Attention-Guided Hybrid Network for Dementia Diagnosis With Structural MR Images

Chunfeng Lian⁽⁾, Member, IEEE, Mingxia Liu⁽⁾, Senior Member, IEEE,

Yongsheng Pan¹⁰, Graduate Student Member, IEEE, and Dinggang Shen¹⁰, Fellow, IEEE

Abstract—Deep-learning methods (especially convolutional neural networks) using structural magnetic resonance imaging (sMRI) data have been successfully applied to computer-aided diagnosis (CAD) of Alzheimer's disease (AD) and its prodromal stage [i.e., mild cognitive impairment (MCI)]. As it is practically challenging to capture local and subtle disease-associated abnormalities directly from the whole-brain sMRI, most of those deep-learning approaches empirically preselect disease-associated sMRI brain regions for model construction. Considering that such isolated selection of potentially informative brain locations might be suboptimal, very few methods have been proposed to perform disease-associated discriminative region localization and disease diagnosis in a unified deep-learning framework. However, those methods based on task-oriented discriminative localization still suffer from two common limitations, that is: 1) identified brain locations are strictly consistent across all subjects, which ignores the unique anatomical characteristics of each brain and 2) only limited local regions/patches are used for model training, which does not fully utilize the global structural information provided by the whole-brain sMRI. In this article, we propose an attention-guided deep-learning framework to extract multilevel discriminative sMRI features for dementia diagnosis. Specifically, we first design a backbone fully convolutional network to automatically localize the discriminative brain regions in a weakly supervised manner. Using the identified disease-related regions as spatial attention guidance, we further develop a hybrid network to jointly learn and fuse multilevel sMRI features for CAD model construction. Our proposed method was evaluated on three public datasets (i.e., ADNI-1, ADNI-2, and AIBL), showing superior performance compared with several state-of-the-art methods in both tasks of AD diagnosis and MCI conversion prediction.

Index Terms-Alzheimer's disease (AD), convolutional neural networks (CNNs), multilevel feature learning, structural MRI, weakly supervised localization.

Manuscript received February 25, 2019; revised August 28, 2019 and May 31, 2020; accepted June 19, 2020. The work of Chunfeng Lian and Mingxia Liu was supported in part by NIH under Grant AG041721 and Grant AG053867. This article was recommended by Associate Editor D. Goldgof. (Chunfeng Lian and Mingxia Liu are co-first authors.) (Corresponding author: Dinggang Shen.)

Chunfeng Lian and Mingxia Liu are with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Yongsheng Pan is with the School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China.

Dinggang Shen is with the Department of Research and Development, Shanghai United Imaging Intelligence Company Ltd., Shanghai 200232, China, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: dinggang.shen@gmail.com).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.3005859

I. INTRODUCTION

S TRUCTURAL magnetic resonance imaging (sMRI) is an important tool in Alzheimer's disease (AD) diagnosis, since it is sensitive to dementia-induced anatomical abnormalities in the brain [1]. Capitalizing on sMRI data, various machine-learning approaches [2]-[12] and modern deep-learning methods [13]-[23] have been proposed for computer-aided diagnosis (CAD) of AD and its prodromal stage [i.e., mild cognitive impairment (MCI)]. Typically, there are three basic steps in constructing a CAD system based on sMRI data [24], including: 1) regions-of-interest (ROIs) identification from the whole-brain sMRI; 2) feature extraction from ROIs; and 3) classifier construction based on sMRI features.

Conventional machine-learning methods [2]-[12], [25]-[27] typically perform these three steps separately. They first predetermine possibly informative ROIs (e.g., coarse brain regions or local image patches) according to domain-specific prior knowledge, and then extract handcrafted features from these ROIs to train classification models. Recently, deep-learning methods [13]-[18], [20], [21] have also been proposed for AD-related brain disease diagnosis using sMRI scans, in which feature extraction and classifier construction are unified to learn high-level features in a task-oriented manner. Although these deep-learning methods greatly improve diagnostic performance, they have a common limitation with the traditional machine-learning methods using handcrafted features. That is, the independent preidentification of ROIs (e.g., coarse brain regions [17] or local image patches [18]) might not be well coordinated with the latter steps of feature extraction and classifier construction, thus leading to suboptimal performance.

To tackle this limitation, in a more recent work [22], a hierarchical fully convolutional network (H-FCN) was proposed to jointly identify multiscale discriminative locations in brain sMRI scans and learn high-level sMRI features for constructing a hierarchical diagnostic model, which yielded state-ofthe-art diagnostic performance, especially in the challenging task of MCI conversion prediction (i.e., whether subjects with MCI would convert to AD within a certain period). However, since H-FCN relies on location proposals that are anatomically consistent across all subjects for discriminative localization and disease diagnosis [22], it still has two inherent disadvantages. First, similar to predetermined ROIs used in the previous CAD approaches for AD diagnosis, the multiscale (i.e., patch and region level) discriminative locations identified

2168-2267 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.



Fig. 1. Illustration of our proposed attention-guided deep-learning framework to automatically extract and integrate multilevel discriminative information from the whole-brain sMRI scans for dementia diagnosis. In the first stage, a backbone FCN is designed for task-oriented localization of discriminative regions in each input sMRI scan, yielding a DAM. In the second stage, with the identified DAM as guidance, we further develop a multibranch hybrid network for dementia diagnosis. In a DAM, the red and blue colors denote strong and weak discriminative capability, respectively.

by H-FCN are consistent across all subjects. This implies that H-FCN has the same assumption as traditional methods, that is, different subjects have a common set of potentially informative ROIs. Therefore, these methods ignore the fact that, for different individuals, the AD-associated brain atrophy may occur at different brain locations or have different degrees of severity in certain brain regions. Second, similar to some other conventional learning methods [8] and deeplearning methods [18], H-FCN employs local image patches as network inputs, thus quantifying global information of the whole brain by integrating local information provided by limited image patches. Notably, a network pruning strategy is used in H-FCN to directly remove uninformative or less informative image patches and brain regions, ignoring the fact that those pruned patches/regions could contain supplementary information (when combined with other distinctive patches/regions) for robust model training. That is, H-FCN using only pruned image patches (although integrated later by the network) may result in incomplete quantification of global information for each brain sMRI; thus, it is highly desired to directly employ global structural information of the brain to boost the diagnostic performance.

