 ^a	57.2 ± 3.7
	EMCI vs. AD	899/592	66.4 ± 4.2 ^a	59.4 ± 4.7 ^a	71.6 ± 7.5 ^a	60.3 ± 3.9 ^a	65.1 ± 3.8 ^a
	LMCI vs. AD	638/592	62.5 ± 3.3 ^a	69.3 ± 7.5	55.9 ± 4.6 ^a	64.6 ± 4.2	62.1 ± 3.0 ^a

Abbreviations. CN: Control normal; AD: Alzheimer's disease; MCI: Mild cognitive impairment; EMCI: Early MCI; LMCI: Late MCI; ACC: Accuracy; SEN: Sensitivity; SPE: Specificity; F1: F1 score; GMean: Geometric mean. Bold indicates a significant improvement of graph-CNN in prediction accuracy.

^a indicates the graph CNN model statistically outperformed MLN, 1D or 2D CNN at $p < 0.05$.

the 2D CNN for CN vs. EMCI was slightly greater than the spectral graph-CNN but not statistically significant. Due to bias to the ratio of the number of subjects in each class, the sensitivity and specificity showed complementary findings to each other. Moreover, the spectral graph-CNN showed a balanced performance for the classifiers with imbalanced datasets, i.e., CN vs. AD and CN vs. LMCI, according to sensitivity and specificity.

3.1.4. Prediction of MCI to AD Conversion

To assess its potential application for predicting the conversion of MCI to AD, we employed the spectral graph-CNN to predict the conversion of MCI subjects to AD based on their cortical thickness graphs at least several months prior to the conversion. Specifically, prediction of MCI to AD conversion was performed only on the cortical thickness graphs of the converted MCI subjects when they were still at EMCI or LMCI conditions. In the ADNI-2 cohort, 24 EMCI and 50 LMCI subjects converted to AD at the followed-up visits. We used their baseline scans and predicted the MCI conversion to AD. The spectral graph-CNN, trained for CN vs. AD task, corrected predicted the conversion of 18/24 EMCI (75%) and 46/50 (92%) LMCI to AD.

3.2. Transfer learning of the graph-CNN to the ADNI-1 cohort

Figs. 2, 3 and 4 provided the transfer classification performance of the spectral graph-CNN models for CN vs. AD, CN vs. MCI and MCI vs. AD tasks, respectively. These models were trained on the ADNI-2 cohort but tested on the ADNI-1 cohort. The fine-tuning trained models were robustly generalized to the ADNI-1 cohort by achieving good transfer performance for CN vs. AD and MCI vs. AD tasks. For these two tasks, the fine-tuning trained models, in general, achieved higher and more consistent performance than the 'learn-from-scratch' model in terms of sensitivity, F1 score and geometric mean with respect to the number of fine-tuning epoch. Specifically, the CN vs. AD accuracy by the fine-tuning trained graph-CNN model on the ADNI-1 cohort was consistently above 88.0%, since 40 epochs with the best performance (Fig. 2, ACC = 89.4%, SEN = 91.4%, SPE = 86.5%, F1 = 91.1%, GMean = 88.9%) was achieved at 100 epochs. The fine-tuning trained

models for MCI vs. AD performed better than the 'learn-from-scratch' models in all four evaluation measures for fine-tuning epoch smaller than 80. The best MCI vs. AD performance (Fig. 4, ACC = 65.2%, SEN = 70.6%, SPE = 60.8%, F1 = 64.8%, GMean = 65.5%) by the fine-tuning trained graph-CNN model on the ADNI-1 cohort was achieved at 20 epochs and was compatible to the LMCI vs. AD performance on the ADNI-2 cohort (Table 4, ACC = 65.2%, SEN = 62.6%, SPE = 68.0%, F1 = 64.1%, GMean = 65.3%). On the other hand, the CN vs. MCI performance by the fine-tuning trained models were better than the learn-from-scratch models in specificity, F1 score and geometric mean only for fine-tuning epoch smaller than 60. The best CN vs. MCI performance (Fig. 3, ACC = 65.0%, SEN = 61.9%, SPE = 70.1%, F1 = 68.8%, GMean = 65.9%) by the fine-tuning trained spectral graph-CNN model was achieved at 40 epochs, but the performance gradually deteriorated when the fine-tuning epoch was increased.

3.3. Transfer learning of the graph-CNN to the Asian cohort

Fig. 5 shows the classification performance of the graph-CNN model that was directly trained based on the Asian cohort (learn-from-scratch model) and was trained based on the ADNI-2 cohort (trained CN vs. LMCI model) but tested on the Asian cohort for CN vs. Moderate MCI task. The trained model consistently achieved higher and more consistent performance than the learn-from-scratch model in terms of sensitivity, F1 score and geometric mean with respect to the number of fine-tuning epoch. Moreover, the best CN vs. Moderate MCI performance (Fig. 5, ACC = 71.1%, SEN = 73.7%, SPE = 69.2%, F1 = 68.3%, GMean = 71.4%) by this trained graph-CNN model on the Asian cohort was achieved at 60 epochs, and such performance was consistently preserved with more epochs. This transfer classification performance is comparable to the performance of the graph-CNN on the ADNI-2 cohort (Table 3; CN vs. LMCI: 69.3%). We then directly applied this trained model (fine-tuned with 60 epochs) to predict the labels of 43 AD subjects of the Asian cohort and achieved a very promising prediction accuracy of 88.4% (38/43 AD subjects). The AD prediction accuracy was consistently higher than 90% when the trained model was fine-tuned with more epochs.

