Multi-stream multi-scale deep convolutional networks for Alzheimer's disease detection using MR images

Chenjie Ge [a,*], Qixun Qu [a], Irene Yu-Hua Gu [a], Asgeir Store Jakola [b,c]

[a] *Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden*
[b] *Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*
[c] *Sahlgrenska University Hospital, Gothenburg, Sweden*

ARTICLE INFO

ABSTRACT

This paper addresses the issue of Alzheimer's disease (AD) detection from Magnetic Resonance Images (MRIs). Existing AD detection methods rely on global feature learning from the whole brain scans, while depending on the tissue types, AD related features in different tissue regions, e.g. grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), show different characteristics. In this paper, we propose a deep learning method for multi-scale feature learning based on segmented tissue areas. A novel deep 3D multi-scale convolutional network scheme is proposed to generate multi-resolution features for AD detection. The proposed scheme employs several parallel 3D multi-scale convolutional networks, each applying to individual tissue regions (GM, WM and CSF) followed by feature fusions. The proposed fusion is applied in two separate levels: the first level fusion is applied on different scales within the same tissue region, and the second level is on different tissue regions. To further reduce the dimensions of features and mitigate overfitting, a feature boosting and dimension reduction method, XGBoost, is utilized before the classification. The proposed deep learning scheme has been tested on a moderate open dataset of ADNI (1198 scans from 337 subjects), with excellent test performance on randomly partitioned datasets (best 99.67%, average 98.29%), and good test performance on subject-separated partitioned datasets (best 94.74%, average 89.51%). Comparisons with state-of-the-art methods are also included.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative brain disease which affects people in various ways. The disease gradually causes cognitive deterioration, and eventually failure to carry out activities of daily life (ADL). Patients with AD frequently also demonstrate behavioural and psychological problems, causing additional distress for patients and caregivers [7]. It has been reported that 26.6 million patients suffer from AD worldwide, among which 56% were at the early stage. It is estimated that the population of the AD patients will grow fourfold to 106.8 million [5]. As a result, accurate diagnosis and treatment of AD is of great importance. Clinical techniques for medical assessment of AD consist of physical and neurobiological exams, Mini-Mental State Examination (MMSE) and so on. Recently, Magnetic Resonance Imaging plays an integral part of diagnosis and differential diagnosis. Studies have also suggested that structural MRIs may play important roles in AD detection [25].

Some diagnostic systems were proposed for extracting features from MR images, followed by classifiers that distinguish different subjects, e.g., AD or normal control (NC) groups. Features from the entire brain scan were used for AD detection [4,22]. Meanwhile, voxel-wise features of segmented brain tissues (GM, WM and CSF) were considered useful for detection of Alzheimer's disease. Fan et al. [10] first segmented the whole brain into GM, WM and CSF, and then voxel-wise densities were calculated and used for classification. Lerch et al. [17] fitted segmented GM and WM surfaces using deformable models. AD classification was based on cortical thickness, measured by calculating distances between two corresponding points in the GM and WM, respectively. These two methods were conducted by using analytical methods with hand-crafted features (i.e. features were designed by human experts) from segmented brain MR images. Since characterization of AD features from MRIs remains a challenging task even to clinicians/medical personnel due to the lack of deep understanding on the pathological changes in the brain, deep learning methods can provide great potential that surpass the methods using hand-crafted features.

Although these methods are promising, significant challenges remain and pose limitations to the clinical usage. One is that most of these deep learning methods use an entire brain scan as the

input. It is observed that changes indicating AD features are different in different tissue regions (e.g. GM, WM and CSF). Another one is that AD features in different tissue regions are best characterized by different scales (or resolution). Using a single scale for features in all tissues is not the best strategy. Furthermore, it is noticed that applying multi-scale feature representation would lead to significant increase in the feature dimension, and hence possible "the curse of dimensionality" to the classifiers. Some feature dimension reduction strategy would be desirable to mitigate this problem.

Motivated by the above issues, we propose a novel deep learning scheme to detect ADs by using multi-stream multi-scale convolution networks followed by multi-level feature fusion, feature boosting and dimension reduction. Three-streams, consisting of GM, WM or CSF tissue regions, are fed separately as the inputs for 3D multi-scale convolutional neural networks (MSCNNs). This is followed by applying feature fusion, feature boosting and dimension reduction before the classifier is applied. The main contributions of the paper include:

- A 3D multi-scale deep convolutional network (3D MSCNN) architecture is proposed to generate multi-scale features with rich semantics and image details for AD detection.
- Multi-stream feature extraction from segmented tissue regions is proposed to generate complementary features for AD detection.
- A two-level feature fusion scheme is proposed, which includes fusion in the scale level and the tissue level.
- A feature boosting and dimension reduction method is proposed for post-processing, where a tree boosting method XG-Boost is used for feature dimension reduction.
- Extensive empirical analysis of the performance is conducted, where the proposed scheme is also compared with the state-of-the-art methods.

The reminder of the paper is organized as follows. Section 2 reviews the related work. Section 3 describes the proposed scheme in detail. Section 4 shows experimental results and performance evaluation. Finally, the conclusion is given in Section 5.

2. Related work

Much effort has been made to use hand-crafted features for AD classification, where feature extraction is based on the knowledge of human researchers. Yang et al. [28] studied potential AD-related MR image features based on independent component analysis. Support vector machine (SVM) was then used for classifying AD and NC subjects. Tong et al. [26] utilized the strategy of multiple instance learning to classify dementia, where features were extracted using bags of MRI voxel patches and graph mapping. Arvesen et al. [3] studied methods of dimensional reduction and variations in the learning task to analyze structural MRI data, where model of decision trees with principal component analysis-based dimensional reduction has achieved the best result for AD detection. Liu et al. [19] proposed to extract multi-view features using selected templates. Encoding features were then obtained by clustering subjects in each view space, followed by an ensemble of SVMs to classify the subjects.

The recent development of deep learning methods for AD detection has drawn significant attention since features are learned automatically. Brosche et al. [6] proposed to learn the manifold of brain images using a deep brief network model, where similar patterns in image groups were used to distinguish AD from NC subjects. Sarraf el al. [22] employed the convolutional neural network (CNN) architectures LeNet and GoogleNet to detect Alzheimer's disease using sMRI and fMRI brain scans. GoogleNet achieved good performance using imbalanced training data where the ratio of AD

and NC scans is 5:1. Bäckström el al. [4] proposed an efficient and simple 3D CNN architecture, where good result was achieved for AD detection on a dataset containing 340 subjects. Auto-encoders (AEs) were shown to be another effective method for learning unsupervised genetic features, followed by fine-tuned task-specific layers for final classification. Suk and Shen [24] used stacked AEs to extract features from MRI and PETs image regions and CSF biomarkers, a multi-kernel SVM was then used for the classification. Gupta et al. [11] extracted slice-wise feature of MR images using 2D CNN. Pre-trained sparse AE was proposed for further performance enhancement. Hosseini-Asl el al. [13] proposed to use a pre-trained 3D convolutional AE to learn generic features, followed by a 3D CNN for refined training. Good performance was achieved for AD/NC classification on a relatively small dataset.