In this article, we propose an attention-guided deep-learning framework to integrate multilevel (i.e., intersubject consistent and individual specific, and local and global) sMRI information for brain disease diagnosis. Using directly the whole-brain sMRI as input, the schematic of our CAD framework is illustrated in Fig. 1. In the first stage, a backbone fully convolutional network (FCN) is trained using solely the subject/image-level supervision (i.e., class labels) for weakly supervised detection of disease-related discriminative locations [i.e., generation of disease attention maps (DAMs)] in each individual (whole-brain) sMRI. Guided by task-oriented DAMs, in the second stage, a multibranch hybrid network (HybNet) is built on the input sMRI as well as the intermediate feature maps of the backbone FCN for brain disease diagnosis. Especially, besides using the local structure information conveyed in image patches, we apply a lightweight convolutional neural network (CNN) on the backbone feature map of each subject, spatially weighted by its corresponding DAM, to directly capture both global and individual-specific discriminative information. We evaluated our proposed method for both the tasks of AD classification and MCI conversion prediction on three datasets, including AD Neuroimaging Initiative-1 (ADNI-1), ADNI-2, and Australian Imaging, Biomarker and Lifestyle Flagship Study of Aging (AIBL). Experimental

results demonstrate that our proposed method performs well in both discriminative localization and brain disease diagnosis, compared with several baseline and state-of-the-art methods.

The remainder of this article is organized as follows. In Section II, we briefly review related work on automated AD diagnosis using sMRI data. In Section III, we introduce the studied datasets (i.e., ADNI-1, ADNI-2, and AIBL) as well as the data preprocessing pipeline. The proposed method is described in detail in Section IV. In Section V, we present the experimental settings, competing methods, and experimental results. We finally conclude this article and briefly discuss the limitations of our current work in Section VII.

II. RELATED WORK

In terms of the ROIs for feature extraction and classifier construction, existing methods for AD diagnosis were usually constructed based on voxel-level, region-level, and/or patch-level pattern analyses. In this section, we briefly review conventional learning approaches as well as deep-learning approaches in the literature that perform AD diagnosis using primarily different scales of sMRI features.

A. Conventional Learning Approaches

Conventional learning approaches typically predetermine ROIs to extract handcrafted features (e.g., brain tissue density map, textural features, and shape features) to construct AD diagnostic models. Using the whole brain (i.e., all voxels in an sMRI) as input, some methods adopt voxel-level feature representations to train classifiers for differentiating between patients and normal controls (NCs). For example, based on voxel-based morphometry (VBM) [28], [29], the gray matter (GM) density map was quantified in [3] and [4] to construct a linear programming boosting (LPboosting) classifier [30] and a linear support vector machine (SVM) classifier [31] for AD diagnosis, respectively.

Another commonly used strategy is to preselect one or multiple brain regions according to anatomical prior knowledge, and then extract regional sMRI features for the construction of a diagnostic model. For example, in [2] and [10], the left and right hippocampi were preidentified from sMRI to extract shape and textural features for AD diagnosis, respectively. By parcellating the whole-brain sMRI into multiple nonoverlapping regions, regional GM volumes were quantified in [5] to train SVM classifiers for AD and MCI classification. Based on anatomical prior knowledge provided by multiple atlases, [6] and [9] extracted multitemplate regional features to construct ensemble classifiers for AD classification and MCI conversion prediction.

Alternatively, considering that structural changes induced by the early stage of AD (e.g., MCI) could be subtle, some methods were proposed to extract features from local image patches. For example, based on local patches extracted at the locations of hippocampus and entorhinal cortex, a previous study [7] defined volumetric and grading biomarkers for AD classification and MCI conversion prediction. In [8], a multiple instance learning (MIL) model [32] was trained to combine the information provided by a set of local patches for AD classification and MCI conversion prediction. By defining AD-related anatomical landmarks, Zhang *et al.* [11] extracted morphological features from local patches centered at these landmarks to train SVM classifiers for AD and MCI classification. The key issue for these conventional learning approaches is that the independent extraction of handcrafted features and construction of classifiers may hamper the diagnostic performance, due to potential heterogeneity between these two steps.

B. Deep-Learning Approaches

Various deep-learning approaches have also been proposed for AD diagnosis [13]–[18], [20]–[22], [33]–[35], among which CNNs were commonly used to extract task-oriented features from sMRI for constructing diagnostic models.

As it is challenging to capture subtle structural changes from the whole brain using directly an end-to-end network without any guidance, similar to conventional learning methods, most of the existing CNN-based diagnostic models require the predetermination of potentially informative ROIs according to domain knowledge and experts' experience. For example, the sMRI hippocampal regions were coarsely extracted in [15] to train a residual network [36] for the multicategory diagnosis of AD. In [18], CNN-based deep MIL models were constructed using local image patches preidentified by anatomical landmarks for AD classification and MCI conversion prediction. In [21], local image patches were extracted from both sMRI and positron emission tomography (PET) scans to construct cascaded CNNs for AD and MCI classification.

Considering that preselection of ROIs isolated to CNNbased feature extraction and classifier construction might not be an optimal solution, a hierarchical FCN (H-FCN) was proposed in [22] to integrate the ROI identification, feature extraction, and classifier construction into a unified taskoriented framework. Specifically, based on the patch-level location proposals widely distributed over the whole brain, H-FCN automatically identified multiscale (i.e., patch level and region level) discriminative locations to construct a hierarchical diagnostic model for AD classification and MCI conversion prediction. However, the H-FCN method employed a network pruning strategy to directly remove uninformative and less informative image patches and brain regions, ignoring the fact that those pruned patches/regions could contain supplementary information for robust model training. This may result in incomplete quantification of global brain structure information, leading to suboptimal diagnostic performance.

III. MATERIALS

A. Data

Three public datasets with 1976 *baseline* sMRI scans were studied in this work, including: 1) AD Neuroimaging Initiative-1 (ADNI-1) [37]; 2) ADNI-2 [37]; and 3) Australian Imaging, and Biomarker and Lifestyle Flagship Study of Aging (AIBL) [38]. The demographic information of the subjects included in these three datasets is summarized in Table I.

TABLE IBASELINE DEMOGRAPHIC INFORMATION OF SUBJECTS INCLUDED IN THETHREE PUBLIC DATASETS (I.E., ADNI-1, ADNI-2, AND AIBL). THEGENDER IS PRESENTED AS MALE/FEMALE. THE AGE, EDUCATIONYEARS, AND MMSE SCORES ARE PRESENTED AS MEAN \pm STANDARDDEVIATION (STD)

| Datasets | Category | Gender | Age | Education | MMSE |
|----------|----------|---------|--------------|----------------|----------------|
| ADNI-1 | NC | 127/102 | 75.8 ± 5.0 | 16.0 ± 2.9 | 29.1 ± 1.0 |
| | sMCI | 151/75 | 74.9 ± 7.6 | 15.6 ± 3.2 | 27.3 ± 1.8 |
| | pMCI | 102/65 | 74.8 ± 6.8 | 15.7 ± 2.8 | 26.6 ± 1.7 |
| | AD | 106/93 | 75.3 ± 7.5 | 14.7 ± 3.1 | 23.3 ± 2.0 |
| ADNI-2 | NC | 113/87 | 74.8 ± 6.8 | 15.7 ± 2.8 | 26.6 ± 1.7 |
| | sMCI | 134/105 | 71.7 ± 7.6 | 16.2 ± 2.7 | 28.3 ± 1.6 |
| | pMCI | 24/14 | 71.3 ± 7.3 | 16.2 ± 2.7 | 27.0 ± 1.7 |
| | AD | 91/68 | 74.2 ± 8.0 | 15.9 ± 2.6 | 23.2 ± 2.2 |
| AIBL | NC | 192/255 | 72.8 ± 6.6 | _ | 28.7 ± 1.2 |
| | AD | 30/42 | 71.2 ± 6.4 | - | 20.5 ± 5.7 |

