frontiers in Neuroscience

Convolutional Neural Networks-based MRI Image Analysis for the Alzheimer's Disease Prediction from Mild Cognitive Impairment

Weiming Lin[1, 2, 3], Tong Tong[4, 2], Qinquan Gao[4, 1, 2], Di Guo[5], Xiaofeng Du[5], Yonggui Yang[6], Gang Guo[6], Min Xiao[3], Min Du[1, 7*], Xiaobo Qu[8*]

[1]College of Physics and Information Engineering, Fuzhou University, China, [2]Fujian Provincial Key Laboratory of Medical Instrument and Pharmaceutical Technology, China, [3]School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, China, [4]Imperial Vision Technology, China, [5]School of Computer and Information Engineering, Xiamen University of Technology, China, [6]Department of Radiology, No.2 Hospital Xiamen, China, [7]Fujian Provincial Key Laboratory of Eco-industrial Green Technology, China, [8]Department of Electronic Science, Xiamen University, China

Provisional

Submitted to Journal:
Frontiers in Neuroscience

Specialty Section:
Neurodegeneration

ISSN:
1662-453X

Article type:
Original Research Article

Received on:
04 Jul 2018

Accepted on:
05 Oct 2018

Provisional PDF published on:
05 Oct 2018

Frontiers website link:
www.frontiersin.org

Citation:
Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, Guo G, Xiao M, Du M and Qu X(2018) Convolutional Neural Networks-based MRI Image Analysis for the Alzheimer's Disease Prediction from Mild Cognitive Impairment. *Front. Neurosci.* 12:777. doi:10.3389/fnins.2018.00777

Copyright statement:
© 2018 Lin, Tong, Gao, Guo, Du, Yang, Guo, Xiao, Du and Qu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Provisional

Convolutional Neural Networks-based MRI Image Analysis for the Alzheimer's Disease Prediction from Mild Cognitive Impairment

Weiming Lin[1,2,3], Tong Tong[3,4], Qinquan Gao[1,3,4], Di Guo[5], Xiaofeng Du[5], Yonggui Yang[6], Gang Guo[6], Min Xiao[2], Min Du[1,7*], Xiaobo Qu[8*], and the Alzheimer's Disease Neuroimaging Initiative

[1]College of Physics and Information Engineering, Fuzhou University, Fuzhou, China

[2]School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen, China

[3]Fujian Provincial Key Laboratory of Medical Instrument and Pharmaceutical Technology, Fuzhou, China

[4]Imperial Vision Technology, Fuzhou, China

[5]College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China

[6]Department of Radiology, Xiamen 2nd Hospital, Xiamen, China

[7]Fujian Provincial Key Laboratory of Eco-industrial Green Technology, Nanping, China

[8]Department of Electronic Science, Xiamen University, Xiamen, China

*** Correspondence:**
Min Du
dm_dj90@163.com
Xiaobo Qu
quxiaobo@xmu.edu.cn

Provisional

Keywords: Alzheimer's disease, deep learning, convolutional neural networks, mild cognitive impairment, magnetic resonance imaging.

Abstract

Mild cognitive impairment (MCI) is the prodromal stage of Alzheimer's disease (AD). Identifying MCI subjects who are at high risk of converting to AD is crucial for effective treatments. In this study, a deep learning approach based on convolutional neural networks (CNN), is designed to accurately predict MCI-to-AD conversion with magnetic resonance imaging (MRI) data. First, MRI images are prepared with age-correction and other processing. Second, local patches, which are assembled into 2.5 dimensions, are extracted from these images. Then, the patches from AD and normal controls (NC) are used to train a CNN to identify deep learning features of MCI subjects. After that, structural brain image features are mined with FreeSurfer to assist CNN. Finally, both types of features are fed into an extreme learning machine classifier to predict the AD conversion. The proposed approach is validated on the standardized MRI datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project. This approach achieves an accuracy of 79.9% and an area under the receiver operating

characteristic curve (AUC) of 86.1% in leave-one-out cross validations. Compared with other state-of-the-art methods, the proposed one outperforms others with higher accuracy and AUC, while keeping a good balance between the sensitivity and specificity. Results demonstrate great potentials of the proposed CNN-based approach for the prediction of MCI-to-AD conversion with solely MRI data. Age correction and assisted structural brain image features can boost the prediction performance of CNN.

1 Introduction

Alzheimer's disease (AD) is the cause of over 60% of dementia cases (Burns and Iliffe, 2009), in which patients usually have a progressive loss of memory, language disorders and disorientation. The disease would ultimate lead to the death of patients. Until now, the cause of AD is still unknown, and no effective drugs or treatments have been reported to stop or reverse AD progression. Early diagnosis of AD is essential for making treatment plans to slow down the progress to AD. Mild cognitive impairment (MCI) is known as the transitional stage between normal cognition and dementia (Markesbery, 2010), about 10% to 15% individuals with MCI progress to AD per year (Grundman et al., 2004). It was reported that MCI and AD were accompanied by losing gray matter in brain (Karas et al., 2004), thus neuropathology changes could be found several years before AD was diagnosed. Many previous studies used neuroimaging biomarkers to classify AD patients at different disease stages or to predict the MCI-to-AD conversion (Cuingnet et al., 2011; Zhang et al., 2011; Tong et al., 2013; Guerrero et al., 2014; Suk et al., 2014; Cheng et al., 2015; Eskildsen et al., 2015; Li et al., 2015; Liu et al., 2015; Moradi et al., 2015; Tong et al., 2017). In these studies, structural magnetic resonance imaging (MRI) is one of the most extensively utilized imaging modality due to non-invasion, high resolution and moderate cost.

To predict MCI-to-AD conversion, we separate MCI patients into two groups by the criteria that whether they convert to AD within 3 years or not (Moradi et al., 2015; Tong et al., 2017). These two groups are referred to as MCI converters and MCI non-converters. The converters generally have more severe deterioration of neuropathology than that of non-converters. The pathological changes between converters and non-converters are similar to those between AD and NC, but much milder. Therefore, it much more difficult to classify converters/non-converters than AD/NC.. This prediction with MRI is challenging because the pathological changes related to AD progression between MCI non-converter and MCI converter are subtle and inter-subject variable. For example, ten MRI-based methods for predicting MCI-to-AD conversion and six of them perform no better than random classifier (Cuingnet et al., 2011). To reduce the interference of inter-subject variability, MRI images are usually spatially registered to a common space (Coupe et al., 2012; Young et al., 2013; Moradi et al., 2015; Tong et al., 2017). However, the registration might change the AD related pathology and loss some useful information. The accuracy of prediction is also influenced by the normal aging brain atrophy, with the removal of age-related effect, the performance of classification was improved (Dukart et al., 2011; Moradi et al., 2015; Tong et al., 2017).

Machine learning algorithms perform well in computer-aided predictions of MCI-to-AD conversion (Dukart et al., 2011; Coupe et al., 2012; Wee et al., 2013; Young et al., 2013; Moradi et al., 2015; Beheshti et al., 2017; Cao et al., 2017; Tong et al., 2017). In recent years, deep learning, as a promising machine learning methodology, has made a big leap in identifying and classifying patterns of images (Li et al., 2015; Zeng et al., 2016; Zeng et al., 2018). As the most widely used architecture of deep learning, convolutional neural networks (CNN) has attracted a lot of attention due to its great success in image classification and analysis (Gulshan et al., 2016; Nie et al., 2016; Shin et al., 2016; Rajkomar et al., 2017; Du et al., 2018). The strong ability of CNN motivates us to develop a CNN-based prediction method of AD conversion.

In this work, we propose a CNN-based prediction approach of AD conversion using MRI images. A CNN-based architecture is built to extract high level features of registered and age-corrected hippocampus images for classification. To further improve the prediction, more morphological information is added by including FreeSurfer-based features (FreeSurfer, RRID:SCR_001847) (Fischl

This is a provisional file, not the final typeset article

and Dale, 2000; Fischl et al., 2004; Desikan et al., 2006; Han et al., 2006). Both CNN and FreeSurfer features are fed into an extreme learning machine as classifier, which finally makes the decision of MCI-to-AD. Our main contributions to boost the prediction performance include: 1) Multiple 2.5D patches are extracted for data augmentation in CNN; 2) both AD and NC are used to train the CNN, digging out important MCI features; 3) CNN-based features and FreeSurfer-based features are combined to provide complementary information to improve prediction. The performance of the proposed approach was validated on the standardized MRI datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI - Alzheimer's Disease Neuroimaging Initiative, RRID:SCR_003007) (Wyman et al., 2013) and compared with other state-of-the-art methods (Moradi et al., 2015; Tong et al., 2017) on the same datasets.