3. Proposed methodology

3.1. Overview of the proposed scheme

The main ideas behind the proposed scheme are to characterize the AD features in a multi-stream multi-scale fashion.

Using multi-stream inputs: This is motivated from the observation that different tissue regions contain different characteristics for Alzheimer's disease thus contribute differently to AD detection. It would be desirable to handle them separately by networks with different settings.

Using multi-scale features in each stream: This is motivated by the observation that AD features for different tissue regions are best described by different scales, e.g. the changes in CSF and GM are very different for ADs. Further, retaining all resolution levels (or, from coarse to fine scales) of features enables one to obtain features both from low-level of volume images and from high-level semantic and structural change information in ADs.

Using multi-level feature fusion: This is motivated by the observation that features from different tissue regions contribute and complement each other for the detection of AD. Furthermore, different scale features contribute coarse-to-fine features for ADs. Hence, the proposed strategy is designed to fuse features in these two levels.

Using feature boosting and dimension reduction: This is motivated by the observation that adopting a multi-stream multi-scale network significantly increases the dimension of features, hence could lead to "the curse of dimensionality" in the classifier given a fixed medium-size dataset. Hence, we apply a feature boosting and dimension reduction method for mitigating this issue.

Based on the above ideas, we proposed a novel 3D multi-stream multi-scale deep convolutional network (3D MSCNN) scheme and architecture that extracts image-level and semantic-level features for AD detection from different types of tissue regions, as shown in the pipeline of Fig. 1. The proposed scheme consists of three streams of 3D MSCNNs, separately applied on the GM, WM and CSF tissue regions (though our experiments later show that the two-stream scheme on GM and CSF leads to the best tradeoff between the performance and the computation, see Section 4.3). This is then followed by a two-level feature fusion method on the extracted features in different scales and tissue regions, respectively. To improve the learned features, we apply a pre-training and refined-training strategy in our scheme for the feature learning. The first round of training is on the whole brain scans to obtain the initial feature description, and the second round of training is then applied on different tissue regions to obtain refined tissue-specific features (see Section 3.6 for details). After these steps, a feature boosting and dimension reduction method is applied that is designed to retain the important/principal features while reducing the dimension of features. Finally, a classification step is used
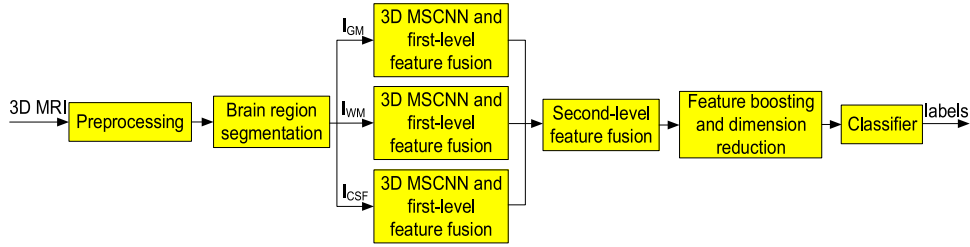
Fig. 1. The pipeline of the proposed Alzheimer's disease detection scheme.

for AD detection by classifying between AD and NC subjects. In the following, we describe these main contributions in details.

3.2. Multi-stream networks on segmented tissue regions

This section describes the individual streams of 3D MSCNN. Since the brain with Alzheimer's disease usually has shrinking gray matter and a larger ventricle than normal ones, it is reasonable to consider WM, GM and CSF as different feature sources since they contain different types of features related to the disease. To obtain richer features for AD detection, we propose to extract brain regions (GM, WM and CSF) as the multi-stream inputs to the deep neural networks in the multi-scale fashion. Let a original brain scan be $\mathbf{I}$, the three segmented brain regions are denoted as $\mathbf{I}_{GM}$, $\mathbf{I}_{WM}$, $\mathbf{I}_{CSF}$, then the segmentation will result in the partitioned regions, such that:

$$\mathbf{I} = \{\mathbf{I}_{GM} \ \mathbf{I}_{WM} \ \mathbf{I}_{CSF}\} \tag{1}$$

where $\mathbf{I}_{GM} \bigcup \mathbf{I}_{WM} \bigcup \mathbf{I}_{CSF} = \mathbf{I}$, $\mathbf{I}_{GM} \bigcap \mathbf{I}_{WM} \bigcap \mathbf{I}_{CSF} = \emptyset$. In the proposed method, the brain segmentation is implemented by an existing open software package FSL FAST [14]. FAST (FMRIB's Automated Segmentation Tool) is one of the softwares in FSL aimed at analyzing FMRI, MRI and DTI brain imaging data using algorithms based on hidden Markov random field model and Expectation-Maximization (EM) algorithm [29]. The output is three masks for GM, WM and CSF. Segmented brain regions are obtained by multiplying the masks with the original image. Example of segmented brain tissue regions is shown in Fig. 2.

3.3. Multi-scale network for individual tissue regions

In different tissue regions, the features are best characterized by different scales since individual tissue regions change differently in AD patients thus cannot be characterized by a single scale. Further, different scale features may describe both image-level and semantic-level features, which enrich the features for the detection of AD. We propose a multi-scale deep CNN network for individual tissue regions, as shown in Fig. 3. Multi-scale convolutional layers are used for extracting multi-scale features, where the input is the 3D brain MR images, followed by multi-scale fusion layers and fully-connected (FC) layers, whose output is the predicted class labels.

The proposed architecture of multi-scale deep convolutional layers consists of 8 layers (Conv1-Conv8) where the last 3 layers (Conv6-Conv8) are obtained from "skip connection" as shown in Fig. 3. The first 5 convolutional layers are designed by using conventional CNN structure, where feature maps are generated in a coarse-to-fine manner. We follow the design of VGG net [23] where multiple small kernels like 3*3 are stacked. This way increases the depth of the network and hence is better at learning more complex features. Since semantically strong features are often associated with relatively low resolution, while detailed image features are often related to fine resolution, it is desirable to obtain both type of features for AD classification. Therefore, skip
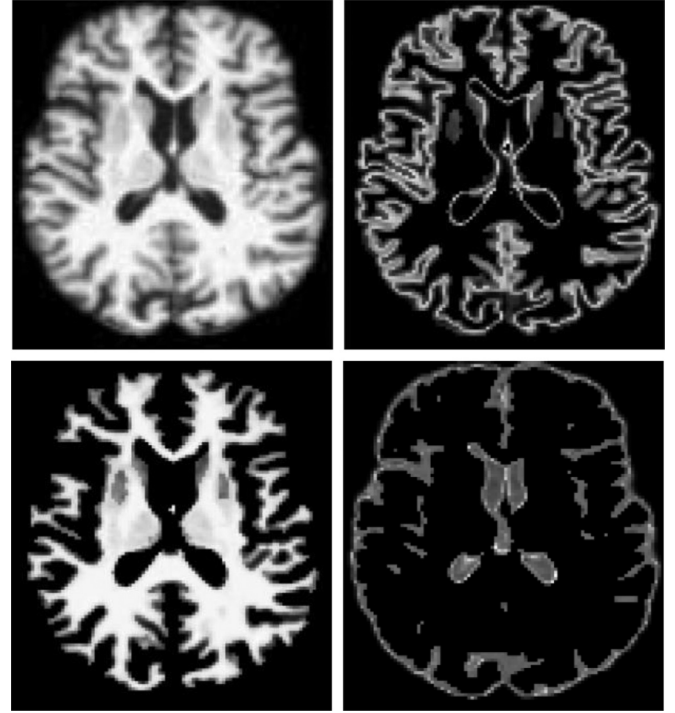