ADNI-1: The *baseline* ADNI-1 dataset consists of 1.5T T1-weighted sMRI scans acquired from totally 821 subjects. These subjects were divided into three categories (i.e., NC, MCI, and AD) in terms of the standard clinical criteria, including mini-mental state examination (MMSE) scores and clinical dementia rating. According to whether MCI subjects would convert to AD within 36 months after the baseline evaluation, the MCI subjects were further specified as stable MCI (sMCI) subjects that were always diagnosed as MCI at all time points (0–96 months), or progressive MCI (pMCI) subjects that finally converted to AD within 36 months after the baseline. To summarize, the baseline ADNI-1 dataset contains 229 NC, 226 sMCI, 167 pMCI, and 199 AD subjects.

ADNI-2: The *baseline* ADNI-2 dataset includes 3T T1weighted sMRI data acquired from 637 subjects. According to the same clinical criteria as those used for ADNI-1, these 637 subjects were further categorized as 201 NC, 239 sMCI, 38 pMCI, and 159 AD subjects. It is worth noting that to make sure ADNI-2 is independent of ADNI-1, subjects that appear in both ADNI-1 and ADNI-2 were removed from ADNI-2. Four demographic factors, including gender, age, education years, and MMSE scores, are available for all subjects in both ADNI-1 and ADNI-2.

AIBL: The *baseline* AIBL dataset consists of 1.5T or 3T T1weighted sMRI scans acquired from 519 subjects, where 72 subjects were diagnosed as AD and the remaining 447 subjects are NCs. Three demographic factors, including gender, age, and MMSE scores, are available for all subjects in AIBL.

B. Image Preprocessing

We preprocessed all sMRI scans using a standard pipeline. Specifically, the anterior commissure (AC)–posterior commissure (PC) correction was performed first using the MIPAV software.¹ Then, the intensity correction of the sMRIs was performed using the N3 algorithm [39]. The brain skull and dura were stripped by the BET method [40] in the FSL package [41], and the cerebellum was further removed by warping a labeled template to each skull-stripped image. The FLIRT method [42] in the FSL package was used to linearly align

¹http://mipav.cit.nih.gov/index.php



Fig. 2. Architecture of the backbone FCN for weakly supervised discriminative localization on 3-D sMRI (size: $144 \times 184 \times 152$), where Conv, GAP, and FC stand for convolutional, GAP, and fully connected layers, respectively. The number of channels (e.g., 16), kernel size (e.g., $3 \times 3 \times 3$), and stride (e.g., 1) in each Conv layer is shown as "16@ $3 \times 3 \times 3$ -1."

all sMRIs to the Colin27 template [43] to remove global linear difference and also to resample all images for having an identical spatial resolution (i.e., $1 \times 1 \times 1 \text{ mm}^3$). Finally, all linearly aligned sMRIs were cropped to have the identical size of $144 \times 184 \times 152$, during which only uninformative backgrounds were removed and all the brain structures were completely preserved without loss of any potentially useful information.

IV. METHOD

As shown in Fig. 1, we propose an attention-guided deep-learning framework to capture multilevel discriminative knowledge from the whole-brain sMRI for automated brain disease diagnosis. We first design an FCN as the backbone to generate DAM for each input brain image, using directly the subject-level class label (e.g., AD/NC) as weakly supervised localization guidance. Each DAM has the same spatial resolution as the input sMRI scan, and each element represents the discriminating power of the corresponding voxel for disease diagnosis. Using the DAMs as guidance, a multibranch HybNet is constructed on the input sMRIs and the intermediate feature maps of the backbone FCN to fuse multilevel sMRI features for AD classification or MCI conversion prediction.

A. Generation of Disease Attention Map

In the first stage, a backbone FCN is trained for weakly supervised discriminative localization, using the whole-brain sMRI (size: $144 \times 184 \times 152$) as the input and the subject's class label (e.g., AD/NC) as the ground truth. Fig. 2 shows the architecture of our FCN, which consists of 12 convolutional (Conv) layers, a global average pooling (GAP) layer, and a classification (i.e., fully connected, FC) layer. All Conv layers use $3 \times 3 \times 3$ kernels (e.g., Conv1 and Conv2), $2 \times 2 \times 2$ kernels (e.g., Conv3), or $1 \times 1 \times 1$ kernels (i.e., Conv12) with zero padding, followed by batch normalization (BN) and rectified linear unit (ReLU) activation. The stride for the $3 \times 3 \times 3$ and $1 \times 1 \times 1$ kernels is set as 1, while the stride for the $2 \times 2 \times 2$ kernels (i.e., Conv3, Conv6, and Conv9) is set as 2 to halve the size of output feature maps. The last Conv layer with $1 \times 1 \times 1$ kernels (i.e., Conv12) is included to squeeze the number of channels. The numbers of channels for Conv1 to Conv11 are 16, 16, 16, 32, 32, 32, 64, 64, 64, 128, 128, and 64, respectively. On the feature maps yielded by Conv12, a GAP operation is performed to produce a feature vector, which is further used in the last FC layer (with softmax activation) to

predict the probability score of the input sMRI belonging to a specific category. Notably, the bias is removed from the FC layer for task-oriented localization after network training.

