[Frontiers In Bioscience, Landmark, 23, 584-596, January 1, 2018]

# Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network

**Ehsan Hosseini Asl[1], Mohammed Ghazal[2,3], Ali Mahmoud[2], Ali Aslantas[2], Ahmed Shalaby[2], Manual Casanova[4], Gregory Barnes[5], Georgy Gimel'farb[6], Robert Keynton[2], Ayman El Baz[2]**

*[1]Department of Electrical and Computer Engineering, University of Louisville, Louisville, KY, [2]Department of Bioengineering, University of Louisville, Louisville, KY, [3]Department of Electrical and Computer Engineering, Abu Dhabi University, UAE, [4]Department of Pediatrics, University of South Carolina, SC, USA, [5]Department of Neurology, University of Louisville, USA, [6]Department of Computer Science, University of Auckland, New Zealand.*

## TABLE OF CONTENTS



## 1. ABSTRACT

Early diagnosis is playing an important role in preventing progress of the Alzheimer's disease (AD). This paper proposes to improve the prediction of AD with a deep 3D Convolutional Neural Network (3D-CNN), which can show generic features capturing AD biomarkers extracted from brain images, adapt to different domain datasets, and accurately classify subjects with improved fine-tuning method. The 3D-CNN is built upon a convolutional autoencoder, which is pre-trained to capture anatomical shape variations in structural brain MRI scans for source domain. Fully connected upper layers of the 3D-CNN are then fine-tuned for each task-specific AD classification in target domain. In this paper, deep supervision algorithm is used to improve the performance of already proposed 3D Adaptive CNN. Experiments on the ADNI MRI dataset without skull-stripping preprocessing have shown that the proposed 3D Deeply Supervised Adaptable CNN outperforms several proposed approaches, including 3D-CNN model, other CNN-based methods and conventional classifiers by accuracy and robustness. Abilities of the proposed network to generalize the features learnt and adapt to other domains have been validated on the CADDementia dataset.

## 2. INTRODUCTION

The Alzheimer's disease (AD), a progressive brain disorder and the most common case of dementia in the late life, causes the death of nerve cells and tissue loss throughout the brain, thus reducing the brain volume dramatically through time and affecting most of its functions (1). The estimated number of affected people will double in the next two decades, so that one out of 85 persons will have the AD by 2050 (2). Because the cost of care for the AD patients is expected to rise dramatically, the necessity of having a computer-aided system for early and accurate AD diagnosis becomes critical (3).

This paper focuses on developing an adaptable deep learning-based system for early diagnosis of the AD. Deep learning helps to solve such a complex diagnostic problem by leveraging hierarchical extraction of input data features to improve classification (4). Several layers of feature extractors are trained to form a model being able to adapt to a new domain by transferring knowledge between different domains and learning distinctive properties of the new data (5), (6). It has been shown that trained features turn from generality to specificity through layers of a deep network (7), which relates

to transferability of features. A robust diagnosis model of a particular disease should be adaptable to various datasets, e.g., MRI scans collected by several patient groups, as to diminish discrepancies in data distributions and biases toward specific groups. Deep learning aims to decrease the use of domain expert knowledge in designing and extracting most appropriate discriminative features (4).

## 3. MATERIALS AND METHODS

The proposed AD diagnostic framework extracts features of a brain MRI with a source-domain-trained 3D-CAE and performs task-specific classification with a deeply supervised target-domain-adaptable CNN (3D-DSA-CNN). The proposed algorithm for AD classification comprises three steps: (i) spatially normalizing brain sMRI and removing the skull on source-domain data (Figure 1); (ii) extracting feature vector by training 3D-CAE on skull-stripped source domain data; (iii) training a 3D-DSA-CNN model on target-domain data for AD diagnosis. The mathematical details of the last two steps are detailed below.

### 3.1. Data collection

Data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), Positron Emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment (MCI) and early Alzheimer's disease (AD). Performance of the proposed 3D-DSA-CNN for AD diagnosis was evaluated on 30 subjects of CADDementia, as source domain, and 210 randomly selected subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, as target domain (demographic information mentioned in Table 1), for five classification tasks: four binary ones (AD vs. NC, AD+MCI vs. NC, AD vs. MCI, MCI vs NC), and three-way classification (AD vs. MCI vs. NC). The CADDementia data set contains structural T1-weighted MRI (T1w) scans of patients with the diagnosis of probable AD, patients with the diagnosis of MCI, and NC without a dementia syndrome (3).