Fig. 2. Segmentation result of one sample image. First row from left to right: whole brain, GM. Second row from left to right: WM, CSF.

connection is utilized to combine features from both shallow and deep convolutional layers. More specifically, the outputs of convolutional layers 2,3,4 are added to the convolutional layers 7,6,5, respectively. The upsampling (by a factor 2) is applied to the convolutional layers 5,6,7, based on the nearest neighbourhood method. By using skip connection, semantically strong features are combined with high-resolution image-level features. Although features of more scales can be generated using the same manner, only four scales are used here because of the limitation of GPU memory (Titan XP 12 GB) when 3D scans are used as the input.

To obtain multi-resolution features from multi-scale deep convolutional layers, we apply max-pooling (7*6*6) to each individual layers to retain the multi-resolution structure, followed by individual fully-connected layers (i.e., 4 FC1's) to refine these feature vectors (after the flatten operation, see Fig. 3). The reason of applying separate pooling is that features from conv5 to conv8 layers have different resolution with image level and semantic level information, therefore global pooling as used in ResNet [12] is not suitable. After this, multi-scale feature vectors in the four FC1 layer boxes are fed to the first-level fusion (see description in the next section). Noting that the two FC layers, FC2 and FC3, are only used for the first-stage end-to-end training for individual tissue regions. The aim is to obtain features from each individual 3D MSCNN. These two FC layers are therefore not used in the final scheme (see Section 3.6 for details).
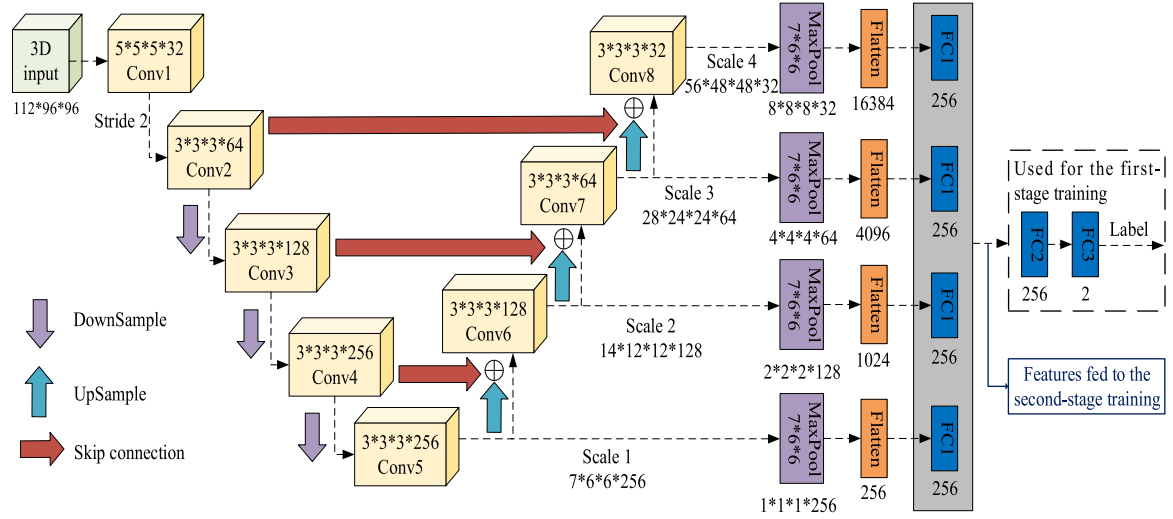
Fig. 3. The pipeline of the proposed 3D multi-scale deep convolutional neural network architecture.

Several existing networks might be looking similar to the proposed architecture, e.g., auto-encoder in [16], or the feature pyramid network in [18]. Therefore, it is worth mentioning their difference here. The proposed scheme differs from Le [16] by applying supervised training rather than unsupervised one. The proposed scheme differs from Lin et al. [18] by applying additional convolutional layers (3*3*3) before merging, resulting in more powerful feature learning. Furthermore, the proposed fusion and classification layers are specially designed for the dedicated task of AD classification.

3.4. Two-level feature fusion

This section describes the fusion method for feature from different scales as well as different tissue regions. Since different features are obtained from different scales and different tissue regions, feature fusion is conducted in two levels, one is on the scale level and another is on the tissue level.

The first-level feature fusion is performed on different scales of each individual tissue type. This is done by concatenating features from different scales, as shown in the gray box around FC1 layers in Fig. 2. Let $\mathbf{f}_j^{s_i}$, $i = 1, \ldots, 4$, be the features from four different scales for a given jth tissue region, $j \in \{GM, WM, CSF\}$, then the fusion may be described by concatenating the feature vectors (from 4 FC1's) as:

$$\mathbf{f}_j = [\mathbf{f}_j^{s_1} \quad \mathbf{f}_j^{s_2} \quad \mathbf{f}_j^{s_3} \quad \mathbf{f}_j^{s_4}] \tag{2}$$

The second-level fusion is performed on features from different tissue regions. This is done by concatenating the feature vectors $\mathbf{f}_j$, $j = GM, WM, CSF$, obtained from different streams of 3D MSCNNs. Let $\mathbf{f}$ be the feature vector after the second-level fusion, then it can be obtained as:

$$\mathbf{f} = [\mathbf{f}_{GM} \quad \mathbf{f}_{WM} \quad \mathbf{f}_{CSF}] \tag{3}$$

For two-level feature fusion, a simple concatenation approach is used as these fused features will be refined by feature boosting and dimension reduction process.

3.5. Feature boosting and dimension reduction

Enriched features from multi-scale convolutional layers and different tissue regions may have high correlations hence redundant for the AD classifier. This may lead to "the curse of dimensionality" in the classifier, as the fused feature dimension is very high.

Among many possible choices of dimension reduction, we chose a feature boosting and dimension reduction method similar to [8] in our scheme, where the features are ranked according to their importance for AD classification. Only features with high importance are chosen for the final classification.