Inspired by [44], we backpropagate the learned weights of the classification layer onto the top Conv feature maps to generate a spatial DAM for respective sMRI. Such DAMs highlight discriminative brain regions that are highly associated with the diagnostic task. Specifically, let the feature maps yielded by Conv12 be { $\mathbf{F}_1, \ldots, \mathbf{F}_M$ }, where each \mathbf{F}_m $(m = 1, \ldots, M)$ has the size of $(X/8) \times (Y/8) \times (Z/8)$ with $X \times Y \times Z$ (i.e., 144 × 184 × 152) be the size of the input sMRI, and M = 64 is the number of channels. Based on the FC weights (i.e., $[\mathbf{w}_1^c, \ldots, \mathbf{w}_M^c]^T$) learned for the *c*th class (e.g., AD), the corresponding DAM (i.e., \mathbf{A}^c) is defined as

$$\mathbf{A}^{c}(x, y, z) = \sum_{m=1}^{M} \mathbf{w}_{m}^{c} \mathbf{F}_{m}(x, y, z).$$
(1)

Determined by the GAP operation and the FC operation without bias, the DAM correlates strongly with the diagnostic task, considering that the classification score s^c (before softmax and normalization) has the form of

$$\mathbf{s}^{c} = \sum_{m=1}^{M} \mathbf{w}_{m}^{c} \sum_{x,y,z} \mathbf{F}_{m}(x,y,z) = \sum_{x,y,z} \mathbf{A}^{c}(x,y,z).$$
(2)

Finally, the DAMs quantified by (1) for different classes are aggregated as

$$\mathbf{A}(x, y, z) = \sum_{c} \mathbf{s}^{c} \mathbf{A}^{c}(x, y, z)$$
(3)

which is simply upsampled via linear interpolation to obtain voxelwise attention map for the whole-brain sMRI.

B. Attention-Guided Hybrid Network

The DAM produced in the first stage can provide critical guidance for discriminative localization of subject specific global brain atrophies induced by dementia, especially considering that it is practically challenging to directly capture local and subtle structural abnormalities from a whole-brain sMRI. On the other hand, by aggregating the DAMs for considerable samples in the training set, it also implies a potentially effective way to identify intersubject-consistent discriminative atrophy locations across different sMRIs. We assume that, since the discriminative knowledge extracted from subject specific and intersubject-consistent DAMs are oriented by the unique diagnostic task, these two kinds of attention maps should be supplements/enhancements to each other. Based on this assumption, in the second stage, we propose an attentionguided HybNet to integrate multilevel discriminative sMRI features for AD diagnosis. The architecture of our HybNet consists of a global branch (GB), a local branch (LB), and a fusion branch, as shown in Fig. 3.

1) Global Branch: The GB attempts to capture subjectspecific discriminative information at the global whole-image level. To this end, the intermediate feature maps of the FCN (i.e., Conv11 in Fig. 2) are reused, considering that they already encode the raw global information automatically extracted from the whole-brain sMRI for the diagnostic task.

LIAN et al.: ATTENTION-GUIDED HYBNET FOR DEMENTIA DIAGNOSIS WITH STRUCTURAL MR IMAGES



Fig. 3. Architecture of our proposed attention-guided HybNet to learn multilevel discriminative sMRI features for dementia diagnosis. Our HybNet consists of a GB, an LB, and a fusion branch. The GB applies a lightweight CNN on the backbone feature map of each subject, spatially weighted by the corresponding DAM, to capturing global and individual-specific discriminative information. In the LB, using local maxima on the mean DAM of the training set as location proposals, an H-FCN [22] is applied on a set of image patches extracted from the input sMRI to capturing local and intersubject-consistent discriminative information. In H-FCN, multiple PSNs share the same network parameters across different patch locations. To further control the number of learnable parameters, the GAP is used in the local and GBes for capturing 1-D feature representations. Finally, the high-level feature representations produced by the local and global branches are further combined in the fusion branch to yield higher level features for brain disease diagnosis.

As the input for the GB, these image-level feature maps are spatially weighted by the corresponding DAM (via voxelwise multiplication) across channels to amplify the influence of input features for potentially informative brain regions. After the voxelwise local enhancement, they are further processed by two additional Conv layers [i.e., $Conv_G1$ and $Conv_G2$ in Fig. 3] to learn more discriminative feature maps. Both $Conv_G1$ and $Conv_G2$ adopt $3 \times 3 \times 3$ kernels, followed by BN and ReLU, with the numbers of channels set as 64 and 32, respectively. Finally, a GAP layer is attached on the $Conv_G2$ layer to produce the global whole-image-level representations encoding subject-specific discriminative information.

2) Local Branch: Determined by the input (i.e., the spatially weighted Conv11 feature maps of the backbone FCN), the GB in our HybNet can only learn global whole-image-level feature representations with limited local description. Also, as the DAMs are quantified independently for each input brain image, the subject-specific discriminative information captured by the proposed GB might be unstable for disease diagnosis [see Fig. 4(b)]. To tackle these limitations, our HybNet includes another branch (i.e., the LB) to extract local discriminative information at the patch level, considering that the early stage of AD may first cause subtle structural changes at very local brain regions. This LB is built on informative patch locations that are spatially consistent across different sMRIs, and thus should be a stable complement to the subject-specific GB.

As shown in Fig. 3, the structure of the LB in our HybNet is in line with H-FCN [22], which is actually a hierarchical network sequentially merging multiscale sMRI features provided by different patches and regions for the diagnostic task. The input for the LB is multiple 3-D image patches (size: $25 \times 25 \times 25$) extracted from the linearly aligned sMRI scan. Each patch location corresponds to a subnetwork (i.e., the right part in Fig. 3), and all patch-level subnetworks (PSNs) share the same architecture and weights to limit the number of learnable parameters. Spatially neighboring patches are grouped as a specific brain region (or second-level patches), and the outputs of the corresponding PSNs (size: $1 \times 1 \times 1 \times 64$) are concatenated according to their spatial relationship to form the input (e.g., $2 \times 2 \times 2 \times 64$ tensor) for the construction of the subsequent region-level subnetwork. Each region-level subnetwork consists of a specific Conv layer with 64 channels (i.e., $Conv_{R}$ -64 in Fig. 3) and a subsequent GAP layer to produce the region-level feature representation (size: $1 \times 1 \times 1 \times 64$). All region-level subnetworks are concatenated in a channelwise fashion, followed by two Conv layers (i.e., Conv_S1-64 and Conv_S2-32 in Fig. 3) with 64 and 32 channels, respectively. A GAP layer is further employed to produce the final feature representation in the LB, encoding intersubject consistent discriminative information from local patches and regions.

Compared with H-FCN [22], the most critical difference of the LB in our HybNet method lies in the definition of the initial inputs (i.e., location proposals). Instead of using local patches that are widely distributed over the whole-brain image as network input, in this study, we select location proposals in a more efficient data-driven way. That is, based on the mean DAM of the training samples in the linearly aligned image space, we first select voxels with the mean DAM values higher than a predefined threshold (set as 0.3 in our implementation) as potential patch locations. Then, we further require that the distance between any two feasible patch locations should not be less than 25 voxels, and use the qualified candidates as the central points to extract image patches (size: $25 \times 25 \times 25$). Considering that the DAM is spatially sparse and strongly correlates with the diagnostic task, the LB in our HybNet requires less location proposals than H-FCN. Also, since the mean DAM defined on the training set [e.g., Fig. 4(c)] is relatively more robust than subject-specific DAMs [e.g., Fig. 4(b)], the corresponding patch locations defined on the mean DAM [e.g.,



Fig. 4. 2-D illustration of a (a) testing subject in ADNI-2 and (b) corresponding subject-specific DAM generated by the backbone FCN trained on ADNI-1. Besides, (c) and (d) present the mean DAM generated on training subjects in ADNI-1 and the resulting intersubject-consistent patch locations, respectively.