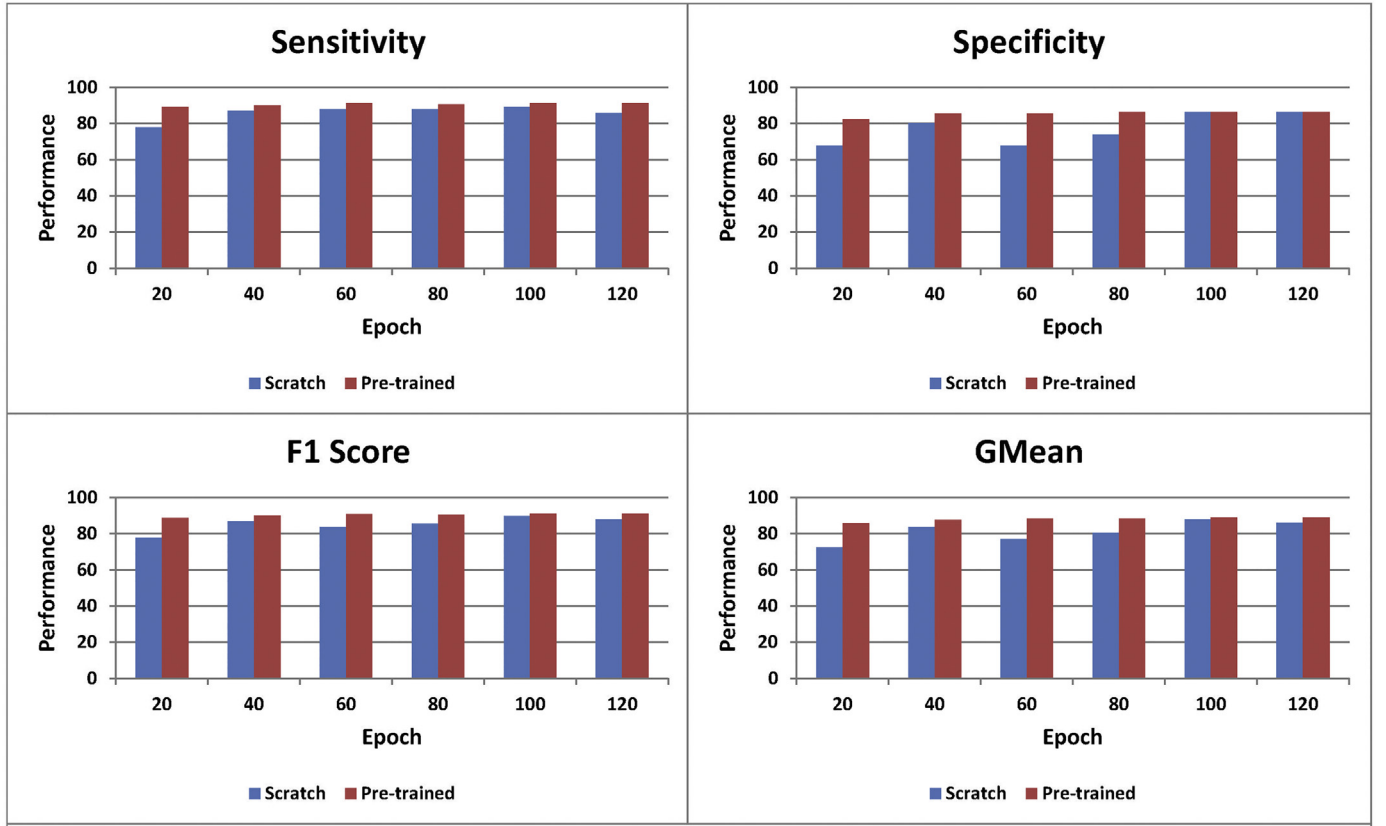


Fig. 2. Classification performance of the graph-CNN model directly trained based on the ADNI-1 cohort (Scratch) and trained based on the ADNI-2 cohort (Pre-trained) with respect to the number of training epoch for the CN vs. AD classification.

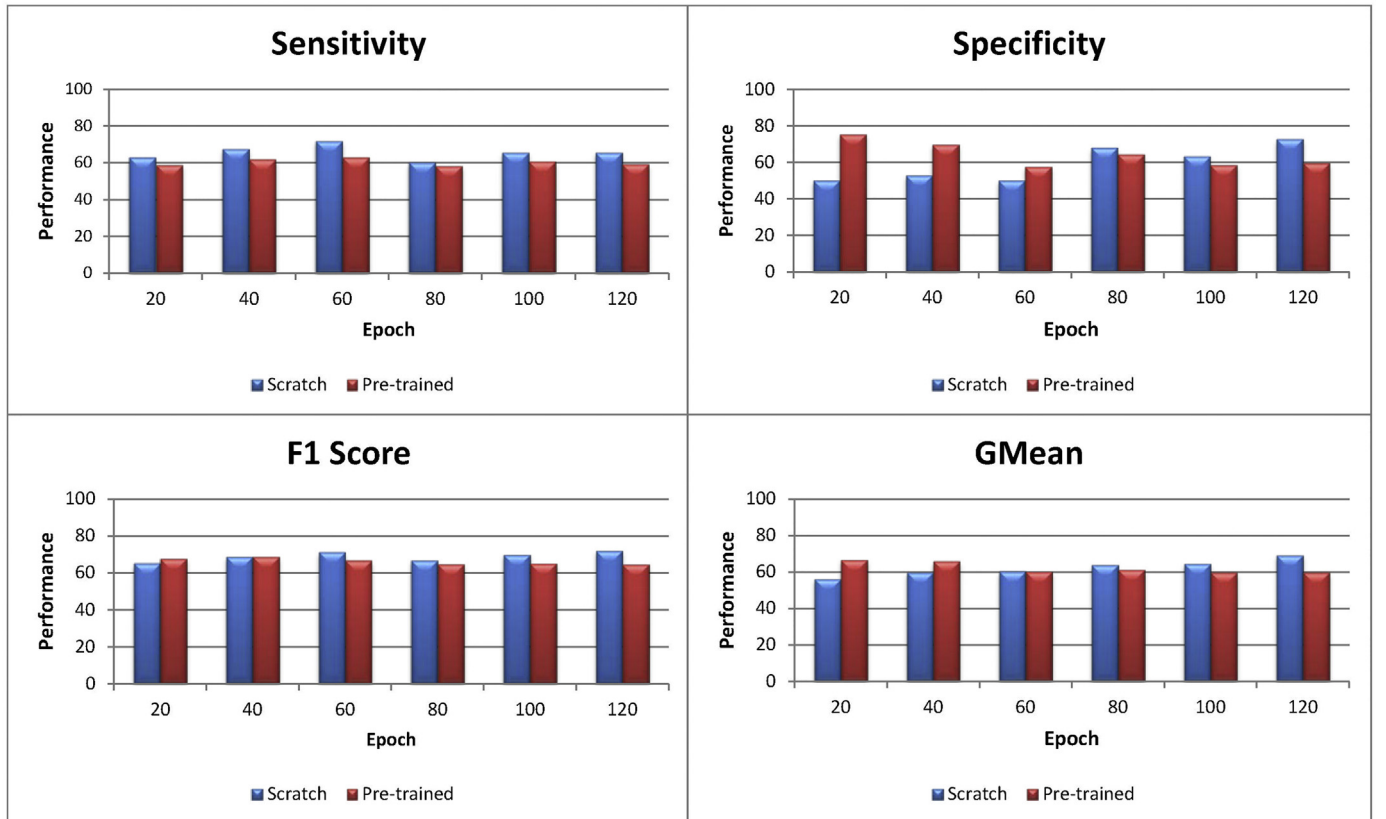


Fig. 3. Classification performance of the graph-CNN model directly trained based on the ADNI-1 cohort (Scratch) and trained based on the ADNI-2 cohort (Pre-trained) with respect to the number of training epoch for the CN vs. MCI classification.