2 Materials and methods

The proposed framework is illustrated in Figure 1. The MRI data were processed through two paths, which extract the CNN-based and FreeSurfer-based image features, respectively. In the left path, CNN is trained on the AD/NC image patches and then is employed to extract CNN-based features on MCI images. In the right path, FreeSurfer-based features which were calculated with FreeSurfer software. These features, which were further mined with dimension reduction and sparse feature selection via PCA and Lasso, respectively, were concatenated as a features vector and fed to extreme learning machine as classifier. Finally, to evaluate the performance of the proposed approach, the leave-one-out cross validation is then used.

ADNI data

Data used in this work were downloaded from the ADNI database. The ADNI is an ongoing, longitudinal study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. The ADNI study began in 2004 and its first six-year study is called ADNI1. Standard analysis sets of MRI data from ADNI1 were used in this work, including 188 AD, 229 NC and 401 MCI subjects (Wyman et al., 2013). These MCI subjects were grouped as: 1) MCI converters who were diagnosed as MCI at first visit, but converted to AD during the longitudinal visits within 3 years (n = 169); 2) MCI non-converters who did not convert to AD within 3 years (n = 139). The subjects who were diagnosed as MCI at least twice, but reverse to NC at last, are also considered as MCI non-converters; 3) Unknown MCI subjects who missed some diagnosis which made the last state of these subjects was unknown (n = 93). The demographic information of the dataset are presented in Table 1. The age ranges of different groups are similar. The proportions of male and female are close in AD/NC groups while proportions of male are higher than female in MCI groups.

Image preprocessing

MRI images were preprocessed following steps in (Tong et al., 2017). All images were first skull-stripped according to (Leung et al., 2011), and then aligned to the MNI151 template using a B-spline free-form deformation registration (Rueckert et al., 1999). In the implementation, we follow the Tong's way to register images (Tong et al., 2017), showing that the effect of deformable registration with a control point spacing between 10 and 5 mm have the best performance in classifying AD/NC and converters/non-converters. After that, image intensities of the subjects were normalized by deform the histogram of each subject's image to match the histogram of the MNI151 template (Nyul and Udupa, 1999). Finally, all MRI images were in the same template space and had the same intensity range.

Age correction

Normal aging has atrophy effects similar with AD (Giorgio et al., 2010). To reduce the confounding effect of age-related atrophy, age correction is necessary to remove age-related effects, which is estimated by fitting a pixel regression model (Dukart et al., 2011) to the subjects' ages. We assume there are N healthy subjects and M voxels in each preprocessed MRI image, and denote $\mathbf{y}_m \in \mathbf{R}^{1\times N}$ as

the vector of the intensity values of N healthy subjects at m^{th} voxel, and $\boldsymbol{\alpha} \in \mathbf{R}^{1\times N}$ as the vector of the ages of N healthy subjects. The age-related effect is estimated by fitting linear regression model $\mathbf{y}_m=\omega_m\boldsymbol{\alpha}+b_m$ at m^{th} voxel. For n^{th} subject, the new intensity of mth voxel can be calculated as $y'_{mn}=\omega_m(C-\alpha_n)+ y_{mn}$, where y_{mn} is original intensity, α_n is age of nth subject. In this study, C is 75, which is the mean age of all subjects.

CNN-based features

A CNN was adopted to extract features from MRI Images of NC and AD subjects. Then, the trained CNN was used to extract image features of MCI subjects. To explore the multiple plane images in MRI, a 2.5D patch was formed by extracting three 32×32 patches from transverse, coronal and sagittal plane centered at a same point (Shin et al., 2016). Then, three patches were combined into a 2D RBG patch. Figure 2 shows an example of constructing 2.5D patch. For a given voxel point, three patches of MRI are extracted from three planes and then concatenated into a three channel cube, following the same way of composing a colorful patch with red/green/blue channels that are commonly used in computer vision. This process allows us to mine fruitful information form 3D views of MRI by feeding the 2.5D patch into the typical color image processing CNN network. Data augmentation (Shin et al., 2016) was used to increase training samples, by extracting multiple patches at different locations from MRI images. The choice of locations has three constraints, 1) The patches must be originated in either left or right hippocampus region which have high correlation with AD (van de Pol et al., 2006); 2) There must be at least two voxels distance between each location; 3) All locations were random chosen. With these constraints, 151 patches were extracted from each image and the sampling positions were fixed during experiments. The number of samples was expanded by a factor of 151, which could reduce over-fitting.

Typically extracted patches are presented in Figure 3. Figure 3(A) shows four 2.5D patches obtained from one subject. These patches are extracted from different positions and show different portions of hippocampus, which means these patches contain different information of morphology of hippocampus. When trained with these patches that spread in whole hippocampus, CNN learns the morphology of whole hippocampus. Figure 3(B) shows patches extracted in same position from four subjects of different groups, demonstrating that the AD subject has the most severe atrophy of hippocampus and expansion of ventricle. This implies that obvious differences are existed between AD and NC. However, the MCI subjects have the medium atrophy of hippocampus, and non-converter is more like NC rather than AD, and converter is more similar to AD. The difference between converter and non-converter is smaller than the difference between AD and NC.

The architecture of the CNN is summarized in Figure 4. The network has an input of 32×32 RGB patch. There are three convolutional layers and three pooling layers. The kernel size of convolutional layer is 5×5 with 2 pixels padding, and the kernel size and stride of pooling layers is 3×3 and 2. The input patch has a size of 32×32 and 3 RBG channels. The first convolutional layer generates 32 feature maps with a size of 32×32. After max pooling, these 32 feature maps were down-sampled into 16×16. The next two convolutional layers and average pooling layers finally generate 64 features maps with a size of 4×4. These features are concatenated as a feature vector, and then fed to full connection layer and softmax layer for classification. There are also rectified linear units layers and local response normalization layers in CNN, but are not shown for simplicity.

The CNN was trained with patches from NC and AD subjects, and there are 62967 (subject number 417 times 151) patches which are randomly split into 417 mini-batches. Mini-batch stochastic gradient descent was used to update the coefficients of CNN. In each step, a mini-batch was fed into CNN, and then error back propagation algorithm was carried out to computer gradient g_j of jth coefficient θ_j, and update the coefficient as $\theta'_j=\theta_j+\nabla\theta_j^n$, in which $\nabla\theta_j^n=m\nabla\theta_{1j}^{n-}-\eta(g_j+\lambda\theta_j)$ is the increment of θ_j at n^{th} step. The momentum m, learning rate η and weight decay λ are set as 0.9, 0.001 and 0.0001, respectively in this work. It is called one epoch with all mini-batches used to train CNN once. The CNN was trained with 30 epochs. Once the network was trained, CNN will be used to extract high level features of MCI subjects' images. The 1024 features output by the last pooling layer were taken as CNN-based features. Thus, CNN generates 154624 (1024×151) features for each image.

This is a provisional file, not the final typeset article

FreeSurfer-based features

The FreeSurfer (version 4.3) (Fischl and Dale, 2000; Fischl et al., 2004; Desikan et al., 2006; Han et al., 2006) was used to mine more morphological information of MRI images, such as cortical volume, surface area, cortical thickness average and standard deviation of thickness in each region of interest. These features can be downloaded directly from ADNI website, and 325 features are used to predict MCI-to-AD conversion after age correction. The age correction for FreeSurfer-based features is similar as described above, but on these 325 features instead of on intensity values of MRI images.

Features selection

Redundant features maybe exist among CNN-based features, thus we introduced the principle component analysis (PCA) (Avci and Turkoglu, 2009; Babaoğlu et al., 2010; Wu et al., 2013) and least absolute shrinkage and selection operator (LASSO) (Kukreja et al., 2006; Usai et al., 2009; Yamada et al., 2014) to reduce the final number of features.