### 3.2. 3D Convolutional Autoencoder (3D-CAE)

Conventional unsupervised autoencoder extracts a few co-aligned scalar feature maps for a set of input 3D images with scalar or vectorial voxel-wise signals by combining data encoding and decoding. The input image is encoded by mapping each fixed voxel neighborhood to a vectorial feature space in the hidden layer and is reconstructed back in the output layer to the original image space. To extract features that capture characteristic patterns of input data variations, training of the autoencoder employs back-propagation and constraints on properties of the feature space to reduce the reconstruction error.

Extracting global features from 3D images with vectorial voxel-wise signals is computationally expensive and requires too large training data sets. This is due to growing fast numbers of parameters to be evaluated in the input (encoding) and output (decoding) layers (32), (33). Moreover, although autoencoders with full connections between all nodes of the layers try to learn global features, local features are more suitable for extracting patterns from high-dimensional images. To overcome this problem, we use a stack of unsupervised CAE with locally connected nodes and shared convolutional weights to extract local features from 3D images with possibly long voxel-wise signal vectors (34)–(36). Each input image is reduced hierarchically using the hidden feature (activation) map of each CAE for training the next-layer of CAE.

The 3D extension of a hierarchical CAE proposed in (34) is shown in Figure 2. To capture the characteristic variations of a 3D image, x, each voxel-wise feature, $h_{i:j:k}$ associated with the $i$-th 3D lattice node, $j$-th component of the input voxel-wise signal vector, and $k$-th feature map; $k = [1, \ldots, K]$, is extracted by a moving-window convolution (denoted below *) of a fixed $n \times n \times n$ neighborhood, $X_{i:neib}$ of this node with a linear encoding filter specified by its weights, $Wk = [W_{j:k} : j = 1, \ldots, J]$ for each relative neighboring location with respect to the node $i$ and each voxel-wise signal component $j$, followed by feature-specific biases, $[b_{j:k} : j = 1, \ldots, J]$ and nonlinear transformations with a certain activation function, $f(.)$

$$h_{i:j:k} = f\,(W_k * X_{i:neib} + b_{j:k}) \qquad (1)$$

The latter function is selected from a rich set of constraining differentiable functions, in particular, the sigmoid, $f(u) = (1 + \exp(-u))^{-1}$ and rectified linear unit (ReLU), $f(u) = max(0,u)$ (37). Since the 3D image x in Eq (1) has the $J$-vectorial voxel-wise signals, the weights $\mathbf{W}_k$ define a 3D moving-window filter convolving the union of J-dimensional signal spaces for each voxel within the window. To simplify the notation, let $h_k = T(x: W_k, b_k, f(.))$ denotes the entire encoding of the input 3D image with $J$-vectorial voxel-wise signals with the $k$-th 3D feature map, $h_k$, such that its scalar components are obtained with Eq. (1) using the weights $\mathbf{W}_k$ and bias vectors $\mathbf{b}_k$ for a given voxel neighborhood. The similar inverse transformation, $\mathbf{T}_{inv}(\ldots)$, with the same voxel neighborhood, but generally with the different convolutional weights, $\mathbf{P}_k$,