Fig. 4(d)] not only are intersubject consistent but also could be more robust to imaging noise.

3) Fusion Branch: Finally, the subject-specific global discriminative features (i.e., the output of the GB) are concatenated with the intersubject-consistent local discriminative features (i.e., the output of the LB). They are further fused via two subsequent FC layers (i.e., FC_F1 and FC_F2 with 64 and 32 units, respectively) followed by ReLU activations to learn a holistic multilevel feature representation with higher discriminative power, based on which the diagnosis of the respective subject is performed by the ultimate softmax classification layer.

C. Implementation Details

The proposed attention-guided deep-learning diagnostic model was implemented in 3-D on a single GPU (i.e., NVIDIA GTX TITAN 12 GB), using Python based on the Keras package.² In the *first* stage, the input of our backbone FCN is the linearly aligned whole-brain sMRI with the size of $144 \times 184 \times 152$. We trained the FCN by setting the minibatch size as 2 and applying 0.5 dropout to multiple Conv layers. The model was trained by the cross-entropy loss with the Adam optimizer. The training set was augmented online by the combination of 1) randomly rescaling the brain images in a very small range and 2) randomly flipping the brain images in the axial plane.

In the *second* stage, the input of our HybNet includes: 1) the intermediate FCN (i.e., Conv11) feature maps spatially enhanced by the respective DAM (size: $18 \times 23 \times 19 \times 128$) and 2) a set of local patches (size: $25 \times 25 \times 25$) identified from the whole-brain image. As the initial patch locations were determined by a data-driven strategy (i.e., based on the mean DAM for all training samples), the size of the input patch set depends on the training data. For example, using ADNI-1 to train the models for AD diagnosis and MCI conversion prediction, the number of location proposals (i.e., initial patches) used in the LB of our HybNet was 36 and 33, respectively. The HybNet was trained with a deep-supervision strategy. The global and local branches of HybNet were first trained independently. That is, the GB was trained with the cross-entropy loss, using the Adam optimizer. And then the LB was trained in a similar way as the implementation of H-FCN in [22], including an initial training step and a network pruning step. After that, both the local and global branches were kept frozen to further update the fusion branch, using the cross-entropy loss and Adam optimizer. During training, the mini-batch size was set as 5, and 0.5 dropout was activated for multiple layers. The training samples were augmented online by the combination of 1) adding the Gaussian noise to the input feature maps and 2) randomly shifting at each patch location within a small neighborhood to extract input image patches.

Two different tasks were studied in this article, including AD diagnosis (i.e., AD versus NC classification) and MCI conversion prediction (i.e., pMCI versus sMCI classification). We first trained our proposed model for AD diagnosis, and then transferred the learned parameters to initialize the training of our proposed model for MCI conversion prediction, mainly considering that these two tasks are highly correlated and the latter task is relatively more challenging.

V. EXPERIMENTS

A. Experimental Settings

To evaluate the generalization capability of different methods, in the experiments, we train classification models on ADNI-1, and evaluate them on the two independent datasets, including ADNI-2 and AIBL. The diagnostic performance was quantitatively evaluated in terms of four criteria, that is, accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

We first compare our proposed HybNet method with two baseline methods, including the region-based method (ROI) [5] and the VBM method [28]. Besides, we further compare HybNet with two state-of-the-art deep-learning methods, including the deep multi-instance learning (DMIL) method [18] and the hierarchical FCN (H-FCN) method [22]. The details of these four competing methods are introduced as follows.

1) Region-Based Method: The ROI method used regionlevel handcrafted features to construct SVM classifiers. Specifically, in line with [5], we first applied the FAST algorithm [45] in the FSL package³ to segment the whole-brain sMRI into GM, white matter (WM), and cerebrospinal fluid (CSF). Using the HAMMER algorithm [46], we then warped the anatomical automatic labeling (AAL) atlas onto each subject to define 90 ROIs. Finally, we quantified the normalized GM volume in each ROI and concatenated them as the features to train linear SVM classifiers based on the training data.

2) Voxel-Based Morphometry: The VBM method used voxel-level handcrafted features quantified from the wholebrain sMRI to construct SVM classifiers. Following [28], we spatially normalized each sMRI scan onto the Colin27 template [43] to extract voxelwise GM density as features. After that, the dimensionality of these voxelwise features was reduced by performing statistical group comparison. Finally,

7

TABLE II

RESULTS OF AD VERSUS NC CLASSIFICATION ON ADNI-2 AND AIBL, RESPECTIVELY, OBTAINED BY THE MODELS TRAINED ON ADNI-1

| Mathad | ADNI-2 | | | AIBL | | | | |
|---------------|--------|-------|-------|-------|-------|-------|-------|-------|
| Methou | ACC | SEN | SPE | AUC | ACC | SEN | SPE | AUC |
| ROI | 0.792 | 0.786 | 0.796 | 0.867 | 0.859 | 0.847 | 0.861 | 0.896 |
| VBM | 0.805 | 0.774 | 0.830 | 0.876 | 0.852 | 0.861 | 0.850 | 0.901 |
| DMIL | 0.911 | 0.881 | 0.935 | 0.959 | 0.900 | 0.902 | 0.899 | 0.943 |
| H-FCN | 0.903 | 0.824 | 0.965 | 0.951 | 0.884 | 0.873 | 0.886 | 0.939 |
| HybNet (Ours) | 0.919 | 0.887 | 0.945 | 0.965 | 0.898 | 0.873 | 0.902 | 0.946 |

we constructed linear SVM classifiers using the selected features.

3) Deep Multi-Instance Learning: The DMIL method constructed a CNN-based multi-instance model on a set of local image patches. In line with [18], we selected the top 40 ADrelated anatomical landmarks from the landmark pool [11]. Using these landmarks as the central points, we extracted $24 \times 24 \times 24$ patches, and each of them was processed by a CNN to obtain a set of patch-level feature representations. Finally, we concatenated all patch-level features, and attached FC layers on them to train softmax classifiers.

4) Hierarchical Fully Convolutional Network: The H-FCN method automatically identified multiscale discriminative locations to construct a hierarchical model for AD diagnosis. Specifically, nonlinear registration was performed first to establish anatomical correspondence across linearly aligned subjects for defining 120 location proposals widely distributed over the whole brain (i.e., to cover all potentially discriminative locations). Then, a set of $25 \times 25 \times 25$ patches was extracted as the input to train an initial H-FCN. We further refined the trained network by pruning the subnetworks to directly remove those uninformative patches and regions.