PCA is an unsupervised learning method that uses an orthogonal transformation to convert a set of samples consisting of possibly correlated features into samples consisting of linearly uncorrelated new features. It has been extensively used in data analysis (Avci and Turkoglu, 2009; Babaoğlu et al., 2010; Wu et al., 2013). In this work, PCA is adopted to reduce the dimensions of features. Parameters of PCA are: 1) For CNN-based features, there are 1024 features for each patch. After PCA, P_C features were left for each patch, since there are 151 patches for one subject, there are still P_C×151 features for each subject.; 2) For FreeSurfer-based features, P_F features were left for each MCI subject.

LASSO is a supervised learning method that uses L_1 norm in sparse regression (Kukreja et al., 2006; Usai et al., 2009; Yamada et al., 2014) as follows:

$$\min_{\boldsymbol{\alpha}} \; 0.5||\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1 \tag{1}$$

Where $\mathbf{y} \in \mathbf{R}^{1\times N}$ is the vector consisting of N labels of training samples, $\mathbf{D} \in \mathbf{R}^{N\times M}$ is the feature matrix of N training samples consisting of M features, λ is the penalty coefficient that was set to 0.1, and $\boldsymbol{\alpha} \in \mathbf{R}^{1\times M}$ is the target sparse coefficients and can be used for selecting features with large coefficients. The LASSO was solved with least angle regression (Efron et al., 2004), and L features are selected after L iterations. Parameters of LASSO are: 1) For CNN-based features, L_C features were selected from P_C×151 features for each MCI subject; 2) For FreeSurfer-based features, L_F features were selected from P_F features. After PCA and LASSO, there were L_C+L_F features.

Figure 5 shows more details of CNN-based features. 151 patches are extracted from all MRI images, including AD, NC and MCI. First, the CNN is trained with patches of all AD and NC subjects. After that, the trained CNN is used to output 1024 features from each MCI patch. The 1024 features of each patch are reduced to P_C features by PCA, and then features of all 151 patches from one subject are concatenated, and Lasso is used to select L_C most informative features from them.

Extreme learning machine

The extreme learning machine, a feed-forward neural network with a single layer of hidden nodes, learns much faster than common networks trained with back propagation algorithm (Huang et al., 2012; Zeng et al., 2017). A special extreme learning machine, that adopts kernel (Huang et al., 2012) to calculates the outputs as formula (2) and avoids the random generation of input weight matrix, is chosen to classify converters/non-converters with both CNN-based features and FreeSurfer-based features. In formula (2), the $\boldsymbol{\Omega}$ is a matrix with elements $\Omega_{i,j}$=$\mathbf{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where $\mathbf{K}(\boldsymbol{a}, \boldsymbol{b})$ is a radial basis function kernel in this study, $[\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N]$ are N training samples, $\boldsymbol{y}$ is the label vector of training samples, and $\boldsymbol{x}$ is testing sample. C is a regularization coefficient and was set to 1 in this study.

Provisional

$$f(\boldsymbol{x}) = \begin{bmatrix} \mathbf{K}(\boldsymbol{x},\ \boldsymbol{x}_1) \\ \vdots \\ \mathbf{K}(\boldsymbol{x}, \boldsymbol{x}_\mathrm{N}) \end{bmatrix}^\mathrm{T} \left(\boldsymbol{\Omega} + {}^{1}/_{C}\right)^\mathrm{T} \boldsymbol{y} \tag{2}$$

Implementation

In our implementation, CNN was accomplished with Caffe[1], LASSO was carried out with SPAMS[2], and extreme learning machine was performed with shared online code[3]. The hippocampus segmentation was implemented with MALPEM[4](Ledig et al., 2015) for all MRI images. Then all hippocampus masks were registered as corresponding MRI images, and then overlapped to create a mask containing hippocampus regions. All image features were normalized to have zero mean and unit variance before training or selection. To evaluate the performance, Leave-one-out cross validation was used as (Coupé et al., 2012; Ye et al., 2012; Zhang et al., 2012).

3 Results

Validation of the robustness of 2.5D CNN

To validate the robustness of the CNN, several experiments have been performed with the CNN. In experiments, the binary decisions of CNN for 151 patches were united to make final diagnosis of the testing subject. We compared the performance in four different conditions:1) The CNN was trained with AD/NC patches and used to classify AD/NC subjects; 2) The CNN was trained with converters/non-converters patches and used to classify converters/non-converters; 3) The CNN was trained with AD/NC patches and used to classify converters/non-converters; 4) The condition is similar with 3), but with different sampling patches in each validation run.

The results are shown in Table 2. The CNN has a poor accuracy of 68.49% in classifying converters/non-converters when trained with converters/non-converters patches, but CNN has obtained a much higher accuracy of 73.04% when trained with AD/NC patches. This means that the CNN learned more useful information from AD/NC data than that from converters/non-converters data. And the prediction performance of CNN is close when different sampling patches are used.

Effect of combining two types of features

In this section, we present the performance of CNN-based features, FreeSurfer-based features, and their combinations. The P_C, P_F, L_C and L_F parameters were set to 29, 150, 35 and 40, respectively, which were optimized in experiments. Finally, 75 features were selected and fed to the extreme learning machine.

Performance was evaluated by calculating accuracy (the number of correctly classified subjects divided by the total number of subjects), sensitivity (the number of correctly classified MCI converters divided by the total number of MCI converters), specificity (the number of correctly classified MCI non-converters divided by the total number of MCI non-converters), and AUC (area under the receiver operating characteristic curve). The performances of the proposed method and the approach with only

[1] http://caffe.berkeleyvision.org/

[2] http://spams-devel.gforge.inria.fr/

[3] http://www.ntu.edu.sg/home/egbhuang/

[4] http://www.christianledig.com/

This is a provisional file, not the final typeset article

one type of features are summarized in Table 3. These results indicates that the approaches with only CNN-based features or FreeSurfer-based features have similar performances, and the proposed method combining both features achieved best accuracy, sensitivity, specificity and AUC. Thus, it is meaningful to combine two features in the prediction of MCI-to-AD conversion. The AUC of the proposed method reached 86.1%, indicating the promising performance of this method. The receiver operating characteristic (ROC) curves of these approaches are shown in Figure 6.

Impact of age correction

We investigated the impact of age correction on the prediction of conversion here. The prediction accuracy in Table 3 and the ROC curves in Figure 6 implied that age correction can significantly improve the accuracy and AUC, Thus, age correction is an important step in the proposed method.

Comparisons to other methods

In this section, we first compared the extreme learning machine with support vector machine and random forest. The performances of three classifiers are shown in Table 4, indicating that extreme learning machine achieves the best accuracy and AUC among three classifiers.

Then we compared the proposed method with other state-of-the-art methods that use the same data (Moradi et al., 2015; Tong et al., 2017), which consists of 100 MCI non-converters and 164 MCI converters. In both methods, MRI images were first preprocessed and registered, but in different ways. After that, features selection was performed to select the most informative voxels among all MRI voxels. Moradi used regularized logistic regression algorithm to select a subset of MRI voxels, and Tong used elastic net algorithm instead. Both methods trained feature selection algorithms with AD/NC data to learn the most discriminative voxels and then used to selected voxels from MCI data. Finally, Moradi used low density separation to calculate MRI biomarkers and to predict MCI converters/non-converters. Tong used elastic net regression to calculate grading biomarkers from MCI features, and SVM was utilized to classify MCI converters/non-converters with grading biomarker.

For fair comparisons, both ten-fold cross validation and leave-one-out cross validation were performed on the proposed method and method of (Tong et al., 2017) with only MRI data was used. Parameters of the compared approaches were optimized to achieve best performance. Table 5 shows the performances of three methods in ten-fold cross validation and Table 6 summarizes the performances in leave-one-out cross validations. These two tables demonstrate that the proposed method achieves the best accuracy and AUC among three methods, which means that the proposed method is more accurate in predicting MCI-to-AD conversion than other methods. The sensitivity of the proposed method is a little lower than the method of (Moradi et al., 2015) but much higher than the method of (Tong et al., 2017), and the specificity of the proposed method is between other two methods. Higher sensitivity means lower rate of missed diagnosis of converters, and higher specificity means lower rate of misdiagnosing non-converters as converters. Overall, the proposed method has a good balance between the sensitivity and specificity.