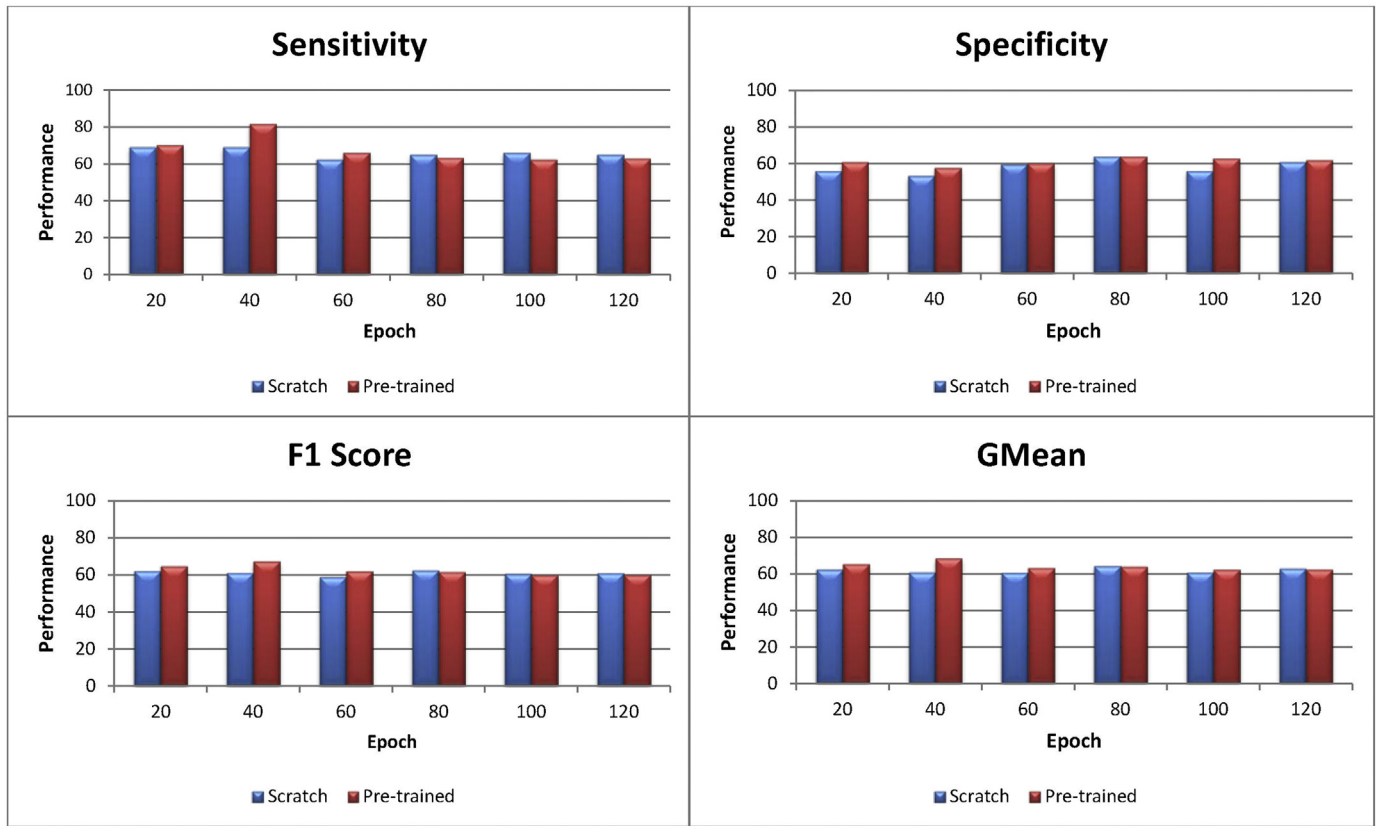


Fig. 4. Classification performance of the graph-CNN model directly trained based on the ADNI-1 cohort (Scratch) and trained based on the ADNI-2 cohort (Pre-trained) with respect to the number of training epoch for the MCI vs. AD classification.

Table 4

Classification performance of the spectral graph-CNN models for the ADNI-1 cohort.

Task	Samples	ACC (%)	SEN (%)	SPE (%)	F1 (%)	GMean (%)
CN vs. AD	654/965	81.0	85.5	74.5	84.3	79.8
CN vs. MCI	661/1210	67.6	71.3	60.7	74.0	65.8
MCI vs. AD	1071/944	65.4	77.5	54.6	67.7	65.1

Abbreviations. CN: Control normal; AD: Alzheimer's disease; MCI: Mild cognitive impairment; ACC: Accuracy; SEN: Sensitivity; SPE: Specificity; F1: F1 score; GMean: Geometric mean.

3.4. The most discriminating brain regions

Fig. 6 graphically illustrates the top ten most discriminating brain regions for the AD and LMCI classifications. The most discriminating regions for AD included the left parahippocampus, the left anterior and middle cingulate gyri, the right superior temporal gyrus, the right Heschl's gyrus, left precuneus, the right postcentral gyrus, the left Rolandic operculum, the right paracentral lobule, and the left lingual gyrus (Fig. 6A). The AD classification accuracies of these regions were between 60–66% (Table S2 in the Supplementary), implying significant drops of performance from that used all brain regions (86%). On the other hand, the most discriminating regions for LMCI were similar to those for AD, mainly including the bilateral parahippocampus, the middle cingulate gyri, the left middle temporal gyrus, the bilateral superior temporal gyri, the bilateral Heschl's gyri, and the left olfactory (Fig. 6B). The LMCI classification accuracies of these regions were between 50–52% (Table S2 in the Supplementary), implying significant drops of performance from that used all brain regions (69%). Consistent with existing literature (Chan et al., 2001; Dickerson et al., 2009; Eskildsen et al., 2015; Fjell et al., 2014; Jack et al., 1998; Lerch et al.,

2008; Singh et al., 2006; Visser et al., 2002), the spectral graph-CNN identified the parahippocampus as one of the major cortical regions distinguishing AD and LMCI from CN.

4. Discussion

In this present study, we endeavored to examine whether classifiers trained on a specific population are transferable to other populations for the AD/MCI diagnosis. Our experimental results demonstrated that the proposed spectral graph-CNN based framework in general performed significantly better than the Euclidean-domain based MLN and 1D CNN for the MCI and AD classification. Moreover, our framework also achieved a promising MCI to AD conversion prediction performance, which is vital for the early diagnosis of AD at its prodromal stage. More importantly, our results demonstrated the feasibility of applying the classifiers built on the ADNI-2 dataset to the Asian dataset. Such finding is crucial to mitigate the burden of building a reliable population-specific classifier from scratch.

We demonstrated that the proposed framework generally outperformed those Euclidean-domain based neural network classifiers, including MLN and 1D CNN with cortical thickness as the input and the 2D CNN with T₁-weighted MR image as the input (Table 3). The improvement in the performance of the spectral graph-CNN over the MLN could be attributed to its preservation of the underlying cortical geometry (as illustrated in Fig. 1). Moreover, compared to the fully connected MLN, the spectral graph-CNN has fewer learnable parameters (i.e., filter weights). A small number of filters (i.e., 8, 16 and 32 filters in the first, second and third convolutional layers, respectively, and each filter has only 3 parameters) that shared over all locations on the cortical graph were used in the study, which makes it more suitable when the training sample size is small.