The gradient boosting-based machine learning method XGBoost [8] uses an ensemble model by summing up the prediction values from multiple decision trees growing during the iterations. Let $\{(\mathbf{f}_m, y_m), m = 1, \ldots, n\}$, be the given dataset for the training, where $\mathbf{f}_m$ is the feature vector of mth brain scan, y_m is the label. Let $\hat{y_m}^{(j)}$ be the predicted label at the jth iteration, then a new tree structure q_j is added to minimize the following objective function $L^{(j)}$:

$$L^{(j)} = \sum_{m=1}^{n} l(y_m, \ \hat{y_m}^{(j-1)} + q_j(\mathbf{f}_m)) + \Omega(q_j) \tag{4}$$

where l is a loss function measuring the difference between the predicted and the true label, $\Omega = \gamma J + 1/2\lambda\|\mathbf{w}\|^2$ is the regularization term that penalizes the complexity of the model, where $\mathbf{w}$ is the leaf weights in the tree T, and J is the number of leaves in the tree T. By applying XGBoost, feature importance can be obtained by a score indicating how valuable each feature is when constructing a boosted decision tree. The more a feature is used for the decision, the higher is its relative importance. The gain after the tree split can be obtained as

$$G = \frac{g_L^2}{h_L + \lambda} + \frac{g_R^2}{h_R + \lambda} - \frac{g^2}{h + \lambda} \tag{5}$$

where g and h are the first and second order gradient on the loss function $l(\cdot)$ with respect to the predicted label, the subscript L and R are the left and right nodes after the tree split. In a single decision tree, the relative influence of the kth feature in the tree T is computed as:

$$U_k(T) = \sum_{i=1}^{J-1} G_i \ \mathbb{1}(v_i = k) \tag{6}$$

where the summation is over all nonterminal nodes i, $i = 1, \ldots, J - 1$, in the tree T, v_i is the feature used for splitting associated with the node i, and G_i is the corresponding gain after the splitting. The feature importance is then averaged across all decision trees T_p, $p = 1, \ldots, M$, within the model.

$$U_k = \frac{1}{M} \sum_{p=1}^{M} U_k(T_p) \qquad (7)$$

Each feature in **f** is ranked according to its relative importance indicated by (7), followed by a threshold to select a certain number of features for the subsequent classification. It is worth noting that important features are selected and recorded based on training data. For testing, feature indices obtained in the training process are directly used. Finally, selected features are fed to a two-layer neural network for final AD classification.

3.6. Implementation issues

Data pre-processing: Pre-processing is an important step in AD detection as the original scan may contain other irrelevant parts to the brain such as neck and skull, which may hamper the feature learning of the brain. The pre-processing consists of three steps, cortical reconstruction, image size normalization and intensity normalization. Cortical reconstruction was conducted by the dataset provider. It is done by using the *recon-all* function from the FreeSurfer software package [27] including a set of pre-processing functions such as motion correction and conform, non-uniform intensity normalization, Talairach transform computation, intensity normalization, skull and neck removal. It is observed that the original images have much zero-value background, which is not helpful for training the network and increases computational cost. Thus, it is necessary to remove background regions and rescale samples. The image size normalization has the following two steps. First each 3D image volume is trimmed according to the largest brain size in the dataset. After that, trimmed 3D images are then resized to 112*96*96 as the inputs. The last step of pre-processing is the intensity normalization that normalizes image intensities to [0,1].

Training issues:

(a) Pre-training and refined-training strategy: To capture the characteristics of different tissue regions, pre-training and refined-training strategy is applied for effective feature learning. That is, 3D scans with whole brain images from the training and the validation subsets are fed to a 3D MSCNN for pre-training, followed by refined-training using separate streams of segmented brain tissue regions (e.g., WM, GM or CSF) as the inputs. The refined-training is done by initializing the network with pre-trained coefficients and using a smaller learning rate and fewer epochs for the training. The advantage of this training strategy is that the network initially trained by the whole brain scans has already learned some prior knowledge from GM, WM and CSF. By pre-training the network using the whole brain scans and refined-training using segmented tissue regions GM, WM and CSF, more effective features can be learned as compared to training the network directly from the scratch with segmented tissue regions.

(b) Multi-stage training: Due to using 3D MSCNNs with multi-stream inputs, and the constraint in GPU memory (Titan XP with 12 GB memory was used), it is not feasible to perform end-to-end training in the proposed scheme. Therefore, the refined-training is further split into two stages. In stage-1 refined-training, each individual stream of 3D MSCNN is trained by the regions from a selected tissue type (e.g. WM; GM or CSF) followed by the first level of future fusion. This is done in an end-to-end fashion on each individual stream. In stage-2 of refined-training, multi-scale feature vectors from all 3D MSCNN streams (where 3D MSCNN coefficients are fixed from the stage-1 refined-training) are fused and then used as the inputs. Stage-2 refined-training includes applying XGBoost for feature boosting and dimension reduction, and FC (fully connected) layers as the classifier. The same way of splitting training, validation and testing subsets is applied in the training, including pre-training and both stages of refined-training.

Algorithm 1 The proposed scheme for AD detection.

DATA PREPARATION

1. Preprocessing of brain scans: cortical reconstruction, image size normalizationand intensity normalization;
2. Segmentation of brain scans into three tissue regions: GM, WM and CSF;
3. Splitting the dataset into training, validation and testing subsets (70%, 15%, 15%).

TRAINING

4. *Pre-training:*
   Pre-training the model on the entire area of brain scans on the training and validation subsets, using 3D MSCNN;
5. *Stage-1 refined training and level-1 fusion:*
   For TissueRegion=GM, WM, CSF do:
   5.1 Stage-1 refined-training on TissueRegion, using 3D MSCNN coefficients obtained from the initial model in pre-training stage;
   5.2 Level-1 fusion on the multiscale feature vector (after 'flatten') related to the TissueRegion;
   End {For}
6. *Level-2 fusion, feature boosting, and stage-2 training:*
   6.1 Level-2 feature fusion on GM, WM, CSF to form a long feature vector;
   6.2 Applying feature boosting XGBoost for reducing feature dimensions;
   6.3 Stage-2 training of FC layers as the classifier.
   Output: trained coefficients in 3DCNN models and FC layers, trained indices in XGBoost.

TESTING

7. For TissueRegion=GM, WM, CSF do:
   7.1 Extract features from TissueRegion on the testing brain scan;
   7.2 Level-1 fusion on the multiscale feature vector (after 'flatten') related to the TissueRegion;
   End {For}
8. Level-2 feature fusion on GM, WM, CSF to form a long feature vector;
9. Apply feature reduction on testing brain scan using the trained XGBoost;
10. Apply trained FC layers for the classification.
Output: class labels

Algorithm 1 summarizes the pseudo code of the proposed AD detection scheme.