4 Discussions

The CNN has a better performance when trained with AD/NC patches rather than MCI patches, we think the reason is that the pathological changes between MCI converters and non-converters are slighter than those between AD and CN. Thus, it is more difficult for CNN to learn useful information directly from MCI data about AD-related pathological changes than from AD/NC data. The pathological changes are also hampered by inter-subject variations for MCI data. Inspired by the work in (Moradi et al., 2015; Tong et al., 2017) which use information of AD and NC to help classifying MCI, we trained the CNN with the patches from AD and NC subjects and improved the performance.

Provisional

After non-rigid registration, the differences between all subject's MRI brain image are mainly in hippocampus (Tong et al., 2017). So we extracted 2.5D patches only from hippocampus regions, that makes the information of other regions lost. For this reason, we included the whole brain features calculated by FreeSurfer as complementary information. The accuracy and AUC of classification are increased to 79.9% and 86.1% from 76.9% and 82.9% with the help of FreeSurfer-based features. To explore which FreeSurfer-based features contribute mostly when they are used to predict MCI-to-AD conversion, we used Lasso to select the most informative features, and the top 15 features are listed in Table 7, in which the features are almost volume and thickness average of regions related to AD. The thickness average of frontal pole is the most discriminative feature. The quantitative features of hippopotamus are not listed, indicating they contribute less than these listed features when predicting conversion. The CNN extract the deep features of hippopotamus morphology, rather than the quantitative features of hippopotamus, which are discriminative for AD diagnosis. Therefore, The CNN-based features and FreeSurfer-based features contain different useful information for classification of converters/non-converters, and they are complementary to improve the performance of classifier.

Different from the two methods used in (Moradi et al., 2015) and (Tong et al., 2017), which directly used voxels as features, the proposed method employs CNN to learn the deep features from the morphology of hippopotamus, and combined CNN-based features with the globe morphology features that were computed by FreeSurfer. We believe that the learnt CNN features might be more meaningful and more discriminative than voxels. When comparing with these two methods, only MRI data was used, but the performances of these two methods were improved when combined MRI data with age and cognitive measures, so investigating the combination of the propose approach with other modality data for performance improvement is also one of our future works.

We have also listed several deep learning-based studies in recent years for comparison in Table 8. Most of them have an accuracy of predicting conversion above 70%, especially the last three approaches (including the proposed one) have the accuracy above 80%. The best accuracy was achieved by (Lu et al., 2018a), which uses both MRI and PET data. However, when only MRI data is used, Lu's method declined the accuracy to 75.44%. Although an accuracy of 82.51% was also obtained with PET data (Lu et al., 2018b), PET scanning usually suffers from contrast agents and more expensive cost than the routine MRI. In summary, our approach achieved the best performance when only MRI images were used and is expected to be improved by incorporating other modality data, e.g. PET, in the future.

In this work, the period of predicting conversion was set to 3 years, that separates MCI subjects into MCI non-converters and MCI converters groups by the criterion who covert to AD within 3 years. But not matter what the period for prediction is, there is a disadvantage that even the classifier precisely predict a MCI non-converters who would not convert to AD within a specific period, but the conversion might still happen half year or even one month later. Modeling the progression of AD and predicting the time of conversion with longitudinal data are more meaningful (Guerrero et al., 2016; Xie et al., 2016). Our future work would investigate the usage of CNN in modeling the progression of AD.

5 Conclusions

In this study, we have developed a framework that only use MRI data to predict the MCI-to-AD conversion, by applying convolutional neural networks (CNN) and other machine learning algorithms. Results show that CNN can extract discriminative features of hippocampus for prediction by learning the morphology changes of hippocampus between AD and NC. And FreeSurfer provides extra structural brain image features to improve the prediction performance as complementary information. Compared with other state-of-the-art methods, the proposed one outperforms others in higher accuracy and AUC, while keeping a good balance between the sensitivity and specificity.

This is a provisional file, not the final typeset article

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Funding

This work was partially supported by National Key R&D Program of China (2017YFC0108703), National Natural Science Foundation of China (61871341, 61571380, 61811530021, 61672335 and 61601276), Natural Science Foundation of Fujian Province of China (2018J06018, 2016J05205 and 2016J05157), Science and Technology Program of Xiamen (3502Z20183053), Fundamental Research Funds for the Central Universities (20720180056), and the Foundation of Educational and Scientific Research Projects for Young and Middle-aged Teachers of Fujian Province (JAT160074 and JAT170406).

8 Acknowledgments

Data used in the preparation of this paper were obtained from the ADNI database[5]. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at the website[6].

9 References

Avci, E., and Turkoglu, I. (2009). An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. *Expert Systems with Applications* 36(2)**,** 2873-2878.

Babaoğlu, I., Fındık, O., and Bayrak, M. (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. *Expert Systems with Applications* 37(3)**,** 2182-2185.

Beheshti, I., Demirel, H., Matsuda, H., and Initiative, A.s.D.N. (2017). Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Computers in Biology and Medicine* 83**,** 109-119.

Burns, A., and Iliffe, S. (2009). Alzheimer's disease. *BMJ* 338**,** b158. doi: 10.1136/bmj.b158.

Cao, P., Shan, X., Zhao, D., Huang, M., and Zaiane, O. (2017). Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease. *Pattern Recognition* 72**,** 219-235.

Cheng, B., Liu, M., Zhang, D., Munsell, B.C., and Shen, D. (2015). Domain transfer learning for MCI conversion prediction. *IEEE Transactions on Biomedical Engineering* 62(7)**,** 1805-1817.

Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Collins, D.L., and Initiative, A.s.d.N. (2012). Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease. *NeuroImage* 59(4)**,** 3736-3747.

Coupe, P., Eskildsen, S.F., Manjon, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., et al. (2012). Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage: Clinical* 1(1)**,** 141-152. doi: 10.1016/j.nicl.2012.10.002.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56(2)**,** 766-781.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31(3)**,** 968-980.

Du, X., Qu, X., He, Y., and Guo, D. (2018). Single image super-resolution based on multi-scale competitive convolutional neural network. *Sensors* 18(3)**,** 789.

[5] www.loni.ucla.edu/ADNI

[6] www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf

Preprint

Dukart, J., Schroeter, M.L., Mueller, K., and Alzheimer's Disease Neuroimaging, I. (2011). Age correction in dementia--matching to a healthy brain. *PLoS One* 6(7), e22193. doi: 10.1371/journal.pone.0022193.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics* 32(2), 407-499.

Eskildsen, S.F., Coupe, P., Fonov, V.S., Pruessner, J.C., Collins, D.L., and Alzheimer's Disease Neuroimaging, I. (2015). Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiology of Aging* 36 Suppl 1, S23-31. doi: 10.1016/j.neurobiolaging.2014.04.034.

Fischl, B., and Dale, A.M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* 97(20), 11050-11055. doi: 10.1073/pnas.200033797.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex* 14(1), 11-22.

Giorgio, A., Santelli, L., Tomassini, V., Bosnell, R., Smith, S., De Stefano, N., et al. (2010). Age-related changes in grey and white matter structure throughout adulthood. *Neuroimage* 51(3), 943-951. doi: 10.1016/j.neuroimage.2010.03.004.

Grundman, M., Petersen, R.C., Ferris, S.H., Thomas, R.G., Aisen, P.S., Bennett, D.A., et al. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology* 61(1), 59-66.

Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., et al. (2016). Instantiated mixed effects modeling of Alzheimer's disease markers. *Neuroimage* 142, 113-125. doi: 10.1016/j.neuroimage.2016.06.049.

Guerrero, R., Wolz, R., Rao, A., Rueckert, D., and Initiative, A.s.D.N. (2014). Manifold population modeling as a neuro-imaging biomarker: application to ADNI and ADNI-GO. *NeuroImage* 94, 275-286.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22), 2402-2410. doi: 10.1001/jama.2016.17216.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., et al. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32(1), 180-194. doi: 10.1016/j.neuroimage.2006.02.051.

Huang, G.B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics* 42(2), 513-529. doi: 10.1109/TSMCB.2011.2168604.