On the other hand, the classification improvement of the spectral

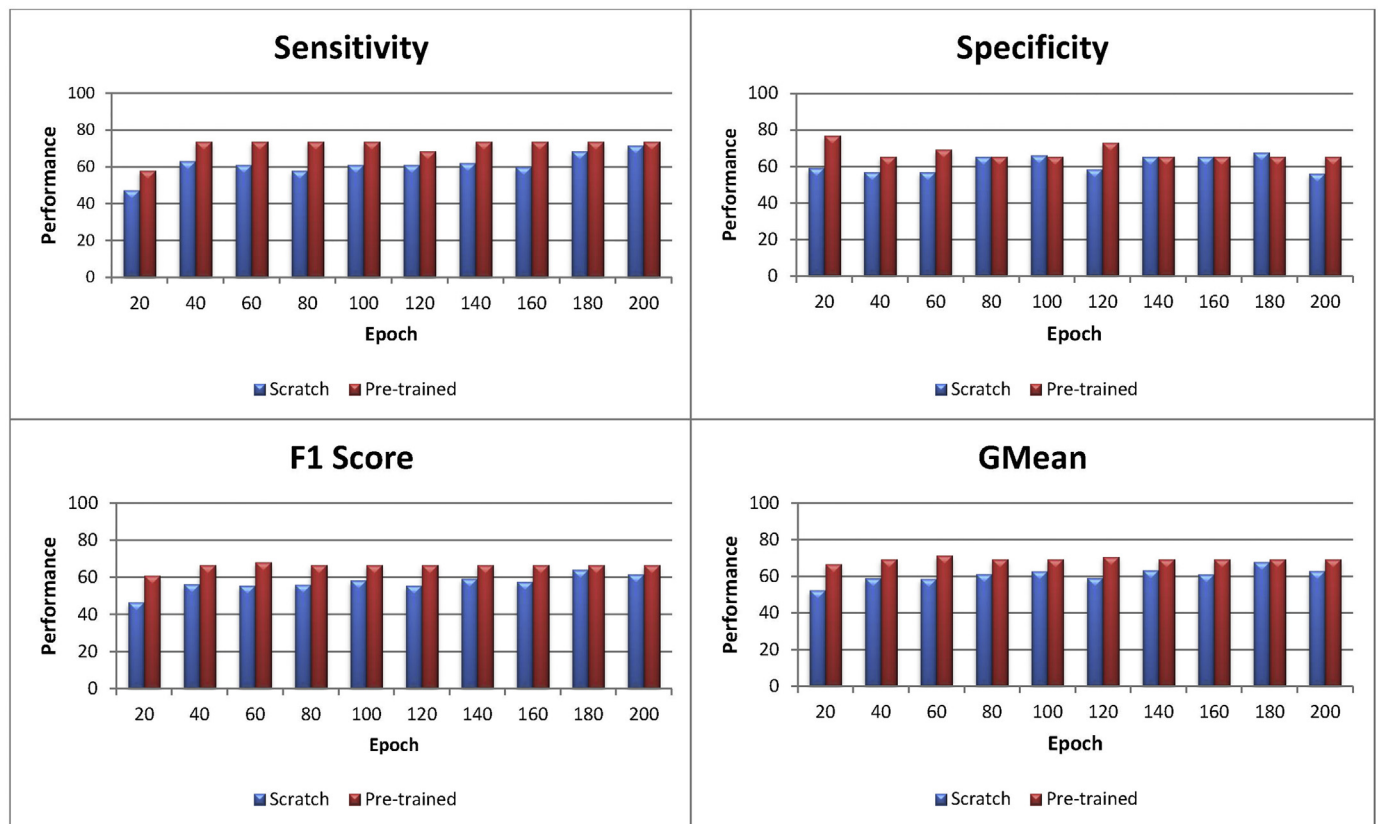


Fig. 5. Classification performance of the graph-CNN model directly trained based on the Asian cohort (Scratch) and trained based on the ADNI-2 cohort (Pre-trained) with respect to the number of training epoch for the CN vs. Moderate MCI classification.

graph-CNN over the conventional 1D CNN could be attributed to meaningful convolution and pooling operations among the “locally-connected” vertices on the cortical graph (see illustration in Fig. 1). This result demonstrated the importance of identifying locally-connected neighboring vertices on a graph for convolution and pooling operations that are analogous to those on a grid-based image in computer vision (Niepert et al., 2016). Also, our finding of better classification performance by the spectral graph-CNN over the 2D CNN might due to the finer morphological information provided by the cortical thickness on the cortical graph compared to the coarser pixel-based image intensity. Nevertheless, this experiment was limited to the CNN where the 2D images of the three principal axes were used in order to employ the trained ResNet-50 model that was specified with three channels of 2D images as inputs. The performance of the 2D CNN model may be improved when incorporating more 2D MRI slices may contain specific statistical property of the sample (Xu et al., 2013).

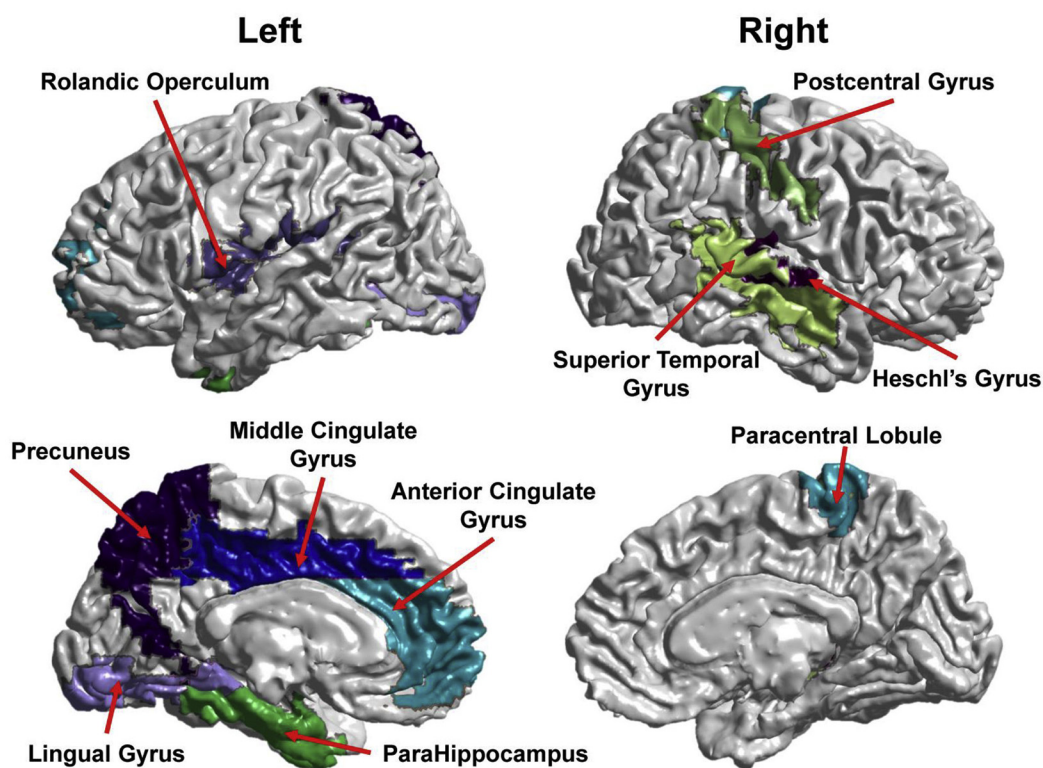
Our proposed framework further achieved a promising MCI to AD conversion prediction performance (75% and 92% for EMCI and LMCI, respectively), which is critical for early diagnosis of AD at its prodromal stage. These high prediction accuracies could be attributed to the spectral graph-CNN's ability to extract dementia-associated subtle changes of brain morphology from the cortical graph. The conversion accuracy for EMCI scans was much lower than that for LMCI scans possibly due to less obvious brain morphological changes at the beginning stage (i.e., EMCI) compared to at the later stage (i.e., LMCI) of the disease (Klöppel et al., 2012; Ledig et al., 2018). Identifying MCI subjects with a high risk of developing AD is crucial as intervention or treatment could be applied at the earliest stage to slow down the progression of disease (Eshkoor et al., 2015).

In the past few years, researchers employed deep learning approaches for predicting MCI and AD, which was mainly based on the ADNI-1 dataset (see review in (Basaia et al., 2019)). There was one

recent study that performed CNN on the ADNI-2 dataset (Korolev et al., 2017). This study incorporated two regular CNN models, denoted as VoxCNN and VoxRes, and evaluated their performance on a small subset of 61 CN, 50 AD, 43 LMCI and 77 EMCI subjects from the ADNI-2 dataset (Korolev et al., 2017). The VoxCNN was a 17-layers regular 3D CNN built upon the VGG architecture (Simonyan and Zisserman, 2014), while the VoxRes was a 21-layers regular 3D CNN built upon the ResNet architecture (He et al., 2016). Both VoxCNN and VoxRes were directly applied on the downsampled unprocessed MRI scans, i.e., no intensity normalization, no skull-stripping, no segmentation, etc. Our spectral graph-CNN on the cortical thickness outperformed these two voxel-based 3D CNN models in all classifiers in terms of classification accuracy, except for CN vs. EMCI where all compared classifiers achieved just above chance performance (see (Korolev et al., 2017)). Our spectral graph-CNN model showed at least 4.9% improvement on classification accuracy for classifiers of CN vs. AD, CN vs. LMCI, EMCI vs. LMCI, and EMCI vs. AD. The robustness of our spectral graph-CNN could be attributed to the utilization of: (1) *finer* cortical morphological features compared to *downsampled coarser* pixel-based image intensity, (2) pre-processed *good quality noise-reduced* MRI scans compared to unprocessed *noisy* original scans, and (3) significantly *larger sample size* (3089 vs. 231) to learn compact latent patterns from high dimensional neuroimaging data. Furthermore, our spectral graph-CNN achieved more consistent performances over multiple repetitions for all six classifiers as shown by much smaller standard deviations in classification accuracy. Our results shed new light on the importance of cortical geometry in deep neural network for improving classification accuracy.