4. Experimental results and performance evaluation

4.1. Setup and dataset used

Setup: KERAS library [9] with TensorFlow [1] backend was used for network training. Network weights were learnt using *Adam* optimizer. The following hyper-parameters were chosen after carefully tuning the networks through experiments. Number of epochs in pre-training was 150. Step-wise learning rate was used, it was set to 0.001 during epoch 1–50, 0.0001 during epoch 51–100, and 0.00001 during epoch 101–150. Number of epochs in refined-training was 50, where learning rate is 0.0001 during epoch 1–25, 0.00001 during epoch 26–50. Dropout rate used in flatten layers and concatenation layer was set to 0.5. For all the layers except the output layer, L2 regularization was used with regularization parameter as 0.00005. Batch size was set to 8. Batch normalization momentum was 0.9. Initialization method for all layers was glorot_uniform. The conventional criteria for accuracy and cross-entropy loss [15] were used for the performance evaluation.

Table 1
Description of ADNI dataset used in our experiments.

Class	# Subjects	#3D MRI Scans
AD	198	600
NC	139	598
Total	337	1198

Table 2
Training accuracies in different phases.

Training phase	Region of tissue type	Accuracy
Pre-training	whole brain images	98.08%
Stage-1 refined-training	GM	98.44%
	WM	99.88%
	CSF	99.52%
Stage-2 refined-training	XGBoost of fused features of all streams	99.99%

Table 3
Classification performance on the test set from using different tissue type combinations: overall performance and the performance on individual class AD.

# Tissue regions	Tissue types	Testing Performance (Overall)	AD Detection rate (%)	AD False alarm (%)
1	GM	88.95	91.01	13.25
	WM	87.21	91.01	16.87
	CSF	92.44	94.38	9.64
2	WM+CSF	88.95	91.01	13.25
	GM+WM	88.37	86.52	9.64
	GM+CSF	**94.74**	**95.50**	**6.02**
3	GM+WM+CSF	91.86	93.26	9.64

Dataset: Experiments were conducted on a ADNI dataset from Alzheimer's Disease Neuroimaging Initiative [2]. In our experiments, T1 MRI scans of 2 classes (AD and NC) are used for classification. Detailed information of the dataset is described in Table 1.

Observing Table 1, one can see that for each subject, there is an average of 3.5 scans, which were made at different times. The whole dataset is then partitioned to 70%, 15%, 15% to form training set, validation set and testing set, respectively.

4.2. Training performance in different stages

As described previously (see Section 3.6), the training process consists of pre-training and two stages of refined-training, Table 2 shows the performance from each of these steps.

Observing the results in Table 2, one can see that the training performance using regions from each individual tissue type is already very high in the stage-1 refined-training, hence, 2nd level feature fusion from different streams and stage-2 refined-training cannot anticipate much gain in the training.

However we shall describe in the subsequent text that significant advantage is obtained on the *test set* through fusion of features in different streams (i.e. second-level fusion, see Table 3) and stage-2 refined-training (e.g., XGBoost for significant feature reduction while maintaining high performance, see Table 12).

One possible reason for very high training performance and the gap to the testing performance (see Section 4.4) is probably due to the relatively moderate size of our dataset available for download and public use.

Since the test performance is related to the generalization performance of the classifier, we only include the evaluation of test performance in the remaining studies.

Furthermore, Fig. 4 shows the overall training accuracies as a function of epochs from the pre-training, stage-1 refined-training on individual streams. Noting that in the training, the learning rate for refined-training was different from pre-training. From the re-
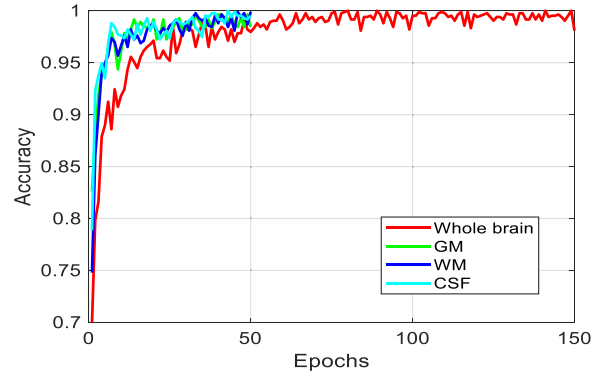

Fig. 4. Overall training accuracy from pre-training, and stage-1 refined-training on individual streams of 3D MSCNNs using different tissue type regions.

sulting curves in Fig. 4, one can observe that refined-training performance converges very fast (50 epochs is sufficient) due to the use of pre-training.

4.3. Determine the number of streams by selecting tissue type combinations

In this section, we study the use of different combination of tissue types in the network. The aim of the tests is to determine the best combination of streams used in the proposed scheme.

To evaluate the overall performance on the testset, Table 3 shows the overall test performance by using different types of tissue combinations, as well as the test performance on individual AD detection and false alarm rate through examining the individual classes (AD vs NC) in the given tissue combination.

Observing the 3rd column in Table 3, one can see that the tissue information fusion on GM and CSF has yielded the best test performance. This is different from the intuitive guess that a combination of all three tissue types would result in the best performance. One possible explanation is due to over-fitting. That is, using fused features from three types of tissues is more likely to result in over-fitting, as tissue information fusion on three types leads to a much higher number of features as compared with that of two type combination, while the training dataset is not sufficiently large.

Observing the last 2 columns of Table 3, tissue information fusion on GM and CSF also achieved the highest AD detection rate and lowest false alarm on the testing set. It indicates that tissue information fusion of GM and CSF are most effective for AD detection, while adding more tissue type (WM) does not improve the overall performance.

Hence, in the subsequent studies, we only focus on using two streams (GM and CSF) for the remaining experiments.

4.4. Performance of the proposed method

To test the effectiveness of the proposed scheme, experiments were conducted on two sets of dataset partition methods: case study-1 using random partition, and case study-2 using subject-separated partition. For case study-1, all brain scans were mixed together without considering subject information when partitioning them into training, validation and testing sets. For case study-2, brain scans from the same subject were used in one kind of set (training, validation or testing) without mixed usage. This means that the same subject in the training set would not appear in the validation set or testing set.

Case study-1: Results using random dataset partition

To test the performance of the proposed method, Table 4 shows the performance of the proposed method when dataset was

Table 4

Overall test accuracies from case study-1, where dataset is randomly partitioned. For each run, the training/validation/test subsets were re-partitioned. All the results were obtained when the training parameters were fixed at epoch=50. The average value and the standard deviation $|\sigma|$ are also included.

Run	1	2	3	4	5	Average ($|\sigma|$)
Accuracy (%)	97.33	99.00	**99.67**	98.56	96.89	98.29 (1.04)

Table 5

Confusion matrix on the testing set for individual classes AD and NC in case study-1 (from the best performance run 3 in Table 4).