It is worth noting that both the two state-of-the-art deeplearning methods (i.e., DMIL and H-FCN) require much more learnable parameters than our HybNet method. Specifically, DMIL contains more than 37 million learnable parameters, since the 40 different patch locations have their own specific subnetworks. H-FCN shares the PSNs across 120 location proposals to control network complexity, resulting in nearly 3 883 930 learnable parameters for the initial network. Compared with H-FCN, our HybNet requires much fewer location proposals for the LB (i.e., 36 for AD diagnosis and 33 for MCI conversion prediction), due to which the whole number of learnable parameters is further reduced to be less than 1 750 000 for the initial network (including the local, global, and fusion branches).

B. Results of AD Diagnosis

In this group of experiments, we compare our HybNet method with four competing methods in the task of AD versus NC classification. We trained the models on ADNI-1 (in which 10% subjects were randomly split for validation), and then applied them to diagnosing the subjects from ADNI-2 and AIBL, respectively. The quantified classification results in terms of four different metrics (i.e., ACC, SEN, SPE, and AUC) are summarized in Table II.

TABLE III Results of PMCI Versus sMCI Classification on ADNI-2, Obtained by the Models Trained on ADNI-1

| Method | ACC | SEN | SPE | AUC |
|---------------|-------|-------|-------|-------|
| ROI | 0.661 | 0.474 | 0.690 | 0.638 |
| VBM | 0.643 | 0.368 | 0.686 | 0.593 |
| DMIL | 0.769 | 0.421 | 0.824 | 0.776 |
| H-FCN | 0.809 | 0.526 | 0.854 | 0.781 |
| HybNet (Ours) | 0.827 | 0.579 | 0.866 | 0.793 |

From Table II, we have at least the following three observations. First, compared with the conventional machinelearning approaches (i.e., ROI and VBM), the deep-learning approaches (i.e., DMIL, H-FCN, and our HybNet) largely improved the diagnostic performance on both ADNI-2 and AIBL datasets, which demonstrates the significance of learning high-level sMRI features for AD diagnosis. Second, compared with DMIL that adopted predetermined informative locations to construct deep networks, the other two deep-learning approaches (i.e., H-FCN and HybNet) based on task-oriented discriminative localization led to competitive or even better diagnostic performance. For example, our proposed HybNet outperformed DMIL in terms of all metrics on the ADNI-2 dataset. It implies that integrating discriminative localization, feature extraction, and classifier construction into a unified deep-learning framework is feasible and beneficial for automated AD diagnosis. Third, compared with H-FCN, our proposed HybNet method yielded better AD classification results on both two datasets. For example, on the ADNI-2 dataset, the ACC and AUC values were improved from 0.903 to 0.919 and 0.951 to 0.965, respectively. It indicates that the combination of multilevel (i.e., subject specific and intersubject consistent, and global and local) discriminative sMRI information in our HybNet method is beneficial for the diagnostic task, compared with identifying solely local image patches that are anatomically consistent across different subjects H-FCN.

C. Results of MCI Conversion Prediction

In this group of experiments, our HybNet method as well as the other four competing methods were evaluated on pMCI versus sMCI classification. The models trained on ADNI-1 were applied to classifying the subjects from ADNI-2, with the classification results presented in Table III. From Table III, we can draw a similar conclusion as that in Section V-B. That is, the DMIL, H-FCN, and our HybNet methods still





Fig. 5. Results of (a) AD diagnosis and (b) MCI conversion prediction on the ADNI-2 dataset, obtained by the first-stage FCN, the global branch (i.e., GB) of HybNet, the local branch (i.e., LB) of HybNet, and the complete HybNet trained on the ADNI-1 dataset, respectively.



Fig. 6. ROC curves for (a) AD diagnosis and (b) MCI conversion prediction on the ADNI-2 dataset, obtained by the first-stage FCN, the global branch (i.e., GB) of HybNet, the local branch (i.e., LB) of HybNet, and the complete HybNet trained on the ADNI-1 dataset, respectively.

largely outperformed the ROI and VBM methods, suggesting that automatically learning high-level sMRI features in a task-oriented deep-learning framework is also beneficial for MCI conversion prediction. On the other hand, refer to the results presented in both Tables II and III, we can observe that the two deep-learning methods capitalizing on automated discriminative localization (i.e., H-FCN and our HybNet) led to more significant improvement of classification performance in the challenging task of MCI conversion prediction, compared with DMIL using predetermined discriminative locations. For example, by comparing DMIL with our HybNet method, the ACC was increased from 0.769 to 0.827 and the AUC was increased from 0.776 to 0.793. It potentially implies that task-oriented detection of discriminative atrophy locations is relatively more important in developing CAD methods for the early stage of AD. Besides, from Table III, we can observe that HybNet still yielded better classification performance than H-FCN, which is in consistent with the results presented in Table III. It further verifies the effectiveness of learning multilevel discriminative sMRI features in the more challenging task of MCI conversion prediction.

D. Ablation Study

In this group of experiments, we performed an ablation study to evaluate the effectiveness of several essential components of our proposed CAD model, including: 1) the weakly supervised attention mechanism; 2) the GB in our HybNet to capture global and subject-specific discriminative information; 3) the LB in our HybNet to capture local and intersubject-consistent discriminative information; and 4) the fusion of multilevel discriminative information in our HybNet. Specifically, on AD diagnosis as well as MCI conversion prediction, we applied our FCN, the GB of our HybNet, the LB of our HybNet, and our complete HybNet trained on ADNI-1 to classifying the subjects from ADNI-2. The classification results in terms of ACC, SEN, SPE, and AUC are summarized in Fig. 5, and the corresponding ROC curves are shown in Fig. 6.

Based on Figs. 5 and 6, we can have the following observations. First, our HybNet method, as well as its components (i.e., GB and LB), consistently outperformed our FCN model. For example, compared with FCN, GB and LB improved the ACC from 0.704 to 0.747 and 0.780, and also improved the AUC from 0.740 to 0.758 and 0.749, respectively. It implies that: 1) the DAMs produced by FCN in the first stage of the proposed CAD model could provide effective guidance for discriminative localization of disease-related brain regions, as it is actually challenging in practice to train a classifier using directly the whole-brain sMRI without any guidance (such as in our FCN) and 2) the global and subject-specific sMRI features learned by GB, as well as the local and intersubjectconsistent sMRI features learned by LB is discriminative for the diagnostic tasks. Second, the complete HybNet led to better performance than GB and LB in both two tasks, for example, the classification results in terms of all criteria were improved due to the fusion of multilevel sMRI features. It indicates that the discriminative information provided by GB and LB are complementary to each other, and the fusion of them to learn holistic multilevel sMRI features are desired for the diagnostic tasks.