The spectral graph-CNN models that were trained based on the ADNI-2 cohort demonstrated better classification performance on the ADNI-1 cohort after fine-tuning as compared to the models that were learned directly based on the ADNI-1 cohort (‘learn-from-scratch’ models) when a small number of training/fine-tuning epochs were

(A) CN vs. AD



(B) CN vs. LMCI

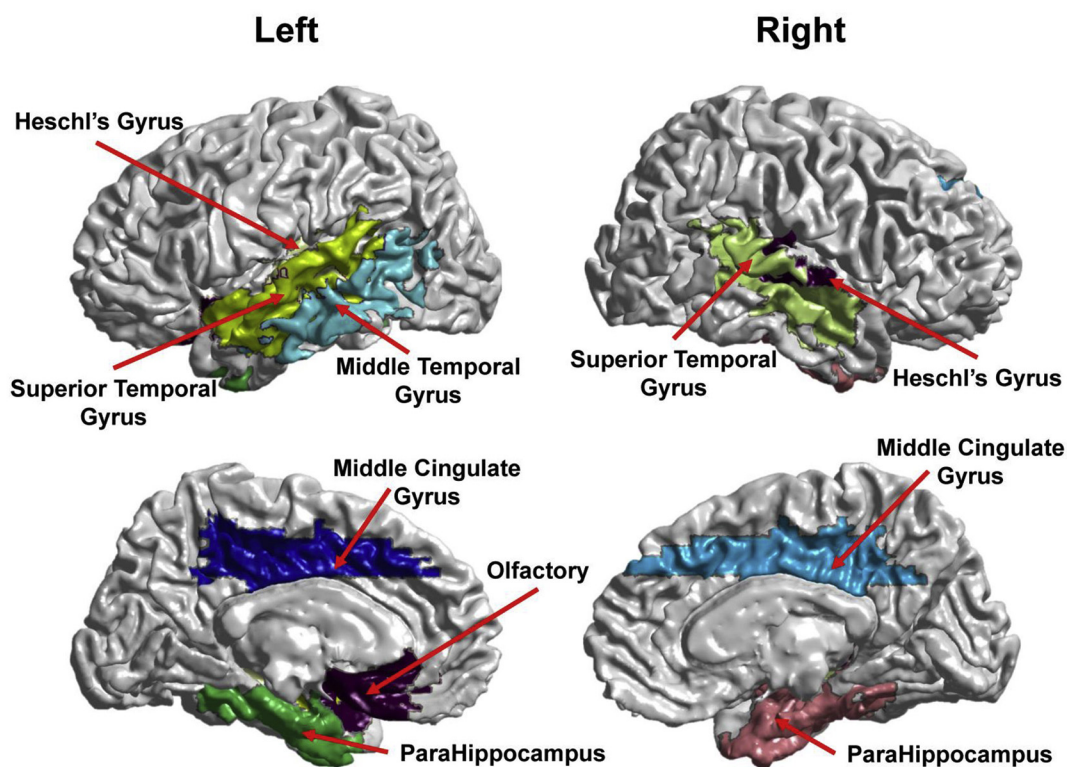


Fig. 6. Top 10 cortical regions for most discriminating (A) AD and (B) LMCI from CN.

used. This performance gain via fine-tuning may suggest that our proposed framework is capable of capturing morphological changes of the brain that are essential to AD/MCI regardless the populations. Our results further demonstrated the applicability of the proposed framework to a small dataset of another population. In neural network, transferring knowledge or features learned from one task/dataset to another is often achieved by updating only the last few layers of a pre-trained network (Yosinski et al., 2014). As the base and target tasks (e.g., MCI and AD classifications) were the same, we ensured that the dementia-associated knowledge learned from the Caucasian population (i.e., ADNI-2 dataset) could be directly transferred to the Asian population by freezing the convolutional layers during the fine-tuning process. The graph convolutional layers play a feature detector role to learn dementia-associated knowledge of the whole brain morphology from the Caucasian population, while the fully connected and logit layers form a reasoning module for making an inference of clinical diagnosis based on the learned knowledge for the Asian population. Our findings showed that the trained model consistently achieved better or comparable performance than the learn-from-scratch model (Fig. 5), suggesting that both the Caucasian and Asian populations shared neural morphological patterns for AD/MCI. Of note, the pre-trained model fine-tuned using Moderate MCI subjects of the Asian population could accurately identify the AD subjects of the same population, implying that relevant knowledge of the target population, even with the less severe condition, is able to efficiently improve the AD prediction. Taking together, our results demonstrated the feasibility of applying the spectral graph-CNN trained based on the ADNI-2 cohort to other population. Such finding is crucial evidence for the generalization of existing knowledge across populations for early diagnosis and prognosis of AD. This sheds light on learning of cross-datasets and cross-populations for image-based brain disease diagnosis. This is in line with the conclusion drawn from previous medical image-based computer-aided detection studies that transfer learning could be a useful technique to mitigate the issue due to a small well-annotated dataset in the medical imaging domain (Li et al., 2014; Shin et al., 2016).

Our spectral graph CNN-based framework achieved decent and comparable performance when compared to state-of-the-art methods for the CN vs. AD classification based on the ADNI-1 or ADNI-2 dataset (Table 5). Our spectral graph-CNN achieved accuracies of 89.4% and 85.8% for the ADNI-1 and ADNI-2 cohort respectively, which are better than the methods with ROI-based cortical thickness as features (< 85%) (Eskildsen et al., 2013; Wee et al., 2012). Compared to the other methods listed in Table 5, the spectral graph-CNN had comparable performance. However, it performed slightly lower than that in (Aderghal et al., 2017; Liu et al., 2012; Suk et al., 2017), perhaps

because of differences in the sample size. Nevertheless, the published results (Table 5) were based on the relatively balanced samples between CN and AD, while our study incorporated the full sample of the ADNI-1 and ADNI-2 cohorts and the imbalanced samples. It has been indicated that classification performance decreases when the sample size increases (Mendelson et al., 2017). However, a classifier built based on a large sample is relatively robust. Our study showed the robustness of the spectral graph-CNN-based framework on the imbalanced samples in AD and CN as both sensitivity and specificity rates were relatively similar.

In the classification of AD and LMCI, the majority of the most discriminating regions were located at the temporal lobe (the temporal and Heschl's gyri), the parahippocampus, and the cingulate gyri. Our findings are in accordance with previous studies on both AD and MCI populations that reported significant atrophy within these brain regions (Chan et al., 2001; Dickerson et al., 2009; Eskildsen et al., 2015; Fjell et al., 2014; Jack et al., 1998; Lerch et al., 2008; Singh et al., 2006; Visser et al., 2002). The medial temporal lobe (MTL), associated with memory loss (Gold and Budson, 2008; Jahn, 2013; Petersen et al., 1994; Squire et al., 2007), was affected during the course of the disease. The parahippocampal atrophy, identified as a biomarker at the early phase of AD (Echavarrri et al., 2011), supports our finding on the most discriminating power of the bilateral parahippocampal gyrus for LMCI. Moreover, cortical thicknesses of the superior temporal gyrus and the parahippocampus was also found effective in predicting the conversion of MCI to AD (Eskildsen et al., 2015). In addition to significant cortical thinning in the parahippocampus and the temporal gyri, AD patients also showed thinning in the anterior and posterior cingulate gyri (Lerch et al., 2005). An existing longitudinal study suggested a significant correlation between cortical thinning in the parahippocampus and the anterior cingulate gyrus with the progression of AD based on their thinning rate from the mild to severe stages of the disease (Lerch et al., 2005). In a volumetric analysis, smaller cortical volumes were observed in the anterior and posterior cingulate gyri in AD compared to healthy controls (Jones et al., 2006).