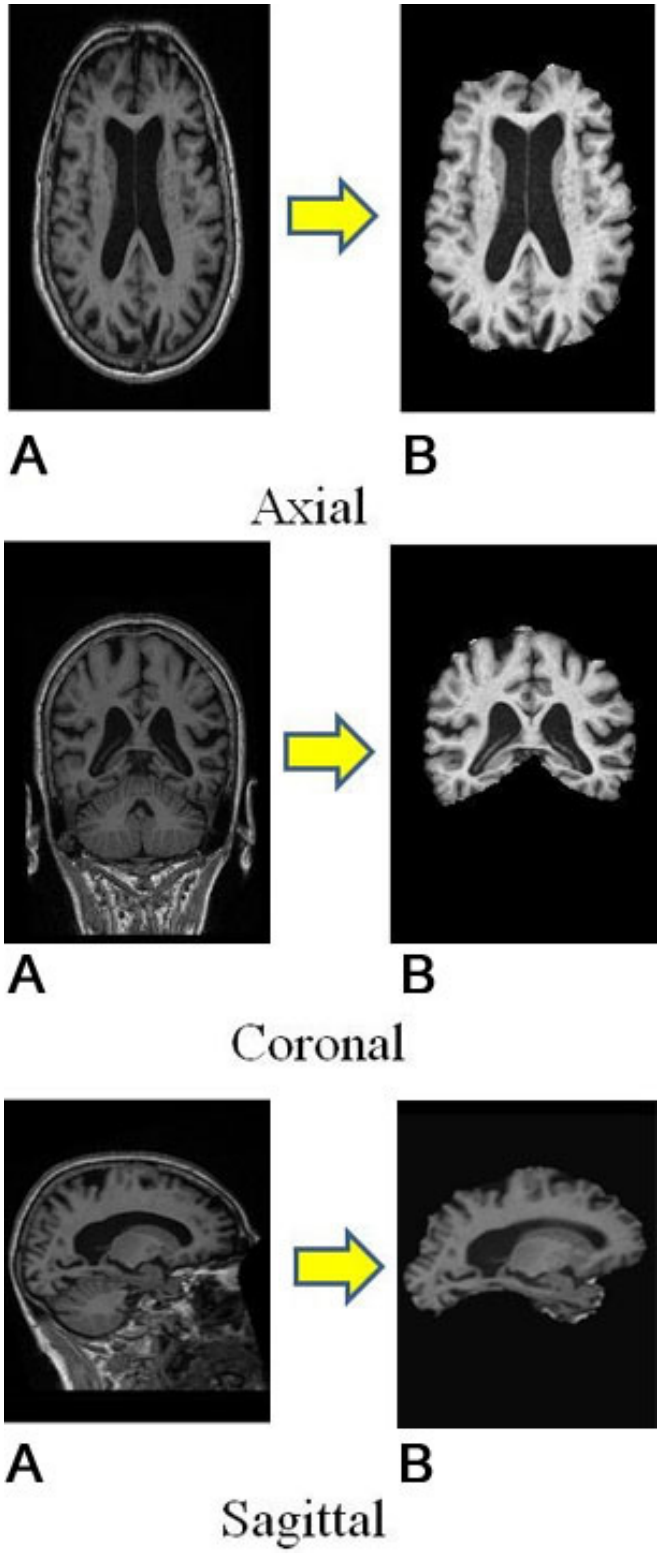


**Figure 1.** Brain sMRI of CADDementia dataset (a) before and (b) after preprocessing depicted in axial, coronal, and sagittal view, by spatially normalizing and removing skull based on mutual information-based rigid registration approach (31). Note that the image intensity is normalized to (0,1) after removing the skull, and the image looks more bright than before preprocessing.

**Table 1.** Demographic data for 210 subjects from the target domain, ADNI database (STD – standard deviation)

| Diagnosis | AD | MCI | NC |
|---|---|---|---|
| Number of subjects | 70 | 70 | 70 |
| Male / Female | 36 / 34 | 50 / 20 | 37 / 33 |
| Age (mean±STD ) | 75.0.±7.9 | 75.9.±7.7 | 74.6.±6.1 |

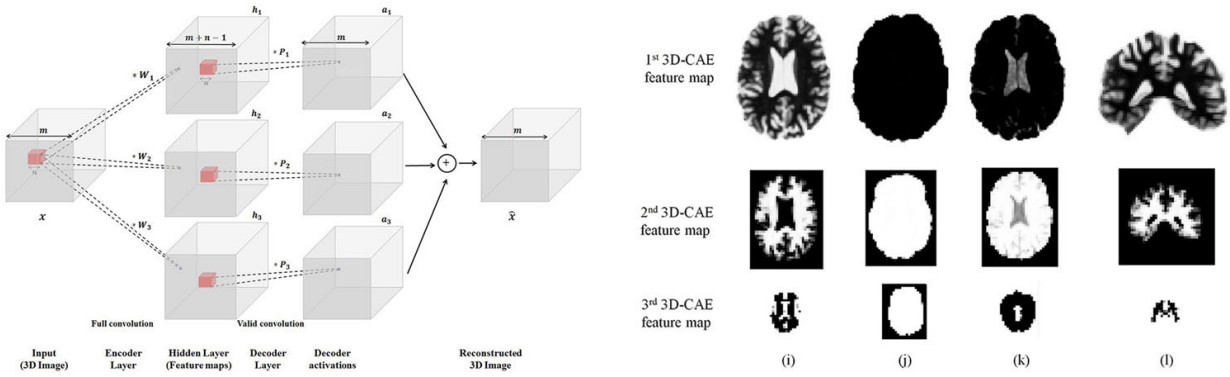


**Figure 2.** Schematic diagram of a 3D-CAE (left) for extracting generic features by convolving and pooling an input 3D image (the encoding feature maps are of a larger size, whereas the decoded image keeps the original size), and axial (i,j,k) and sagittal (l) Slices (right) of hierarchical 3D feature maps extracted from the source domain, CADDementia brain sMRI at three layers of the stacked 3D-CAE: cortex thickness and volume (i), brain size (j), ventricle size (k), and hippocampus model (l). The feature maps are down-sampled at each layer by max-pooling to reduce their size and detect the higher-level features.