Karas, G.B., Scheltens, P., Rombouts, S.A., Visser, P.J., van Schijndel, R.A., Fox, N.C., et al. (2004). Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 23(2), 708-716. doi: 10.1016/j.neuroimage.2004.07.006.

Kukreja, S.L., Löfberg, J., and Brenner, M.J. (2006). A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC Proceedings Volumes* 39(1), 814-819.

Ledig, C., Heckemann, R.A., Hammers, A., Lopez, J.C., Newcombe, V.F., Makropoulos, A., et al. (2015). Robust whole-brain segmentation: application to traumatic brain injury. *Medical Image Analysis* 21(1), 40-58.

Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., et al. (2011). Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. *Neuroimage* 55(3), 1091-1108.

Li, F., Tran, L., Thung, K.H., Ji, S., Shen, D., and Li, J. (2015). A robust deep model for improved classification of AD/MCI patients. *IEEE Journal of Biomedical and Health Informatics* 19(5), 1610-1616. doi: 10.1109/JBHI.2015.2429556.

Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., et al. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Transactions on Biomedical Engineering* 62(4), 1132-1140.

Lu, D., Popuri, K., Ding, G.W., Balachandar, R., and Beg, M.F. (2018a). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Scientific Reports* 8(1), 5697.

Lu, D., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F., and Initiative, A.s.D.N. (2018b). Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Medical Image Analysis* 46, 26-34.

Markesbery, W.R. (2010). Neuropathologic alterations in mild cognitive impairment: a review. *Journal of Alzheimer's disease* 19(1), 221-228. doi: 10.3233/JAD-2010-1220.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and Alzheimer's Disease Neuroimaging, I. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104, 398-412. doi: 10.1016/j.neuroimage.2014.10.002.

Nie, D., Wang, L., Gao, Y., and Shen, D. (2016). "Fully convolutional networks for multi-modality isointense infant brain image segmentation", in: *Proc. IEEE Int. Symp. Biomed. Imaging*, 1342-1345.

Nyul, L.G., and Udupa, J.K. (1999). On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine* 42(6), 1072-1081.

Provisional

This is a provisional file, not the final typeset article

Ortiz, A., Munilla, J., Gorriz, J.M., and Ramirez, J. (2016). Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *International Journal of Neural Systems* 26(07), 1650025.

Rajkomar, A., Lingam, S., Taylor, A.G., Blum, M., and Mongan, J. (2017). High-throughput classification of radiographs using deep convolutional neural networks. *Journal of Digital Imaging* 30(1)**,** 95-101.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., and Hawkes, D.J. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8)**,** 712-721.

Shi, J., Zheng, X., Li, Y., Zhang, Q., and Ying, S. (2018). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics* 22(1), 173-183.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging* 35(5)**,** 1285-1298. doi: 10.1109/TMI.2016.2528162.

Singh, S., Srivastava, A., Mi, L., Caselli, R.J., Chen, K., Goradia, D., et al. (2017). "Deep-learning-based classification of FDG-PET data for Alzheimer's disease categories", in: *13th International Conference on Medical Information Processing and Analysis*, (International Society for Optics and Photonics), 105720J.

Suk, H.I., Lee, S.W., Shen, D., and Alzheimer's Disease Neuroimaging, I. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage* 101**,** 569-582. doi: 10.1016/j.neuroimage.2014.06.077.

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., et al. (2017). A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Transactions on Biomedical Engineering* 64(1)**,** 155-165. doi: 10.1109/TBME.2016.2549363.

Tong, T., Wolz, R., Gao, Q., Hajnal, J.V., and Rueckert, D. (2013). Multiple instance learning for classification of dementia in brain MRI. *Medical Image Analysis* 16(Pt 2)**,** 599-606.

Usai, M.G., Goddard, M.E., and Hayes, B.J. (2009). LASSO with cross-validation for genomic selection. *Genetics Research* 91(6)**,** 427-436. doi: 10.1017/S0016672309990334.

van de Pol, L.A., Hensel, A., van der Flier, W.M., Visser, P.J., Pijnenburg, Y.A., Barkhof, F., et al. (2006). Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry* 77(4)**,** 439-442. doi: 10.1136/jnnp.2005.075341.

Wee, C.Y., Yap, P.T., and Shen, D. (2013). Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping* 34(12)**,** 3411-3425.

Wu, P.H., Chen, C.C., Ding, J.J., Hsu, C.Y., and Huang, Y.W. (2013). Salient region detection improved by principle component analysis and boundary information. *IEEE Transactions on Image Processing* 22(9)**,** 3614-3624. doi: 10.1109/TIP.2013.2266099.

Wyman, B.T., Harvey, D.J., Crawford, K., Bernstein, M.A., Carmichael, O., Cole, P.E., et al. (2013). Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's & Dementia* 9(3)**,** 332-337. doi: 10.1016/j.jalz.2012.06.004.

Xie, Q., Wang, S., Zhu, J., Zhang, X., and Initiative, A.s.D.N. (2016). Modeling and predicting AD progression by regression analysis of sequential clinical data. *Neurocomputing* 195**,** 50-55.

Yamada, M., Jitkrittum, W., Sigal, L., Xing, E.P., and Sugiyama, M. (2014). High-dimensional feature selection by feature-wise kernelized Lasso. *Neural Computation* 26(1)**,** 185-207. doi: 10.1162/NECO_a_00537.

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., et al. (2012). Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurology* 12(1)**,** 46.

Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S., et al. (2013). Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *NeuroImage: Clinical* 2**,** 735-745. doi: 10.1016/j.nicl.2013.05.004.

Zeng, N., Wang, Z., Zhang, H., Liu, W., and Alsaadi, F.E. (2016). Deep belief networks for quantitative analysis of a gold immunochromatographic strip. *Cognitive Computation* 8(4), 684-692.

Zeng, N., Zhang, H., Liu, W., Liang, J., and Alsaadi, F.E. (2017). A switching delayed PSO optimized extreme learning machine for short-term load forecasting. *Neurocomputing* 240, 175-182.

Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., and Dobaie, A.M. (2018). Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* 273, 643-649.

Zhang, D., Shen, D., and Initiative, A.s.D.N. (2012). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PloS one* 7(3), e33182.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., and Alzheimer's Disease Neuroimaging, I. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55(3), 856-867. doi: 10.1016/j.neuroimage.2011.01.008.

Provisional

Table 1
The demographic information of the dataset used in this work

	AD	NC	MCIc	MCInc	MCIun
Subjects' number	188	229	169	139	93
Age range	55-91	60-90	55-88	55-88	55-89
Males/Females	99/89	119/110	102/67	96/43	60/33

Note: MCIc means MCI converters. MCInc means MCI non-converters, MCIun means MCI unknown.

Table 2
The performance of the 2.5D CNN

	Classifying: AD/NC Trained with: AD/NC	Classifying: MCIc/MCInc Trained with: MCIc/MCInc	Classifying: MCIc/MCInc Trained with: AD/NC	Different patch Sampling
Accuracy	88.79%	68.68%	73.04%	72.75 %
Standard deviation	0.61%	1.63%	1.31%	1.20 %
Confidence interval	[0.8862, 0.8897]	[0.6821, 0.6914]	[0.7265, 0.7343]	[0.7252, 0.7299]

Note: MCIc means MCI converters. MCInc means MCI non-converters. The results were obtained with ten-fold cross validations, and averaged over 50 runs.

Table 3
The performance of different features used, and the performance without age correction.

Method	Accuracy	Sensitivity	Specificity	AUC
Proposed method (both features)	**79.9%**	**84%**	**74.8%**	**86.1%**
Only CNN-based features	76.9%	81.7%	71.2%	82.9%
Only FreeSurfer-based features	76.9%	82.2%	70.5%	82.8%
Without age correction	75.3%	79.9%	69.8%	82.6%

Table 4
Comparison of extreme learning machine with other two classifiers.

Method	Accuracy	Sensitivity	Specificity	AUC
SVM	**79.87%**	83.43 %	**75.54 %**	83.85**%**
Random forest	75.0%	82.84 %	65.47 %	81.99 %
Extreme learning machine	**79. 87%**	**84.02 %**	74.82%	**86.14 %**

Note: Implementation of SVM was performed using third party library LIBSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/), and the random forest was utilized with the third party library (http://code.google.com/p/randomforest-matlab). Both classifiers used the default settings.