The present study posed several limitations. First, the ADNI project is a continuous longitudinal study with most of the individuals have been followed up for years and scanned for multiple times. In this study, we, however, treated each scan as an individual sample to increase the number of training sample. Longitudinal brain morphological changes have been proven as a useful neuroimaging biomarker not only for AD/MCI identification (Li et al., 2012) but also potentially for predicting conversion and time-to-conversion of an individual to other clinical conditions (Li et al., 2012; Thung et al., 2018). We believe that incorporating the longitudinal morphological pattern would be beneficial

Table 5
Classification accuracy between Alzheimer's disease (AD) patients and normal controls from the ADNI cohorts.

Study	Feature type	Classifier	Samples (AD/CN)	ACC (%)	SEN (%)	SPE (%)
(Coupé et al., 2012)	HP/EC volume	QDA	60/60	90.0	88.0	92.0
(Schmitter et al., 2015)	10 volumes	SVM	221/276	–	86.0	91.0
(Zhang and Shen, 2012)	GM volume	SVM	45/50	84.8	–	–
(Suk et al., 2015)	GM volume	SAE + SVM	51/52	88.2	–	–
(Suk et al., 2017)	GM volume	JLLR + DeepESM	186/226	91.0	92.7	89.9
(Luo et al., 2017)	Whole brain (patch-based)	2D CNN	49/30	–	69.0	98.0
(Liu et al., 2012)	GM voxels	Ensemble SRC	198/229	90.8	86.3	94.8
(Casanova et al., 2013)	GM voxels	RLR	171/188	87.1	84.3	88.9
(Aderghal et al., 2017)	Bilateral HP	2D CNN	188/228	91.4	93.8	89.1
(Wee et al., 2012)	ROI-based CT	SVM	200/198	84.7	82.8	86.5
(Eskildsen et al., 2013)	ROI-based CT	LDA	194/226	84.5	79.4	88.9
(Cho et al., 2012)	Vertex-based CT	PCA + LDA	128/160	–	82.0	93.0
Proposed	CT graph	Graph-CNN	592/960	85.8	83.5	87.5

Abbreviations. GM: Gray Matter; SVM: Support Vector Machine; CT: Cortical Thickness; ROI: Region-Of-Interest; PCA: Principal Component Analysis; LDA: Linear Discriminant Analysis; ROI: Region-Of-Interest; QDA: Quadratic Discriminant Analysis; SRC: Sparse Regression Classifier; RLR: Regularized Linear Regression; JLLR: Joint Linear and Logistic Regression; SAE: Stacked Auto-Encoder; DeepESM: Deep Ensemble Sparse Model; HP: Hippocampus; EC: Entorhinal Cortex; ACC: Accuracy; SEN: Sensitivity; SPE: Specificity.

to the proposed cortical graph based CNN model to improve the AD/MCI diagnosis and prognosis accuracy. Second, the most discriminating regions for AD and LMCI in this study were determined based on a leave-one-ROI approach, in which only the contribution of a brain region was quantified at each time. Although the most discriminating regions have been consistently reported to be associated with the AD pathology, the leave-one-ROI approach may be less comprehensive in reflecting the contribution of a group of the brain regions as the nature of deep neural network is to characterize multivariate features and their interactions (Du et al., 2018; Vieira et al., 2017). In addition, we identified discriminating brain regions of the spectral graph-CNN classifiers based on the AAL atlas. Other atlases can also be used. However, our approach may not work to identify discriminating regions at a vertex level. This could be due to a limited contribution of each vertex to the discriminating power of the disease classes. Furthermore, we noted that the classification accuracy of MCI/AD in existing literature based on multi-modal imaging data (e.g., (Yu et al., 2016; Zhang and Shen, 2012; Zhang et al., 2011; Zhu et al., 2014; Zhu et al., 2017)) may be higher than that provided by this study based only on cortical thickness. Our study could not incorporate multi-modal imaging data partly because other imaging modalities, such as PET, DTI, and fMRI, were collected in a subset of the subjects in ADNI. This restricts the robustness of the training of the spectral graph-CNN. Nevertheless, our framework can be extended to a multi-channel spectral graph-CNN to incorporate multi-modal imaging data represented on the cortical surface.

5. Conclusions

In this paper, we employed the spectral graph-CNN based framework that utilized cortical thickness and its underlying geometric information for AD and MCI diagnosis and MCI-to-AD conversion. We evaluated the effectiveness of our framework using 3089 MRI scans from the publicly available ADNI-2 cohort and demonstrated state-of-the-art dementia classification performance. The spectral graph-CNN on cortical thickness outperformed the voxel-based CNN models. Furthermore, our spectral graph-CNN was able to achieve a relatively balanced prediction for the two classes with the imbalanced sample sizes. We achieved this via training mini-batch with balanced samples in two classes and oversampling of the class with a small sample size, and determining the optimal network parameters based on the adjusted geometric mean that emphasizes balanced performance for both the majority and minority classes. The spectral graph-CNN also achieved high prediction accuracy for the conversion of MCI to AD. When transferring the spectral graph-CNN model trained on the ADNI-2 cohort to the Asian cohort, the spectral graph-CNN consistently achieved better classification performance compared to the same model that trained directly on the Asian cohort. Furthermore, the trained spectral graph-CNN model that fine-tuned on MCI subjects was able to accurately identify AD subjects from the Asia cohort, suggesting the feasibility of applying the existing robust classifiers for early diagnosis and prognosis of MCI and AD of new populations. The spectral graph-CNN model proposed in this study is relatively general and can be applied to other brain imaging data for the early diagnosis of brain disorders. Further improvement can be obtained when integrating the spectral graph-CNN on cortical thickness with other classification approaches on multi-modal brain image data.

Acknowledgments

Singapore Ministry of Education (Academic research fund tier 1; NUHSRO/2017/052/T1-SRP-Partnership/01), and NUS Institute of Data Science, Singapore.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2019.101929>.