biases, $\mathbf{b}_{inv:k}$, and, possibly, activation function, $g(.)$, decodes, or reconstructs the initial 3D image:

$$\hat{X} = \sum_{k=1}^{K} \underbrace{T_{inv}(h_k : P_k, b_{inv:k}, g(.))}_{a_k} \quad (2)$$

Given $L$ encoding layers, each layer l generates an output feature image, $h(l) = [h_{(l):k} : k = 1, \ldots, K_l]$, with $K_l$- vectorial voxel-wise features and receives the preceding output, $h(l) = [h_{(l-1):k} : k = 1, \ldots, K_{l-1}]$ as the input image (i.e., h(0) = x).

The 3D-CAE of Eqs. (1) and (2) is trained by minimizing the mean squared reconstruction error for $T; T \geq 1$ given training input images, $x^{[t]}$; $t = 1,\ldots,T$,

$$E(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{x} - x^{[t]} \right\|_2^2 \quad (3)$$

where $\theta = [w_k; p_k; b_k; b_{inv:k} : k = 1,\ldots,K]$ and $\|\ldots\|_2^2$ denote all free parameters and the average vectorial $l_2$-norm over the $T$ training images, respectively. To reduce the number of free parameters, the decoding weights $\mathbf{P}_k$ and encoding weights $\mathbf{W}_k$ were tied by flipping over all their dimensions as proposed in (34). The cost function Eq. (3) was minimized in the parameter space by using the stochastic gradient descent search which computed by error back-propagation.

In order to obtain translational invariance, the feature maps, $\mathbf{h}_{(j)}$, are down-sampled by max-pooling, i.e., extracting the maximum value of non-overlapping sub-regions. For entangling shape variations in the higher-level feature maps of reduced size, the max-pooling output is used for training the higher layer CAE, as shown in Figure 2(a). Stacking the 3D-CAE's encoding layers (abbreviated 3D-CAES below), known as greedy layer-wise training (38), halves the size of the feature map at each level of their hierarchy (34).

### 3.3. Transfer Learning and Domain Adaptation

To achieve good performance, supervised learning of a classifier often requires a large training set of labeled data. If this set is, in principle, of a too limited size, additional knowledge from building a similar classifier can be involved via so-called transfer learning. In particular, the goal classifier based on a deep CNN might employs initial weights, been learned for solving similar task (39)–(42).

We focus on domain adaptation (43)–(45), or source-to-target adaptation, when a trained classifier on a source data, is adapted (fine-tuned) to the

target data. Unlike usual supervised learning, when a classifier is trained from scratch, by minimizing a total quantitative loss from errors on the training data, the domain adaptation minimizes the same loss over the target domain by updating the classifier, which has been trained on the source domain. We leverage the unsupervised feature learning to transfer features, found in the source domain, to the target domain, in order to boost the predictive performance of the deep CNN models (46), (47).

### 3.4. 3D Deeply Supervised Adaptive CNN (3D-DSA-CNN)

While the lower layers of a goal predictive 3D-CNN extract the generalized features, the upper layers have to facilitate task-specific classification using those features (6). The proposed classifier extracts the generalized features by using a stack of locally-connected (convolutional) lower layers, while performing task-specific classification, by fine-tuning the parameters of the upper fully-connected layers. Training the proposed hierarchical 3D-CNN consists of pre-training, initial training of the lower convolutional layers, followed by task-specific fine-tuning. At the pre-training stage, the convolutional layers for generic feature extraction are formed as a stack of 3D-CAEs, which were trained in the source domain. Then these layers are initialized by stacking the encoding layers of the 3D-CAE (5), and finally, the deep-supervision based (14) fine-tuning of the upper fully-connected layers, which is performed for each task-specific binary or multi-class sMRI classification.

Due to the pre-training on the source domain data, the bottom convolutional layers can extract generic features related to the AD biomarkers, such as the ventricular size, hippocampus shape, and cortical thickness, as shown in Figure 2(b). We use the Net2Net initialization (5), which allows for different convolutional kernels and pooling sizes of the 3D-CNN layers, comparing to those in the 3D-CAE, based on the target-domain image size and imaging specifications. This is to facilitate adapting the 3D-CNN across different domains. To classify the extracted features in a task-specific way, weights of the upper fully-connected layers of 3D-CNN are fine-tuned, on the target-domain data, by minimizing a specific loss function. The loss depends explicitly on the weights, and is proportional to a negative log-likelihood (NLL) of the true output classes, given the input features extracted from the target-domain images by the pre-trained bottom layers of the network.