E. Identified Disease Attention Maps

Our proposed method can automatically identify discriminative brain regions from the whole-brain sMRIs by learning task-oriented DAMs (i.e., Section IV-A). In Figs. 7 and 8, we present some example DAMs (for the test subjects in ADNI-2) generated by the FCN models constructed in the tasks of AD versus NC classification and pMCI versus sMCI classification, respectively, where each DAM is shown in 2-D from two different views.

From Figs. 7 and 8, we can have the following observations. *First*, our proposed method consistently highlighted multiple parts at the locations of the hippocampus, frontal lobe, fusiform gyrus, amygdala, and ventricle for different subjects with AD or MCI. It suggests that the automated-localization results yielded by our proposed methods are reliable, considering that the discriminative power of these brain regions for dementia diagnosis has already been validated in previous studies [1], [2], [5], [7]. *Second*, while the discriminative brain

LIAN et al.: ATTENTION-GUIDED HYBNET FOR DEMENTIA DIAGNOSIS WITH STRUCTURAL MR IMAGES



Fig. 7. Illustration of the DAMs for six test AD subjects in ADNI-2, which were generated by the weakly supervised FCN model trained on ADNI-1 in the task of AD diagnosis.



Fig. 8. Illustration of the DAMs for six test MCI subjects in ADNI-2, which were generated by the weakly supervised FCN model trained on ADNI-1 in the task of MCI conversion prediction.



Fig. 9. 2-D illustration of (a) two different AD subjects from ADNI-2, (b) respective AD-related anatomical landmarks identified by [11], and (c) respective DAMs produced by our proposed method.

regions (e.g., in Fig. 7) detected by our method are consistent, they are not totally the same across different subjects, for example, the same brain region may have been emphasized differently. It implies that our proposed method is feasible for individualized localization of brain atrophies associated with AD/MCI, which should be a valuable property in practice. *Third*, by comparing Fig. 7 with Fig. 8, we can see that the brain regions identified for AD diagnosis and MCI conversion prediction are partially different, although they are largely consistent, for example, the ventricular regions were highlighted for some MCI subjects in Fig. 8. It is worth noting that previous studies [1], [47] have shown that ventricular

enlargement is an important biomarker for analyzing the progression of AD, which in some sense implies that our proposed method is effective in task-oriented discriminative localization to predict the disease progression for at risk AD subjects.

9

We also compare the DAMs generated by our method with the top-ranked (120) AD-related anatomical landmarks defined by [11]. The corresponding 2-D illustrations for two different ADNI-2 subjects are shown in Fig. 9. From Fig. 9, we can see that the highlighted regions in our generated DAMs are largely consistent with those regions covered by anatomical landmarks defined in [11], which further verifies the effectiveness of our method for task-oriented discriminative localization in sMRIbased dementia diagnosis.

VI. DISCUSSION

A. Evaluation on Other Brain Diseases With sMRI

Apart from AD, the proposed method may also be used for sMRI-based diagnosis of other brain diseases, for example, Parkinson's disease (PD) and autism spectrum disorder (ASD). To verify the feasibility, we conducted several preliminary experiments on two challenging public datasets, that is, the Parkinson's Progression Markers Initiative (PPMI) [48] and National Database for Autism Research (NDAR) [49], for PD and ASD diagnosis, respectively. Briefly, the PPMI dataset consists of the T1-weighted sMR images acquired from 374



Fig. 10. 2-D illustration of the DAMs produced by our proposed method on (a) two different PD subjects from PPMI and (b) two different ASD subjects from NDAR, respectively.

PD and 169 NC subjects, while the NDAR dataset consists of the T1-weighted sMR images acquired from 61 Autism/ASD and 215 NC subjects. The detailed demographic information of these two datasets can be found in Section 1 (including Tables S1 and S2) of the supplementary material. For each diagnostic task, we randomly selected 10% subjects as the test set and used the remaining 80% and 10% subjects as the training and validation sets, respectively. In addition, considering the small sizes of both PPMI and NDAR, we transferred the network parameters learned on ADNI (for AD diagnosis) to initialize the training of the networks for PD and ASD diagnosis. The corresponding results on both tasks are presented in Table S3 of the supplementary material, suggesting that our method is potentially feasible for other brain diseases.

In Fig. 10, we present some example DAMs generated by our method in the tasks of PD versus NC classification [i.e., Fig. 10(a)] and ASD versus NC classification [i.e., Fig. 10(b)]. From Fig. 10, we can have the following observations. First, the DAMs demonstrate large interdisease differences and intradisease consistency, reliably uncovering the fact that different brain diseases actually cause abnormal changes at different brain locations. Second, the DAMs for the same brain disease consistently highlighted some biologically meaningful brain regions. For example, in the task of PD diagnosis [i.e., Fig. 10(a)], our method robustly highlighted multiple parts at the locations of the brain stem and middle cingulate gyrus. It further suggests the reliability of our method in automated discriminative localization, considering that the predictive value of those brain regions in PD diagnosis has been validated by previous studies [50], [51].

B. Limitations and Future Work

While our proposed method performs well in automated discriminative localization and disease diagnosis, several limitations should be carefully addressed in the future to further improve its performance and practical value. *First*, our current model was implemented to perform discriminative localization and disease diagnosis independently, which means the diagnostic results in the second stage could not be used to help refine the discriminative localization in the first stage. It should be an interesting and promising direction to integrate these two stages into a purely end-to-end framework, so that the modules designed for both tasks could provide complementary guidance to each other. *Second*, discriminative

locations detected by our current model are relatively coarse, mainly due to the reason that the DAMs were defined on high-semantic (but low-resolution) features maps from the top convolutional layers of the backbone FCN. To generate higherresolution DAMs without loss of semantic information, we could potentially extend our current model by using the topdown strategy to combine high-semantic information (from top layers) with local details (from bottom layers) for weakly supervised discriminative localization at multiple scales. Third, the intersubject-consistent patch locations used in the LB of our HybNet model were relatively coarse, as they were determined by averaging the DAMs for training samples directly in the linearly aligned image space. To define patch locations more precisely, similar to [11] and [22], we could alternatively identify the voxel-to-voxel anatomical correspondence across all subjects via nonlinear registration [52] and then calculate the mean DAM in a template space, which will be one of our future work.

VII. CONCLUSION

In this work, an attention-guided deep-learning model was proposed to automatically identify the AD-related abnormal locations from the whole-brain sMRI, and then extract multilevel discriminative sMRI features at these locations to construct the classifiers for AD diagnosis. The effectiveness of our proposed method has been evaluated on three public datasets consisting of 1976 subjects. The experimental results have demonstrated the superior performance of our method compared with several state-of-the-art methods in both the tasks of AD diagnosis and MCI conversion prediction.