	ÂD	N̂C
AD	99.56%	0.44%
NC	0.22%	99.78%

Table 6

Overall test accuracies from case study-2, where dataset was randomly partitioned according to subjects. For each run, the training/validation/test subsets were randomly re-partitioned according to subjects. All results were obtained when the training parameters were fixed at epoch=50. The average value and the standard deviation $|\sigma|$ are also included.

Run	1	2	3	4	5	Average ($|\sigma|$)
Accuray (%)	**94.74**	88.07	85.64	93.53	85.56	89.51 (3.90)

Table 7

Confusion matrix on the testing set for individual classes AD and NC in case study-2 (from the best performance run 1 in Table 6).

	ÂD	N̂C
AD	95.50%	4.50%
NC	6.02%	93.98%

Table 8

Summarizing the results (in tables) from different tests in Section 4.5. These tests were aimed at examining different settings and their impact to the proposed scheme.

	Setting	Results
(1)	With/without brain segmentation	Table 9
(2)	Fusion of multi-scale features	Table 10
(3)	Network with/without pre-training	Table 11
(4)	With/without feature reduction	Table 12

Table 9

Test performance with/without segmentation: overall performance, and performance on individual class AD.

Segmentation	Overall test accuracy (%)	AD detection rate (%)	AD false alarm (%)
With	**94.74**	**95.50**	**6.02**
Without	90.12	91.01	10.84

Table 10

Test performance from different scale levels: overall performance and the performance on the individual class AD.

Scale level	Overall test accuracy (%)	AD Detection rate (%)	AD false alarm (%)
1	91.28	94.38	12.05
2	84.88	94.38	25.30
3	90.70	93.26	12.05
4	89.53	89.89	10.88
1+2+3+4	**94.74**	**95.50**	**6.02**

expected since scans of the same person may appear in both training and testing set in case study-1, and these scans may have high correlation even though they were made at different times.

4.5. Performance influenced by different network/parameter settings

In this subsection, empirical test results are included which were obtained from using different network and settings. The aim of these tests is to gain deep insight into the proposed scheme through different types of empirical tests. Table 8 summarizes the list of tables in the section below where results from different types of tests were included.

(1) With/without brain tissue segmentation:

To examine the impact of brain tissue segmentation, comparison was made with the method using whole brain images as the input, followed by only scale-level feature fusion, feature boosting and feature dimension reduction. Table 9 shows the comparison of overall test performance, as well as the test performance on the individual AD class.

Observing Table 9, the proposed scheme with segmentation is shown to have achieved better test accuracy (in the 2nd column), with higher AD detection rate and lower false alarm rate (in the 3rd and 4th column). It indicates that segmenting the brain into different tissue regions for feature learning, combined with feature fusion is effective for AD detection.

(2) Fusion of multi-scale features:

To examine the impact of fusion of multi-scale features, tests were conducted using single-scale features and multi-scale combined features, Table 10 shows the overall test performance in different scale levels, as well as the test performance on the individual class AD in different scale levels.

Observing Table 10, the proposed scheme with multi-scale feature fusion resulted in the highest overall testing accuracy, also the highest detection rate (95.50%) and lowest false alarm rate (6.02%) for the individual class AD.

(3) Network with/without pre-training:

partitioned randomly for 5 times. To examine the performance on the individual class, especially on AD, Table 5 shows the confusion matrix for the test performance from the best performance run.

Observing Tables 4 and 5, the proposed method is shown to be effective with high classification accuracy (99.67%), and high AD detection rate (99.56%) with low false alarm rate (0.22%) on the testing set. Observing Table 4, the standard deviation of the test accuracies on different partitions is small, which shows the robustness of the proposed scheme.

Case study-2: Results from using subject-separated dataset partition

The performance of the proposed scheme was tested for multiple runs, where partitions of training, validation and test subsets in the dataset were done randomly according to subjects, i.e., brain scans of each subject belong only to either the training or the test subset but not both. A random partition was done in each of the 5 runs. Table 6 shows the performance on the testing set. To examine the performance on the individual class, especially on AD, Table 7 shows the confusion matrix for the performance on the testing set from the best performance run.

Observing Table 6, the proposed scheme is shown to be effective with relatively high average classification accuracy (89.51%), where the highest is 94.74% and lowest 85.56%. This shows the proposed method is still rather robust, however, the performance is much lower than that in case study-1. It is also noticed that the standard deviation of test accuracies from different runs has increased to 3.90%, this is probably due to the moderate size of our training dataset. The results in Table 7 show that AD and NC detection rates are relatively balanced, also the AD detection rate was high (95.50%) and the false alarm rate was relatively low (6.02%) though there is a significant increase as compared with that in case study-1. Since in case study-1 the dataset was partitioned randomly according to brain scans, such drop in performance is

Table 11
Test performance with/without pre-training: overall performance and the performance on the individual class AD.

Pre-training	Overall test accuracy (%)	AD detection rate (%)	AD false alarm (%)
With	**94.74**	**95.50**	**6.02**
Without	91.28	94.38	12.05

Table 12
Test performance with/without feature boosting and dimension reduction: overall performance and the performance on the individual class AD.

XGBoost	Dimension of features	Overall testing accuracy (%)	AD detection rate (%)	AD false alarm (%)
with	**207**	**94.74**	**95.50**	**6.02**
without	2048	94.19	94.38	6.02

To examine the impact of pre-training using whole brain scans, comparisons were made with the scheme directly using GM and CSF for training without pre-training. Table 11 shows the comparison of the overall test performance as well as the performance on the individual AD class, from the proposed schemes with and without pre-training.

Observing Table 11, applying pre-training has improved the overall accuracy (by 3.46%) on the testing set. Further, the performance on the individual class AD showed high detection rate (improved by 1.12%) with a low false alarm rate (decreased by 6.03%). These results also suggest that the whole brain scan contains some prior knowledge about different tissue regions. By pre-training the network using whole brain scan and refined training the network using GM and CSF, more effective features can be learned compared to training the network from scratch with GM and CSF.

(4) With/without feature boosting and dimension reduction:

To investigate the impact of applying XGBoost for feature boosting and dimension reduction, comparisons were made for the proposed scheme with and without applying this step. Table 12 shows the comparison of the overall test performance using the proposed scheme with/without feature boosting and dimension reduction. To further examine the performance on individual classes, especially on the AD class, Table 12 also shows the AD detection rate and false alarm rate on the testing set from the scheme with/without feature boosting and dimension reduction.

Comparing Table 12, the performance from the proposed scheme by using feature boosting and dimension reduction has resulted some improvement though not significant. Observing Table 12, an improvement of 1.12% on AD detection accuracy was obtained, which indicates that feature boosting and dimension reduction has reduced potential over-fitting by removing irrelevant features thus improved the classification performance.

4.6. Comparisons with state-of-the-art methods

Comparisons were made with eight existing methods on two kinds of settings, random dataset partition and subject-separated dataset partition. Results are shown in Tables 13 and 14.