## 4. RESULTS

This section addresses the experimental results of the proposed framework. To pretrain 3D-CAES on CADDementia, sMRI are preprocessed by spatially normalizing using rigid registration approach (31). Then skull is removed and image intensities are normalized to (0, 1), resulting in sMRI of size (200×150×150), as shown in Figure 1. Classification accuracy for each task was evaluated by ten-fold cross-validation. The Theano library (51) was used to develop the deep CNN implemented for our experiments on the Amazon EC2 g2.8.xlarge instances with GPU GRID K520.

### 4.1. Generic and task-specific feature evaluation

Special 2D projections of the extracted features in Figure. 2(b) illustrate generalization and adaptation abilities of the pro- posed 3D-DSA-CNN (Figure. 3). Selected slices of the three feature maps from each layer of stacked 3D-CAE (abbreviated 3D-CAES below) in Figure. 2(b), indicate that the trained generic convolutional filters can capture related features of AD biomarkers, e.g., the ventricle size, cortex thickness, and hippocampus model. These feature maps were generated by the pre-trained 3D-CAES for the CADDementia database.

According to these projections, different convolutional filters of the 3D-CAES can extract the cortex thickness, the brain size (related to the patient gender), ventricle size, and hippocampus model, as the discriminative AD features, in each encoding layers. The 3D Convolutional Autoencoder (3D-CAE) tries to find the variational factors in dataset (CADDementia) without using labels (unsupervised learning). The intuition is that the hidden factorial variation in AD are related to AD biomarkers. Each 3D-CAES layer combines the extracted lower-layer feature maps, in order to train a higher-level feature, describing more detailed anatomical variations of the brain sMRI. Both the ventricle size and cortex thickness features are combined in upper layers, to extract a conceptually higher-level features at the next layers. Visualized in Figure. 4, projections demonstrate the capability of the extracted higher-layer features to separate the AD, MCI, and NC brain sMRI's in the low-dimensional feature space.

Visualization of the manifold distributions of hidden activation of the proposed 3D-DSA-CNN on training ADNI sMRIs, shown in Figure. 4, illustrate the discriminative abilities of the generic and task-specific features. The generic feature-extraction layers (conv1, conv2, and conv3 in Figure. 4(a–c)) gradually enhance the AD, MCI and NC discriminability along their hierarchy. The subsequent task-specific classification layers further enhance the discriminability of these three ADNI classes, as shown in Figure. 4(d–h). The task-specific features are highlighted in Figure. 4(d), depicting the distribution of all three classes when the AD+MCI subjects are to be distinguished from the NC subjects. Obviously, the AD, MCI, and NC cases are