References

- Aderghal, K., Benois-Pineau, J., Afdel, K., 2017. Classification of sMRI for Alzheimer's disease diagnosis with CNN: single siamese networks with 2D + ? In: Approach and Fusion on ADNI. Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval. ACM, Bucharest, Romania, pp. 494–498.
- Alzheimer's Association, 2015. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement.* 11, 332–384.
- Barandela, R., Sánchez, J.S., García, V., Rangel, E., 2003. Strategies for learning in class imbalance problems. *Pattern Recogn.* 36, 849–851.
- Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., Filippi, M., Alzheimer's Disease Neuroimaging, I., 2019. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin.* 21, 101645.
- Casanova, R., Hsu, F.-C., Sink, K.M., Rapp, S.R., Williamson, J.D., Resnick, S.M., Espeland, M.A., for the Alzheimer's Disease Neuroimaging Initiative, 2013. Alzheimer's Disease risk assessment using large-scale machine learning methods. *PLoS ONE* 8, e77949.
- Chan, D., Fox, N.C., Scallan, R.I., Crum, W.R., Whitwell, J.L., Leschziner, G., Rossor, A.M., Stevens, J.M., Cipolotti, L., Rossor, M.N., 2001. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann. Neurol.* 49, 433–442.
- Chin, J.J., 2002. Screening for cognitive impairment in primary care: the role of objective tests. *Singapore Fam. Phys.* 28, 62–68.
- Cho, Y., Seong, J.-K., Jeong, Y., Shin, S.Y., 2012. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* 59, 2217–2230.
- Cotter, A., Shamir, O., Srebro, N., Sridharan, K., 2011. Better Mini-Batch Algorithms via Accelerated Gradient Methods. *NIPS*, pp. 1647–1655.
- Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Collins, D.L., 2012. Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease. *Neuroimage* 59, 3736–3747.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Neuroimaging, A.D., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56, 766–781.
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 3844–3852.
- Dickerson, B.C., Feczko, E., Augustinack, J.C., Pacheco, J., Morris, J.C., Fischl, B., Buckner, R.L., 2009. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol. Aging* 30, 432–440.
- Du, A.-T., Schuff, N., Kramer, J.H., Rosen, H.J., Gorno-Tempini, M.L., Rankin, K., Miller, B.L., Weiner, M.W., 2007. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain J. Neurol.* 130, 1159–1166.
- Du, J., Younes, L., Qiu, A., 2011. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. *Neuroimage* 56, 162–173.
- Du, Y., Fu, Z., Calhoun, V.D., 2018. Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Front. Neurosci.* 12, 525.
- Dyrba, M., Ewers, M., Wegrzyn, M., Kilimann, I., Plant, C., Oswald, A., Meindl, T., Pievani, M., Bokde, A.L.W., Fellgiebel, A., Filippi, M., Hampel, H., Klöppel, S., Hausenstein, K., Kirste, T., Teipel, S.J., E.S.G., 2013. Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. *PLoS ONE* 8, e64925.
- Dyrba, M., Barkhof, F., Fellgiebel, A., Filippi, M., Hausner, L., Hausenstein, K., Kirste, T., Teipel, S.J., 2015. Predicting prodromal Alzheimer's disease in subjects with mild cognitive impairment using machine learning classification of multimodal multicenter diffusion-tensor and magnetic resonance imaging data. *J. Neuroimaging* 25, 738–747.
- Echavarrri, C., Aalten, P., Uylings, H.B.M., Jacobs, H.I.L., Visser, P.J., Gronenschild, E.H.B.M., Verhey, F.R.J., Burgmans, S., 2011. Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. *Brain Struct. Funct.* 215, 265–271.
- Eshkoor, S.A., Hamid, T.A., Mun, C.Y., Ng, C.K., 2015. Mild cognitive impairment and its management in older people. *Clin. Interv. Aging* 10, 687.
- Eskildsen, S.F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* 65, 511–521.
- Eskildsen, S.F., Coupé, P., Fonov, V.S., Pruessner, J.C., Collins, D.L., 2015. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiol. Aging* 36, S23–S31.
- Ewers, M., Sperling, R.A., Klunk, W.E., Weiner, M.W., Hampel, H., 2011. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci.* 34, 430–442.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Dale, A.M., Walhovd, K.B., Alzheimer Disease Neuroimaging, I., 2014. Accelerating