ACKNOWLEDGMENT

Data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The investigators within the ADNI did not participate in analysis or writing of this study. A complete listing of ADNI investigators can be found online.

REFERENCES

- G. B. Frisoni, N. C. Fox, C. R. Jack, Jr, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nat. Rev. Neurol.*, vol. 6, no. 2, p. 67, 2010.
- [2] L. Wang *et al.*, "Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type," *IEEE Trans. Med. Imag.*, vol. 26, no. 4, pp. 462–470, Apr. 2007.
- [3] S. Klöppel et al., "Automatic classification of MR scans in Alzheimer's disease," Brain, vol. 131, no. 3, pp. 681–689, 2008.
- [4] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138–149, 2009.
- [5] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [6] J. Koikkalainen, J. Lötjönen, L. Thurfjell, D. Rueckert, G. Waldemar, and H. Soininen, "Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease," *NeuroImage*, vol. 56, no. 3, pp. 1134–1144, 2011.
- [7] P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, and D. L. Collins, "Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease," *NeuroImage*, vol. 59, no. 4, pp. 3736–3747, 2012.

- [8] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," *Med. Image Anal.*, vol. 18, no. 5, pp. 808–818, 2014.
- [9] M. Liu, D. Zhang, and D. Shen, "Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1463–1474, Jun. 2016.
- [10] L. Sørensen *et al.*, "Early detection of Alzheimer's disease using MRI hippocampal texture," *Human Brain Map.*, vol. 37, no. 3, pp. 1148–1161, 2016.
- [11] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.
- [12] T. Zhou, K.-H. Thung, M. Liu, and D. Shen, "Brain-wide genome-wide association study for Alzheimer's disease via joint projection learning and sparse regression model," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 165–175, Jan. 2019.
- [13] H.-I. Suk, S.-W. Lee, D. Shen, and ADNI, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, Apr. 2017.
- [14] S. Liu *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.
- [15] H. Li, M. Habes, and Y. Fan. (2017). Deep Ordinal Ranking for Multi-Category Diagnosis of Alzheimer's Disease Using Hippocampal MRI Data. [Online]. Available: https://arxiv.org/abs/1709.01599
- [16] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention*, 2018, pp. 455–463.
- [17] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline. (2018). 3D CNN-Based Classification Using sMRI and MD-DTI Images for Alzheimer's Disease Studies. [Online]. Available: https://arxiv.org/abs/1801.05968
- [18] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multiinstance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, 2018.
- [19] C. Lian *et al.*, "Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images," *Med. Image Anal.*, vol. 46, pp. 106–117, May 2018.
- [20] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.
- [21] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, pp. 295–308, Mar. 2018.
- [22] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.
- [23] M. Liu, J. Zhang, C. Lian, and D. Shen, "Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3381–3392, Jul. 2020.
- [24] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017.
- [25] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.
- [26] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 186–200, Jan. 2020.
- [27] P. Yang *et al.*, "Fused sparse network learning for longitudinal analysis of mild cognitive impairment," *IEEE Trans. Cybern.*, early access, Sep. 30, 2019, doi: 10.1109/TCYB.2019.2940526.
- [28] J. Ashburner and K. J. Friston, "Voxel-based morphometry: The methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
- [29] Z. Xue, D. Shen, and C. Davatzikos, "CLASSIC: Consistent longitudinal alignment and segmentation for serial image computing," *NeuroImage*, vol. 30, no. 2, pp. 388–399, 2006.

- [30] A. Demiriz, K. P. Bennett, and J. Shawe-Taylor, "Linear programming boosting via column generation," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 225–254, 2002.
- [31] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [32] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artif. Intell.*, vol. 201, pp. 81–105, Aug. 2013.
- [33] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Human Brain Map.*, vol. 40, no. 3, pp. 1001–1016, 2019.
- [34] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 46, pp. 26–34, May 2018.
- [35] W. Yan, H. Zhang, J. Sui, and D. Shen, "Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2018, pp. 249–257.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [37] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," J. Magn. Resonance Imag., vol. 27, no. 4, pp. 685–691, 2008.
- [38] K. A. Ellis et al., "The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," Int. Psychogeriatrics, vol. 21, no. 4, pp. 672–687, 2009.
- [39] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
- [40] S. M. Smith, "Fast robust automated brain extraction," Human Brain Map., vol. 17, no. 3, pp. 143–155, 2002.
- [41] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.
- [42] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, 2002.
- [43] C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *J. Comput. Assisted Tomography*, vol. 22, no. 2, pp. 324–333, 1998.
- [44] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.
- [45] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectationmaximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [46] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
- [47] S. M. Nestor *et al.*, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain*, vol. 131, no. 9, pp. 2443–2454, 2008.
- [48] K. Marek et al., "The Parkinson progression marker initiative (PPMI)," Progr. Ueurobiol., vol. 95, no. 4, pp. 629–635, 2011.
- [49] D. Hall, M. F. Huerta, M. J. McAuliffe, and G. K. Farber, "Sharing heterogeneous data: The national database for autism research," *Neuroinformatics*, vol. 10, no. 4, pp. 331–339, 2012.
- [50] H. Braak, K. Del Tredici, U. Rüb, R. A. De Vos, E. N. J. Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiol. Aging*, vol. 24, no. 2, pp. 197–211, 2003.
- [51] E. Adeli *et al.*, "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *NeuroImage*, vol. 141, pp. 206–219, Nov. 2016.
- [52] G. Wu, F. Qi, and D. Shen, "Learning-based deformable registration of MR brain images," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1145–1157, Sep. 2006.



Chunfeng Lian (Member, IEEE) received the B.S. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2010, and the Ph.D. degree in computer engineering from Univiersité de Technologie de Compiègne, CNRS, Heudiasyc (UMR 7253), Compiègne, France, in 2017.

His research interests include medical image analysis, pattern recognition, and machine learning.



Yongsheng Pan (Graduate Student Member, IEEE) received the B.E. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering.

His research areas include neuroimage analysis, computer vision, and machine learning.

Mingxia Liu (Senior Member, IEEE) received the B.S. and M.S. degrees from Shandong Normal University, Shandong, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015.

Her current research interests include machine learning, pattern recognition, and neuroimaging analysis.

Dinggang Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1995.

He has published more than 1100 peer-reviewed papers in the international journals and conference proceedings, with H-index 99. His research interests include medical image analysis, computer vision, and pattern recognition.

Dr. Shen served on the Board of Directors, the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, in 2012–2015, and

was the General Chair for MICCAI 2019. He is a fellow serves as an editorial board member for eight international journals. He is a Fellow of the American Institute for Medical and Biological Engineering and the International Association for Pattern Recognition.