Observing Table 13, the proposed scheme achieved the best result compared to other state-of-the-art methods. It is worth mentioning that AE+ [20] used both MRI and PET data while the proposed method used only T1 MRI scans. Still, the proposed method achieved the best result. Observing Table 14, when dataset was partitioned by separating different subjects, the proposed scheme showed comparable performance to other two methods. Although most methods in Tables 13 and 14 include Mild Cognitive Impairment (MCI) patients in their experiments, we only compare the

Table 13
Comparison of the proposed scheme with 6 existing state-of-the-art methods on random dataset partition.

Method	# Subjects AD/NC/MCI	# Brian Scans AD/NC/MCI	Accuracy (%) AD vs. NC
3D-AE-CNN [13]	70/70/70	–	97.60
AE+ [20]	65/77/169	–	87.76
SAE [11]	200/232/411	755/1278/2282	94.74
ICA [28]	202/236/410	–	85.70
MIL [26]	198/231/405	–	88.80
3DCNN [4]	198/139/–	600/598	98.74
Proposed	**198/139/–**	**600/598**	**99.67**

Table 14
Comparison of the proposed scheme with 2 existing state-of-the-art methods on subject-separated dataset partition.

Method	# Subjects AD/NC/MCI	#Scans AD/NC/MCI	Accuracy (%) AD vs. NC
SAE-CNN [21]	**755/755/755**	**755/755/755**	**95.39**
3DCNN [4]	198/139/–	600/598	90.11
Proposed	198/139/–	600/598	94.74

Table 15
Processing time for the training and testing.

Processing	Step	Time (second)
Training	Tissue segmentation	1.14 (one 3D brain scan)
	Pre-training on each 3D brain	50.05 (one epoch on training set)
	Stage-1 refined-training on each individual stream of tissue region	50.00 (one epoch on training set)
	Feature fusion	0.14 (one 3D scan on training set)
	XGBoost feature reduction	1.0 (whole training set)
	FC layers classifier training	1.0 (whole training set)
Testing	Tissue segmentation	1.14 (one 3D brain scan)
	Feature extraction + classification	0.141 (one 3D brain scan)

proposed method with their AD vs. NC performance, so such comparison is fair.

4.7. Processing Time

All experiments were conducted on a workstation with Intel-i7 3.40 GHz CPU, 48 G RAM and an NVIDIA Titan Xp 12 GB GPU, without code optimization. Detailed processing time of each step in the training and testing process is shown in Table 15, where the most time-consuming part is the training. The total training time was $7508 + 5000 + 117 + 2 = 12,627$ s (or, 3h 30 min and 27 s), where the pre-training using 150-epochs took $50.05*150 = 7508$ s, the 1st-stage refined-training using 50 epochs on 2 streams of GM and CSF regions took $50*50*2 = 5000$ s, feature fusion took $0.14*838 = 117.4$ s, and finally XBGoost plus FC layer training took 2.0 seconds. For testing time, it took $1.14 + 0.141 = 1.281$ s for each input of one 3D brain scan.

4.8. Discussion

From various empirical test results given above, we have demonstrated that the proposed scheme is effective for Alzheimers' disease detection. From different sets of experiments and the results, we may also draw the following conclusions:

- Overall performance: The proposed scheme is effective, with an excellent performance in case study-1 (highest 99.67%, average 98.29%), and good performance in case study-2 (highest 94.74 %, average 89.51%).

There is a drop of performance (average 8.78%) in case study-2 as comparing with case study-1, since AD detection is more challenging when training and testing subsets are partitioned according to subjects rather than 3D image scans.

- Individual class performance: Classification rates (also false alarms) are well balanced in the AD and NC classes in both case studies. For AD class, the best detection rate is 99.56% (or 95.50%), with false alarm 0.22% (or 6.02%) in 5 runs of case study-1 (or case study-2). There is a drop in AD detection rate and a big increase in false alarm in case study-2.
- Combining different tissue regions: Empirical tests using two streams, GM and CSF, have generated the best performance (94.74%), as comparing with the performance (91.86%) using three streams (GM, WM and CSF). The latter could probably be explained as due to overfitting, since three streams of multi-scale CNNs generate a very high dimension of feature vector, while the size of training dataset is moderate.
- Using segmented tissue regions: Empirical tests show that using segmented tissue regions in separate streams has led to an increase in overall test performance (about 4.62%).
- Using multiscale CNNs: Empirical tests show an increase in the test performance as comparing with those using only one scale ($\geq$ 3.46%). This is probably due to different levels containing semantic and many detailed tissue features.
- Pre-training: Whole 3D scan-based pre-training followed by tissue region-based refined-training is effective. It led to a fast convergence in the refined-training (50 epochs), and also an increase in performance (3.46%). The latter is probably due to adding cross-correlation information among different tissues.
- XGBoost: is shown to be rather efficient. Empirical tests show that XGBoost not only significantly reduced the feature dimension from its original size (10%) but also led to a small increase in performance (0.55%). This can probably be explained as XGBoost reduces overfitting when the dimension of multistream and multiscale features is high while the training dataset is moderate.
- Comparison to state-of-the-art: shows that proposed scheme achieved the best performance on randomly partitioned dataset, and second best on subject-separated dataset partition.

Limitations: Currently, the dataset used in our experiments is moderate in size, due to the limitation to public users. This has led to some overfitting especially when several streams of multiscale features lead to a very high dimension of feature vector. This can be observed from a very high training performance, and the gap between the training and testing performance. Apart from increasing the dataset, more sophisticated data augmentation such as Generative Adversarial Networks (GAN) shall be adopted in the near future work. Another limitation is that the proposed scheme currently cannot be trained end-to-end since the limitation of our computer's GPU memory. This could also introduce some performance degradation.

5. Conclusion

The proposed multi-stream multi-scale deep CNN scheme has been tested on Alzheimer's disease (AD) detection. Results from our experiments have shown that using two-stream brain regions (GM and CSF) as inputs, followed by multi-scale feature learning using 3D MSCNNs, is effective for AD detection. Two-level feature fusion is shown to be effective to capture the characteristics of AD. Furthermore, feature boosting and dimension reduction are shown to have improved performance by removing irrelevant features while retaining important features for classification. A range

of empirical tests have also been conducted for examining the impact of tissue segmentation, combination of tissue streams, multi-scale CNNs, pre-training, XGBoost. Among them, using segmented tissue regions and using multi-scale CNNs are the main contributors to a large performance improvement (4.62% and 3.46%, respectively), pre-training mainly contributes to the fast convergence in refined-training, XGBoost mainly contributes to significantly reducing feature dimension without compromising the performance. Comparing with eight existing methods, the proposed scheme has achieved the best performance on randomly partitioned dataset, and second best on subject-separated dataset partition. Some limitations are also discussed. Our future work will be focused on training the proposed scheme using large training dataset (e.g., with more real or GAN augmented data) to reduce possible overfitting in the training, seeking end-to-end training by GPUs with large internal memory, and adding MCI subjects into the proposed scheme which could be clinically important.