Table 5
Comparison with others methods on the same dataset in ten-fold cross validation.

Method	Accuracy	Sensitivity	Specificity	AUC
MRI biomarker in (Moradi et al., 2015)	74.7%	**88.9%**	51.6%	76.6%
Global grading biomarker in (Tong et al., 2017)	78.9%	76.0%	**82.9%**	81.3%
Proposed method	**79.5%**	86.1%	68.8%	**83.6%**

Provisional

This is a provisional file, not the final typeset article

Note: The performances of MRI biomarker and global grading biomarker are described in (Moradi et al., 2015) and (Tong et al., 2017). The results are averages over 100 runs, and the standard deviation/confidence intervals of accuracy and AUC of the proposed method are 1.19% / [0.7922, 0.7968] and 0.83% / [0.8358, 0.8391].

Table 6
Comparison with others methods on the same dataset in leave-one-out cross validation.

Method	Accuracy	Sensitivity	Specificity	AUC
MRI biomarker in (Moradi et al., 2015)	-	-	-	-
Global grading biomarker in (Tong et al., 2017)	78.8%	76.2%	**83%**	81.2%
Proposed method	**81.4%**	**89.6%**	68%	**87.8%**

Note: The global grading biomarkers was download from the web described in (Tong et al., 2017) and the experiment was performed with same method as in (Tong et al., 2017).

Table 7
The 15 most informative FreeSurfer-based features for predicting MCI-to-AD conversion

Number	FreeSurfer-based feature
1	Cortical Thickness Average of Left FrontalPole
2	Volume (Cortical Parcellation) of Left Precentral
3	Volume (Cortical Parcellation) of Right Postcentral
4	Volume (WM Parcellation) of Left AccumbensArea
5	Cortical Thickness Average of Right CaudalMiddleFrontal
6	Cortical Thickness Average of Right FrontalPole
7	Volume (Cortical Parcellation) of Left Bankssts
8	Volume (Cortical Parcellation) of Left PosteriorCingulate
9	Volume (Cortical Parcellation) of Left Insula
10	Cortical Thickness Average of Left SuperiorTemporal
11	Cortical Thickness Standard Deviation of Left PosteriorCingulate
12	Volume (Cortical Parcellation) of Left Precuneus
13	Volume (WM Parcellation) of CorpusCallosumMidPosterior
14	Volume (Cortical Parcellation) of Left Lingual
15	Cortical Thickness Standard Deviation of Right Postcentral

Provisional

Table 8
Results of previous deep learning based approaches for predicting MCI-to-AD conversion

Study	Number of MCIc/MCInc	Data	Conversion time	Accuracy	AUC
(Li et al., 2015)	99/56	MRI+PET	18 months	57.4%	-
(Singh et al., 2017)	158/178	PET	-	72.47%	-
(Ortiz et al., 2016)	39/64	MRI+PET	24 months	78%	82%
(Suk et al., 2014)	76/128	MRI+PET	-	75.92%	74.66%
(Shi et al., 2018)	99/56	MRI+PET	18 months	78.88%	80.1%
(Lu et al., 2018a)	217/409	MRI+PET	36 months	**82.93%**	-
(Lu et al., 2018a)	217/409	MRI	36 months	75.44%	-
(Lu et al., 2018b)	112/409	PET	-	82.51%	-
This study	164/100	MRI	36 months	81.4%	**87.8%**

Note: MCIc means MCI converters. MCInc means MCI non-converters. Different subjects and modalities of data are used in these approaches. All the criteria are copied from the original literatures.

Figure 1
Framework of proposed approach. The dashed arrow indicates the CNN was trained with 2.5D patches of NC and AD subjects. The dashed box indicates Leave-one-out cross validation was performed by repeat LASSO and extreme learning machine 308 times, in each time one different MCI subject was leaved for test, and the other subjects with their labels were used to train LASSO and extreme learning machine.

Figure 2
The demonstration of 2.5D patch extraction from hippocampus region. (A, B, C): 2D patches extracted from transverse (red box), coronal (green box) and sagittal (blue box) plane; (D): The 2.5D patch with three patches at their spatial locations, red dot is the center of 2.5D patch; (E): Three patches are combined into RGB patch as red (red box patch), green (green box patch) and blue (blue box patch) channels.

Figure 3
(A) Four random chosen 2.5D patches of one subject (who is normal control, female and 76.3 years old), indicating that these patches contain different information of hippocampus; (B) The comparison of correspond 2.5D patches of four subjects from four groups, the different level of hippocampus atrophy can be found.

Figure 4
The overall architecture of the CNN used in this work.

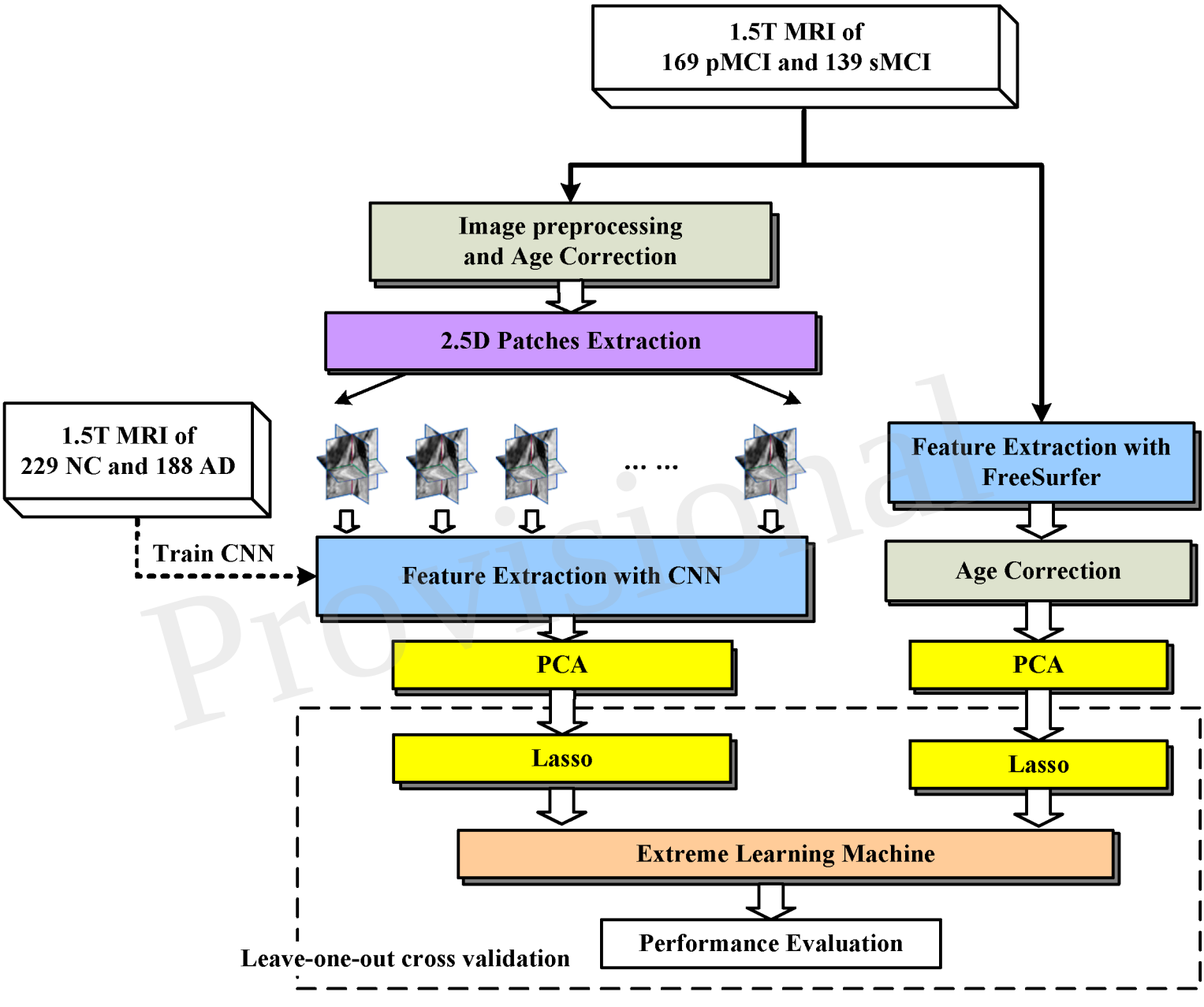
Figure 5
The workflow of extracting CNN-based features. The CNN was trained with all AD/NC patches, and used to extract deep features from all 151 patches of MCI subject. The feature number of each patch is reduced to P_C (P_C=29) from 1024 by PCA. Finally, Lasso selects L_C (L_C=35) features from P_C×151 features for each MCI subject.