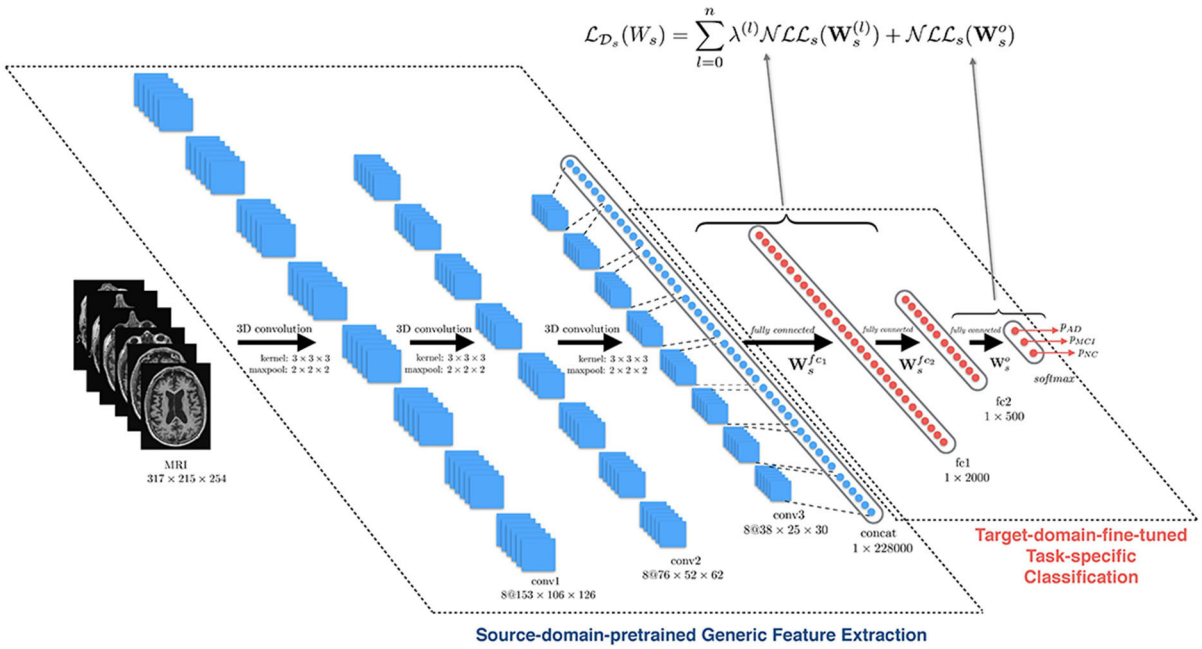


**Figure 3.** Architecture of the deeply supervised and adapTable CNN (3D-DSA-CNN) for AD diagnosis.

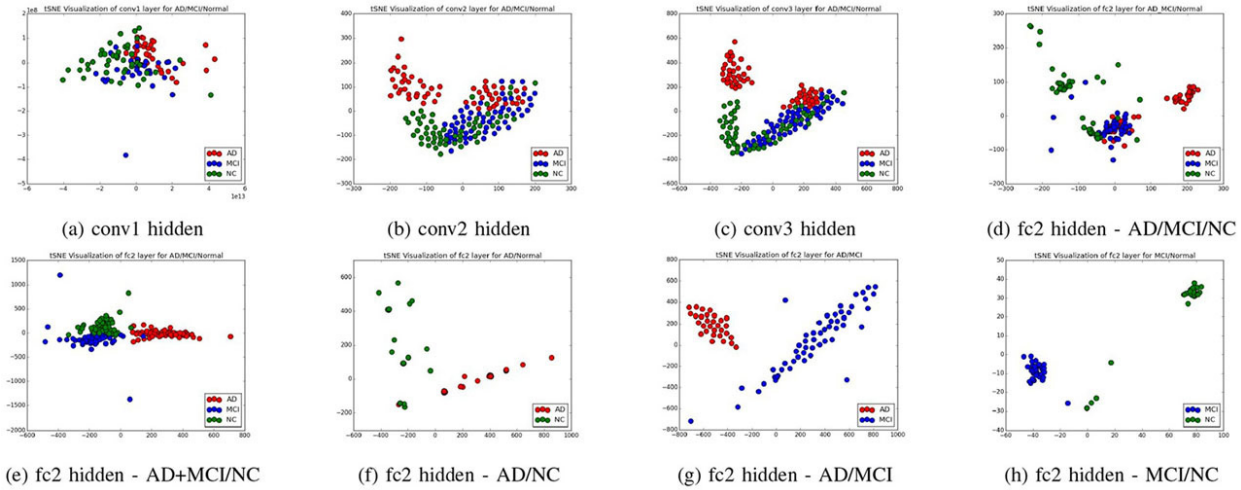


**Figure 4.** Manifold visualization of target domain (ADNI) training data, by t-SNE projection (50), in pre-trained generic (a,b,c) and fine-tuned task-specific (d,e,f,g,h) 3D-DSA-CNN layers.

projected at closer distances and well separated in the task-specific feature space.

The three-class manifold distribution of the test dataset for ternary (AD vs. MCI vs. NC) classification in Figure. 5 indicates a sound discriminative ability of the trained features, to distinguish between these classes. Subjects' distribution in these manifolds indicate the correlation between the disease severity and the extracted features. For example, the most severe AD cases are distributed at the right-most side of the AD manifold, and the most normal (NC) cases are at the bottom-left of the NC manifold.