- cortical thinning: unique to dementia or universal in aging? *Cereb. Cortex* New York, N.Y. 24, 919–934.
- Gold, C.A., Budson, A.E., 2008. Memory loss in Alzheimer's disease: implications for development of therapeutics. *Expert. Rev. Neurother.* 8, 1879–1891.
- Gupta, A., Ayhan, M.S., Maida, A.S., 2013. Natural image bases to represent neuroimaging data. In: *ICML*, pp. III-987–III-994.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. 2016 IEEE Conf. Comp. Vis. Pattern R. (CVPR) 770–778.
- Jack, C.R., Petersen, R.C., Xu, Y., Brien, P.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1998. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* 51, 993.
- Jahn, H., 2013. Memory loss in Alzheimer's disease. *Dialogues Clin. Neurosci.* 15, 445–454.
- Jones, B.F., Barnes, J., Uylings, H.B.M., Fox, N.C., Frost, C., Witter, M.P., Scheltens, P., 2006. Differential regional atrophy of the cingulate gyrus in Alzheimer disease: a volumetric MRI study. *Cereb. Cortex* 16, 1701–1708.
- Karakaya, T., Fußer, F., Schröder, J., Pantel, J., 2013. Pharmacological treatment of mild cognitive impairment as a prodromal syndrome of Alzheimer's disease. *Curr. Neuropharmacol.* 11, 102–108.
- Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourao-Miranda, J., Vemuri, P., 2012. Diagnostic neuroimaging across diseases. *Neuroimage* 61, 457–463.
- Korolev, S., Safiullin, A., Belyaev, M., Dodonova, Y., 2017. Residual and plain convolutional neural networks for 3D brain MRI classification. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 835–838.
- Lawlor, B., Segurado, R., Kennelly, S., Olde Rikkert, M.G.M., Howard, R., Pasquier, F., Björksson-Hanson, A., Tsolaki, M., Lucca, U., Molloy, D.W., Coen, R., Riepe, M.W., Kálmán, J., Kenny, R.A., Clegg, F., O'Dwyer, S., Walsh, C., Adams, J., Banzi, R., Breuili, L., Daly, L., Hendrix, S., Aisen, P., Gaynor, S., Sheikh, A., Taekema, D.G., Verhey, F.R., Nemni, R., Nobili, F., Franceschi, M., Frisoni, G., Zanetti, O., Konsta, A., Anastasios, O., Nenopoulou, S., Tsolaki-Tagaraki, F., Pakaski, M., Dereeper, O., de la Sayette, V., Sènechal, O., Lavenex, I., Devendeville, A., Calais, G., Crawford, F., Mullan, M., for the N.S.G., 2018. Nilvadipine in mild to moderate Alzheimer disease: a randomised controlled trial. *PLoS Med.* 15, e1002660.
- Ledig, C., Schuh, A., Guerrero, R., Heckemann, R.A., Rueckert, D., 2018. Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Sci. Rep.* 8.
- Lerch, J.P., Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J., Evans, A.C., 2005. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb. Cortex* 15, 995–1001.
- Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol. Aging* 29, 23–30.
- Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S., 2014. Deep learning based imaging data completion for improved brain disease diagnosis. *MICCAI* 17, 305–312.
- Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D., 2012. Alzheimer's Disease Neuroimaging, I, 2012. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol. Aging* 33 (427.e415–427.e430).
- Liu, M., Zhang, D., Shen, D., 2012. Ensemble sparse classification of Alzheimer's disease. *Neuroimage* 60, 1106–1116.
- Liu, X., Tosun, D., Weiner, M.W., Schuff, N., for the Alzheimer's Disease Neuroimaging Initiative, 2013. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. *Neuroimage* 83, 148–157.
- Liu, M., Zhang, D., Shen, D., the Alzheimer's Disease Neuroimaging Initiative, 2014. Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. *Hum. Brain Mapp.* 35, 1305–1319.
- Liu, M., Zhang, D., Shen, D., 2015. View-centralized multi-atlas classification for Alzheimer's disease diagnosis. *Hum. Brain Mapp.* 36, 1847–1865.
- López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* 250, 113–141.
- Luo, S., Li, X., Li, J., 2017. Automatic Alzheimer's disease recognition from mri data using deep learning method. *J. Appl. Math. Phys.* 5, 1892–1898.
- Mendelson, A.F., Zuluaga, M.A., Lorenzi, M., Hutton, B.F., Ourselin, S., 2017. Selection bias in the reported performances of AD classification pipelines. *NeuroImage* 14, 400–416.
- Möller, C., Pijnenburg, Y.A.L., van der Flier, W.M., Versteeg, A., Tijms, B., de Munck, J.C., Hafkemeijer, A., Rombouts, S.A.R.B., van der Grond, J., van Swieten, J., Dopfer, E., Scheltens, P., Barkhof, F., Vrenken, H., Wink, A.M., 2016. Alzheimer disease and behavioral variant frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. *Radiology* 279, 838–848.
- Nair, V., Hinton, G.E., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. *ICML*, pp. 807–814.
- Niepert, M., Ahmed, M., Kutzkov, K., 2016. Learning Convolutional Neural Networks for Graphs. *ICML*, pp. 2014–2023.
- Nir, T.M., Villalon-Reina, J.E., Prasad, G., Jahanshad, N., Joshi, S.H., Toga, A.W., Bernstein, M.A., Jack, C.R., Weiner, M.W., Thompson, P.M., 2015. DTI-based maximum density path analysis and classification of Alzheimer's disease. *Neurobiol. Aging* 36, S132–S140.
- Payan, A., Montana, G., 2015. Predicting Alzheimer's Disease: A Neuroimaging Study with 3d Convolutional Neural Networks. *ICPRAM*.
- Pellegrini, E., Ballerini, L., Hernandez, M., Chappell, F.M., Gonzalez-Castro, V., Anlagan, D., Danso, S., Munoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T.J., Trucco, E., Wardlaw, J.M., 2018a. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement.(Amst)* 10, 519–535.
- Pellegrini, E., Ballerini, L., Hernandez, M.V., Chappell, F.M., Gonzalez-Castro, V., Anlagan, D., Danso, S., Maniega, S.M., Job, D., Pernet, C., Mair, G., MacGillivray, T., Trucco, E., Wardlaw, J., 2018b. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement.* 10, 519–535.
- Petersen, R.C., Smith, G.E., Ivnik, R.J., Kokmen, E., Tangalos, E.G., 1994. Memory function in very early Alzheimer's disease. *Neurology* 44, 867–872.
- Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, M., Alzheimer's Disease International, 2015. World Alzheimer report 2015. In: *The Global Impact of Dementia. An analysis of Prevalence, Incidence, Costs and Trends*.
- Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., The Alzheimer's Disease Neuroimaging, I, 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 132, 2036–2047.
- Racine, A., Brickhouse, M., Wolk, D., Dickerson, B., Alzheimer's Disease Neuroimaging Initiative, 2018. The personalized Alzheimer's disease cortical thickness index predicts likely pathology and clinical progression in mild cognitive impairment. *Alzheimers Dement.(Amst)* 10, 301–310.
- Rathore, S., Habes, M., Ifthikhar, M.A., Shacklett, A., Davatzikos, C., 2017. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* 155, 530–548.
- Rizzi, L., Rosset, I., Roriz-Cruz, M., 2014. Global epidemiology of dementia: Alzheimer's and vascular types. *Biomed. Res. Int.* 2014, 8.
- Robinson, L., Tang, E., Taylor, J.-P., 2015. Dementia: timely diagnosis and early intervention. *BMJ* 350.
- Schaerer, J., Sampat, M., Roche, F., Oh, J., Bracoud, L., Suh, J., 2016. Assessing regional cortical thickness for predicting MCI conversion to Alzheimer's disease. *Alzheimers Dement.* 12, P1096–P1097.
- Schmitter, D., Roche, A., Maréchal, B., Ribes, D., Abdulkadir, A., Bach-Cuadra, M., Daducci, A., Granziera, C., Klöppel, S., Maeder, P., Meuli, R., Krueger, G., 2015. An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *NeuroImage* 7, 7–17.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 61–80.
- Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*.
- Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E., Kabani, N.J., 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain* 129, 2885–2893.
- Squire, L.R., Zola-Morgan, J., Clark, R.E., 2007. Recognition memory and the medial temporal lobe: a new perspective. *Nat. Rev. Neurosci.* 8, 872–883.
- Suk, H.-I., Lee, S.-W., Shen, D., The Alzheimer's Disease Neuroimaging Initiative, 2015. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct. Funct.* 220, 841–859.
- Suk, H.-I., Lee, S.-W., Shen, D., 2017. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med. Image Anal.* 37, 101–113.
- Tan, M., Qiu, A., 2016. Large deformation multiresolution diffeomorphic metric mapping for multiresolution cortical surfaces: a coarse-to-fine approach. *IEEE Trans. Image Process.* 25, 4061–4074.
- Tan, M., Qiu, A., 2018. Multiscale frame-based kernels for large deformation diffeomorphic metric mapping. *IEEE Trans. Med. Imaging* 37 (10), 2344–2355.
- Thong, J.Y.J., Hilal, S., Wang, Y., Soon, H.W., Dong, Y., Collinson, S.L., Anh, T.T., Ikram, M.K., Wong, T.Y., Venketasubramanian, N., Chen, C., Qiu, A., 2013. Association of silent lacunar infarct with brain atrophy and cognitive impairment. *J. Neurol. Neurosurg. Psychiatry* 84, 1219–1225.
- Thong, J.Y.J., Du, J., Ratnarajah, N., Dong, Y., Soon, H.W., Saini, M., Tan, M.Z., Tuan Ta, A., Chen, C., Qiu, A., 2014. Abnormalities of cortical thickness, subcortical shapes, and white matter integrity in subcortical vascular cognitive impairment. *Hum. Brain Mapp.* 35, 2320–2332.
- Thung, K.-H., Yap, P.-T., Adeli, E., Lee, S.-W., Shen, D., 2018. Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. *Med. Image Anal.* 45, 68–82.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliet, M., 2002. Automated anatomical Labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- Vieira, S., Pinaya, W.H.L., Mechelli, A., 2017. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci. Biobehav. Rev.* 74, 58–75.
- Visser, P.J., Verhey, F.R.J., Hofman, P.A.M., Scheltens, P., Jolles, J., 2002. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *J. Neurol. Neurosurg. Psychiatry* 72, 491.
- Wee, C.-Y., Yap, P.-T., Shen, D., 2012. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Hum. Brain Mapp.* 34, 3411–3425.
- Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, A.M., Winblad, B., Jönsson, L., Liu, Z., Prince, M., 2017. The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimers Dement.* 13, 1–7.
- Xu, C., Tao, D., Xu, C., 2013. A Survey on Multi-View Learning. *CoRR*.
- Yiannopoulou, K.G., Papageorgiou, S.G., 2013. Current and future treatments for Alzheimer's disease. *Ther. Adv. Neurol. Disord.* 6, 19–33.

- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How Transferable are Features in Deep Neural Networks? NIPS, pp. 3320–3328.
- Yu, G., Liu, Y., Shen, D., 2016. Graph-guided joint prediction of class label and clinical scores for the Alzheimer's disease. *Brain Struct. Funct.* 221, 3787–3801.
- Zhang, D., Shen, D., 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* 59, 895–907.
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Alzheimer's Disease Neuroimaging, I., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867.
- Zhong, J., Phua, D.Y., Qiu, A., 2010. Quantitative evaluation of LDDMM, FreeSurfer, and CARET for cortical surface mapping. *Neuroimage* 52, 131–141.
- Zhu, X., Suk, H.-I., Shen, D., 2014. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *Neuroimage* 100, 91–105.
- Zhu, X., Suk, H.-I., Wang, L., Lee, S.-W., Shen, D., 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Med. Image Anal.* 38, 205–214.