Figure 6
The ROC curves of classifying converters/non-converters when different features used or without age correction.

Provisional

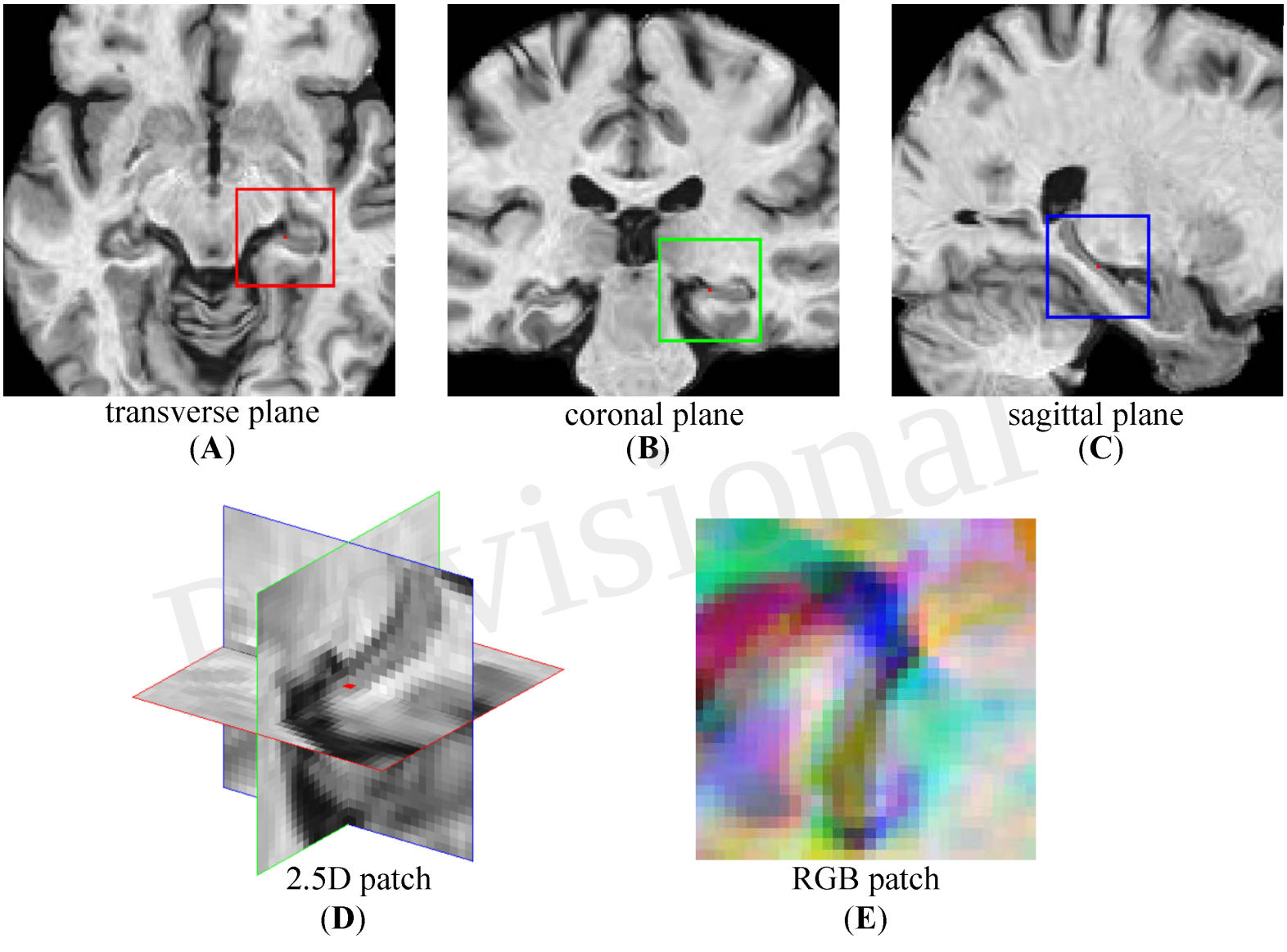
This is a provisional file, not the final typeset article