### 4.2. Classification performance evaluation

Performance of the proposed 3D-DSA-CNN classifier for each specific task, listed in Section 4.1, was evaluated and compared to competing approaches (19), (22)–(26) by using eight evaluation metrics. Let TP, TN, FP, and FN denote, respectively, numbers of true positive, true negative, false positive,

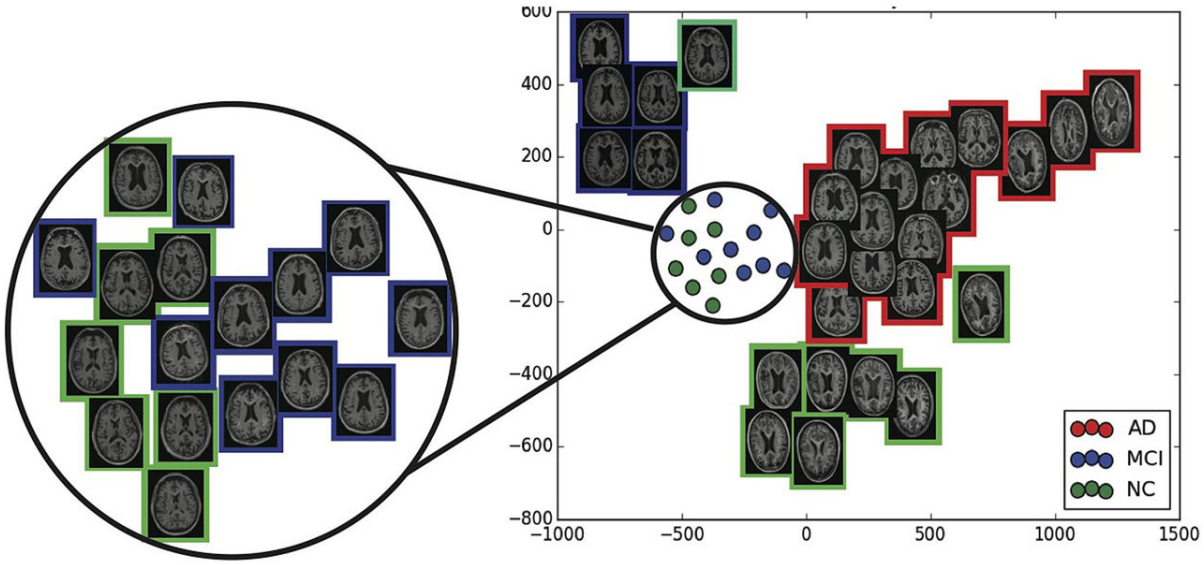


**Figure 5.** Hidden manifold distribution of the target domain (ADNI) test data (two test folds randomly selected for cross-validation) in the fc2 layer, visualized by t-SNE projection (50).

**Table 2.** Task-specific performance of the proposed classifier on target domain (ADNI) for a selected cross-validation fold.

| Class | PPV | SEN | F1 | PPV | SEN | F1 | PPV | SEN | F1 | PPV | SEN | F1 | PPV | SEN | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AD | 1.0.0 | 1.0.0 | 1.0.0 | - | - | - | 0.8.8 | 1.0.0 | 0.9.4 | 1.0.0 | 1.0.0 | 1.0.0 | - | - | - |
| MCI | 0.6.0 | 0.8.0 | 0.6.9 | - | - | - | - | - | - | 1.0.0 | 1.0.0 | 1.0.0 | 0.9.2 | 0.9.7 | 0.9.4 |
| AD+MCI | - | - | - | 0.9.4 | 0.9.7 | 0.9.5 | - | - | - | - | - | - | - | - | - |
| NC | 0.7.0 | 0.4.7 | 0.5.6 | 0.9.3 | 0.8.7 | 0.9.0 | 1.0.0 | 0.8.7 | 0.9.3 | - | - | - | 0.9.7 | 0.9.1 | 0.9.4 |
| Mean | 0.7.7 | 0.7.6 | 0.7.5 | 0.9.3 | 0.9.3 | 0.9.3 | 0.9.4 | 0.9.3 | 0.9.3 | 1.0.0 | 1.0.0 | 1.0.0 | 0.9.5 | 0.9.4 | 0.9.4 |

and false negative classification results for a given set of data items. Then the performance is measured with the following metrics (52): accuracy (ACC); sensitivity (SEN), or recall; specificity (SPE); balanced accuracy (BAC); positive predictive value (PPV), or precision; negative predictive value (NPV), and F1-score, detailed in Equation. (4):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}; \quad F1 = \frac{2TP}{2TP + FP + FN};$$
$$SEN = \frac{TP}{TP + FN}; \quad SPE = \frac{TN}{TN + FP};$$
$$PPV = \frac{TP}{TP + FP}; \quad NPV = \frac{TN}{TN + FN}; \tag{4}$$
$$BAC = \frac{1}{2}(SEN + SPE)$$

In addition, after building a Receiver Operating Characteristic (ROC) of the classifier, its performance is evaluated by the area under the ROC curve (AUC).