References

[1] M. Abadi, A. Agarwal, in: TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. https://www.tensorflow.org/.
[2] ADNI, in: Alzheimer's disease neuroimaging initiative, 2017. https://adni.loni.usc.edu/about/.
[3] E. Arvesen, Automatic classification of Alzheimer's disease from structural MRI, 2015 Master's thesis.
[4] K. Bäckström, M. Nazari, I.Y.-H. Gu, A.S. Jakola, An efficient 3d deep convolutional network for Alzheimer's disease diagnosis using mr images, in: Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, IEEE, 2018, pp. 149–153.
[5] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H.M. Arrighi, Forecasting the global burden of Alzheimer's disease, Alzheimer's Dementia J. Alzheimer's Assoc. 3 (3) (2007) 186–191.
[6] T. Brosch, R. Tam, A.D.N. Initiative, et al., Manifold learning of brain Mris by deep learning, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 633–640.
[7] C. Carrion, F. Folkvord, D. Anastasiadou, M. Aymerich, Cognitive therapy for dementia patients: a systematic review., Dement. Geriatr. Cogn. Disord. 46 (1–2) (2018) 1–26.
[8] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 785–794.
[9] F. Chollet, et al., in: Keras, 2015. https://github.com/fchollet/keras.
[10] Y. Fan, D. Shen, R.C. Gur, R.E. Gur, C. Davatzikos, Compare: classification of morphological patterns using adaptive regional elements, IEEE Trans. Med. Imaging 26 (1) (2007) 93–105.
[11] A. Gupta, M. Ayhan, A. Maida, Natural image bases to represent neuroimaging data, in: Proceedings of the International Conference on Machine Learning, 2013, pp. 987–994.
[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[13] E. Hosseini-Asl, R. Keynton, A. El-Baz, Alzheimer's disease diagnostics by adaptation of 3D convolutional network, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 126–130.
[14] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith, FSL, Neuroimage 62 (2) (2012) 782–790.
[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[16] Q.V. Le, Building high-level features using large scale unsupervised learning, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2013, pp. 8595–8598.
[17] J.P. Lerch, J. Pruessner, A.P. Zijdenbos, D.L. Collins, S.J. Teipel, H. Hampel, A.C. Evans, Automated cortical thickness measurements from mri can accurately separate Alzheimer's patients from normal elderly controls, Neurobiol. Aging 29 (1) (2008) 23–30.

[18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the CVPR, 1, 2017, p. 4.

[19] M. Liu, D. Zhang, E. Adeli, D. Shen, Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 63 (7) (2016) 1473–1482.

[20] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: Proceedings of the IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, 2014, pp. 1015–1018.

[21] A. Payan, G. Montana, Predicting Alzheimer's disease: a neuroimaging study with 3d convolutional neural networks, in: ICPRAM, 2, SciTePress, 2015, pp. 355–362.

[22] S. Sarraf, G. Tofighi, et al., Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri, Biorxiv (2016) 070441.

[23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations, 2015.

[24] H.-I. Suk, D. Shen, Deep learning-based feature representation for Ad/MCI classification, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 583–590.

[25] S. Teipel, A. Drzezga, M.J. Grothe, H. Barthel, G. Chételat, N. Schuff, P. Skudlarski, E. Cavedo, G.B. Frisoni, W. Hoffmann, et al., Multimodal imaging in alzheimer's disease: validity and usefulness for early detection, Lancet Neurol. 14 (10) (2015) 1037–1053.

[26] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J.V. Hajnal, D. Rueckert, Multiple instance learning for classification of dementia in brain Mri, Med. Image Anal. 18 (5) (2014) 808–818.

[27] Website, in: Freesurfer, 2017. https://surfer.nmr.mgh.harvard.edu/fswiki.

[28] W. Yang, R.L. Lui, J.-H. Gao, T.F. Chan, S.-T. Yau, R.A. Sperling, X. Huang, Independent component analysis-based classification of Alzheimer's disease mri data, J. Alzheimer's Dis. 24 (4) (2011) 775–783.

[29] Y. Zhang, M. Brady, S. Smith, Segmentation of brain mr images through a hidden Markov random field model and the expectation-maximization algorithm, IEEE Trans. Med. Imaging 20 (1) (2001) 45–57.
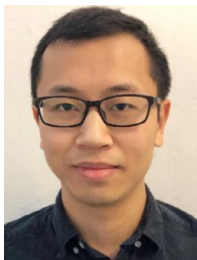
Chenjie Ge received his bachelor degree from Harbin Institute of Technology (HIT), China in 2013. Currently he is a dual doctoral degree student at Department of Electrical Engineering, Chalmers University of Technology, Sweden and Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University (SJTU), China, under the supervision of Prof. Irene Yu-Hua Gu and Prof. Jie Yang. His research interests are computer vision, image processing, machine learning and deep learning, with applications to, e.g. medical image analysis, human activity classification and visual saliency detection.
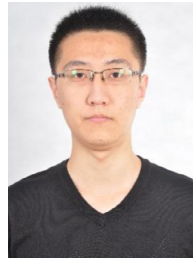
Qixun Qu receivd the M.Sc degree in biomedical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 2018. Since then, he has been an algorithm engineer focusing on health care field. His research and work interests cover the designing and applying machine learning methods to health management in daily routine.

Irene Yu-Hua Gu received the Ph.D. degree in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992. From 1992 to 1996, she was Research Fellow at Philips Research Institute IPO, Eindhoven, The Netherlands, and post dr. at Staffordshire University, Staffordshire, U.K., and Lecturer at the University of Birmingham, Birmingham, U.K. Since 1996, she has been with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, where she has been a full Professor since 2004. Her research interests include statistical image and video processing, object tracking and video surveillance, machine learning and deep learning, and signal processing with applications to electric power systems. Dr. Gu was an Associate Editor for the IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, and Part B: Cybernetics from 2000 to 2005, and an Associate Editor for the EURASIP Journal on Advances in Signal Processing from 2005 to 2016. She was the Chair of the IEEE Swedish Signal Processing Chapter from 2001 to 2004. She has been with the Editorial board of the Journal of Ambient Intelligence and Smart Environments since 2011.

Asgeir Store Jakola, a medical doctor since 2006, received PhD in medicine in 2013 from Norwegian University of Science and Technology, Trondheim, Norway. He is a clinically active neurosurgeon with a special interest in brain tumor research, medical technology and clinical research with the fields of neurology, neurosurgery and oncology. He was affiliated with Sahlgrenska Academy, Institution of Neuroscience and physiology, Gothenburg, Sweden as adjunct lecturer in 2015 and as associate professor from 2016. Dr. Jakola has since 2015 been editorial board member in ActaNeurochirurgica and he was elected as executive board member of EANO in 2018.