Figure 01.TIF



Provisional

Figure 02.TIF



Provisional

Figure 03.TIF

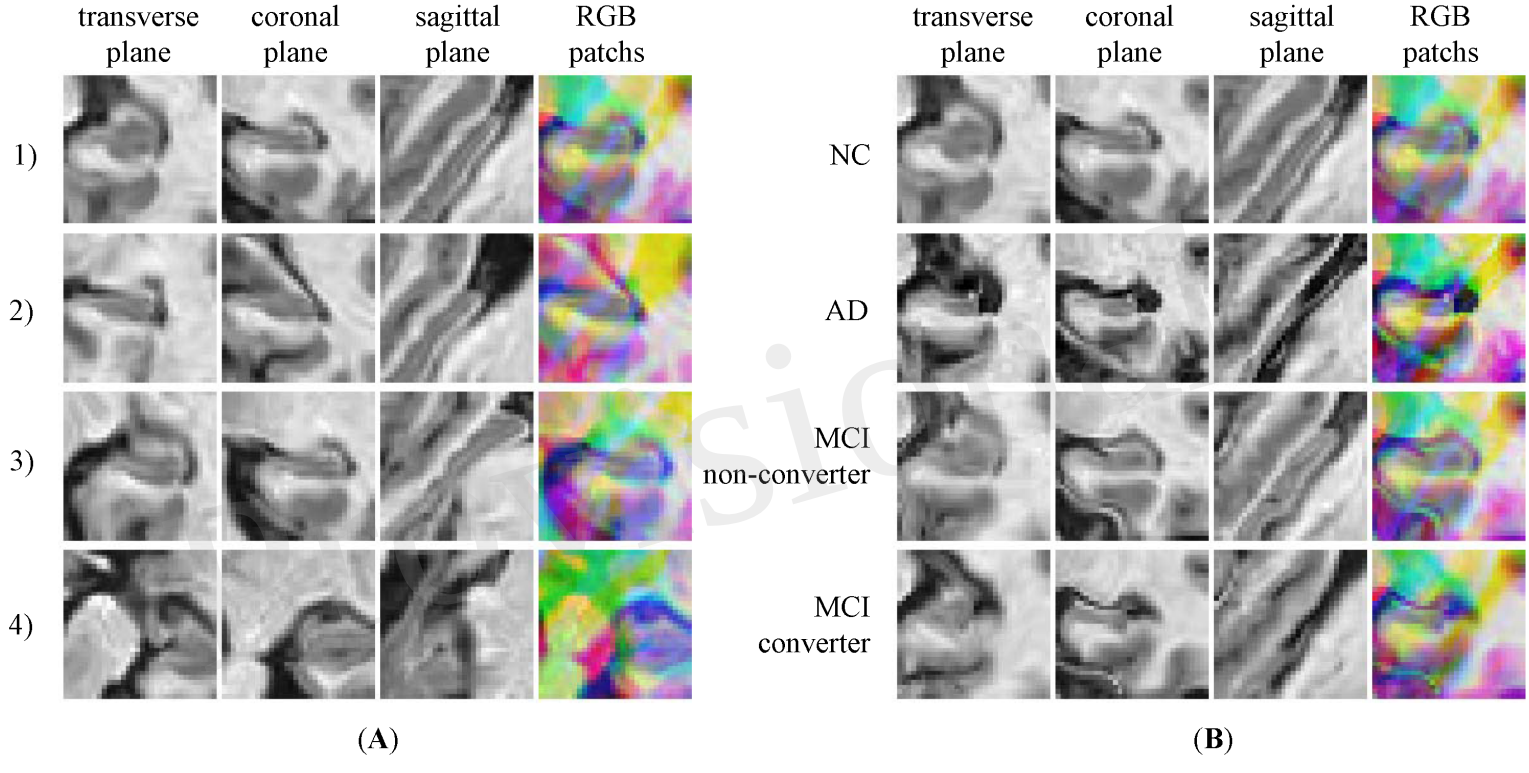
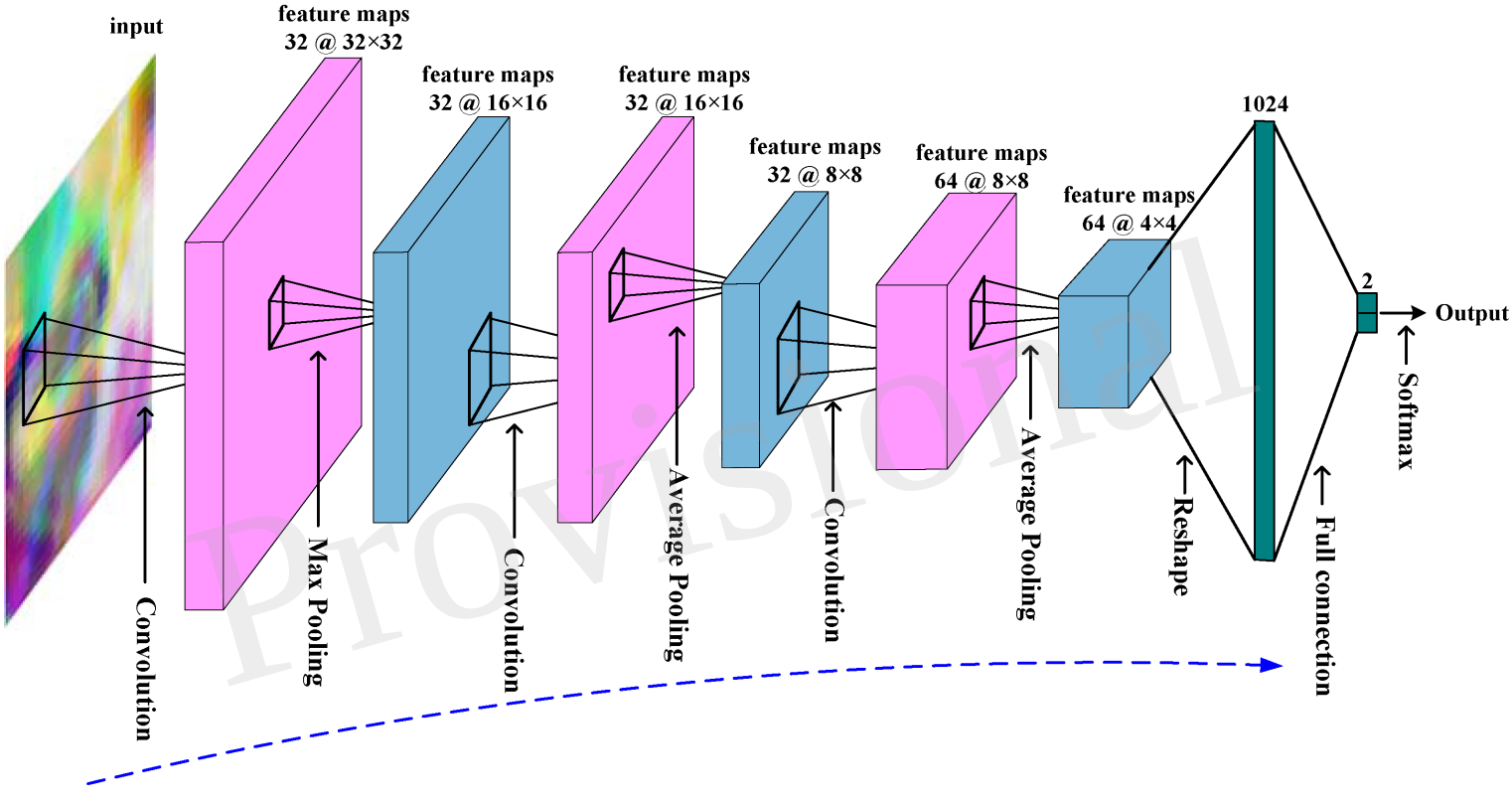
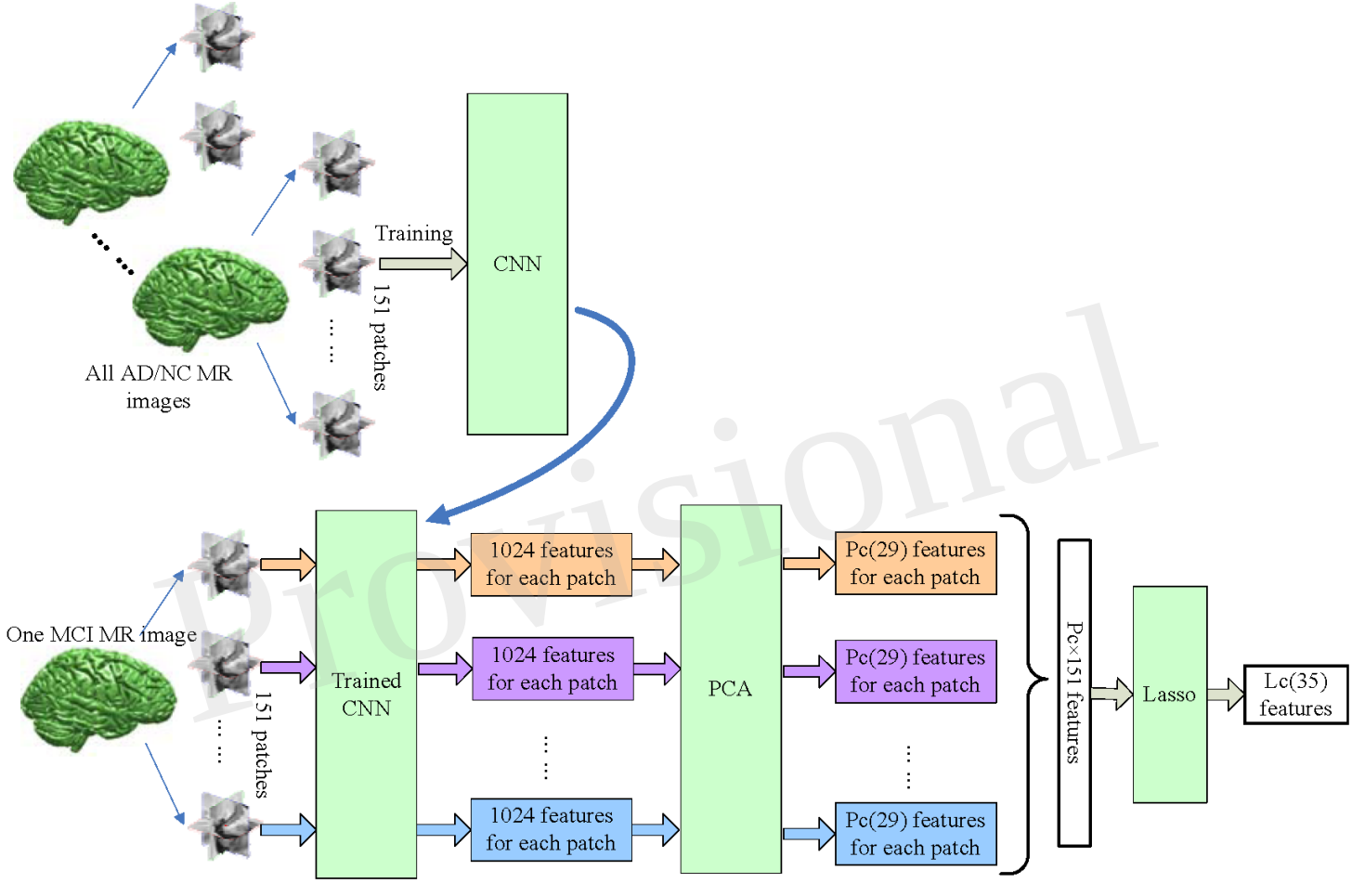


Figure 04.TIF



Provisional

Figure 05.TIFF



Provisional

Figure 06.TIFF