Table 2 details the class-wise performance of our 3D-DSA-CNN classifier for a selected cross-validation fold and five specific classification tasks. The ROCs / AUCs of these tests in Figure. 6 and the means and standard deviations of all the metrics of Equation. (4) in Table 3, indicate high robustness and confidence of the AD predictions by the proposed task-specific 3D-DSA-CNN classifier. Its accuracy (ACC) is compared in Table 4 with seven other known approaches that use either just the same, or even additional inputs (imaging modalities). Table 4 presents the average results of ten-fold cross-validation of proposed classifier. Comparing these and other aforementioned experiments, the proposed 3D-DSA-CNN outperforms other approaches in all five task-specific cases. This is in spite of employing only a single imaging modality (sMRI) and performing no prior skull-stripping. Moreover, compared to proposed 3D-CNN approaches in (28), (29), 3D-DSA-CNN

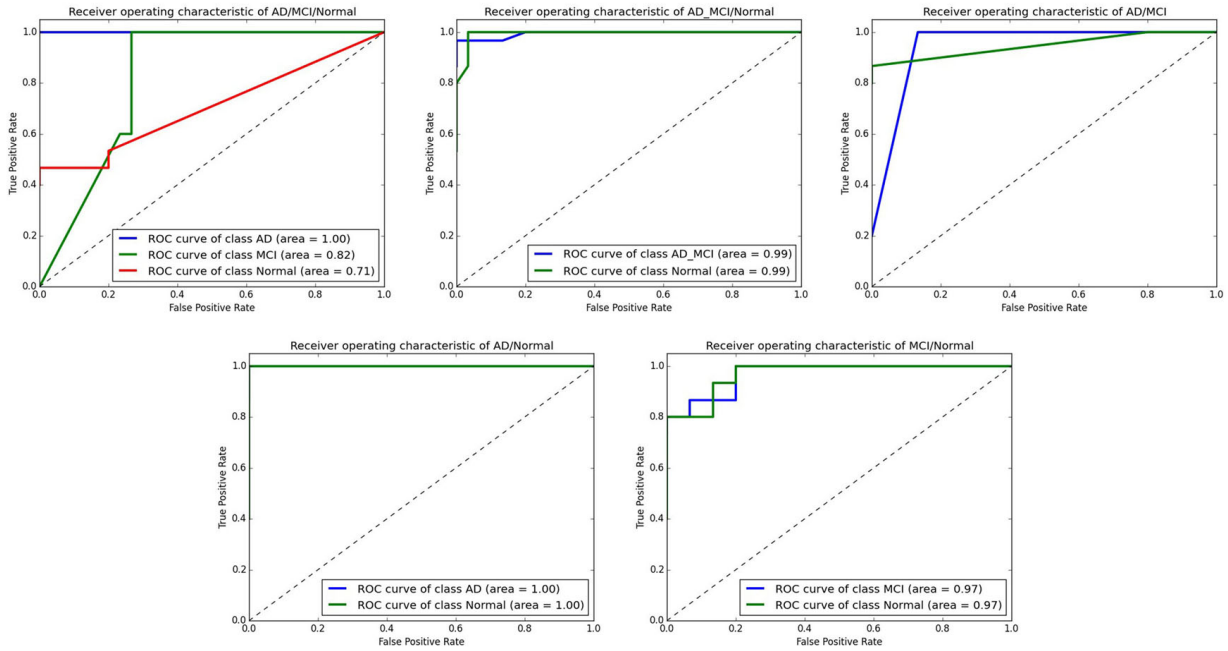


**Figure 6.** ROCs and AUC performance scores for the 3D-DSA-CNN classifier after fine-tuning to the specific task of distinguishing between (left-to-right) AD / MCI / NC; AD+MCI / NC; AD / NC; AD / MCI, and MCI / NC subjects on target domain (ADNI). Note: For each task, ROC curve for each class is computed, by assuming other classes as minor.

**Table 3.** Performance of the proposed 3D-DSA-NCC classifier on target domain (ADNI) [$mean_{STD}$ ,%]

| | **Performance Metrics** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Task** | **ACC** | **SEN** | **SPE** | **BAC** | **PPV** | **NPV** | **AUC** | **F1-Score** |
| AD / MCI / NC | $94.8_{2.6}$ | - | - | - | - | - | - | - |
| AD+MCI / NC | 95.73.1 | 94.84.1 | 97.23.8 | 96.02.9 | 98.42.2 | 91.06.8 | 96.12.9 | 93.94.4 |
| AD / NC | 99.31.6 | 1000 | 98.63.1 | 98.63.1 | 98.63.1 | 1000 | 99.32. | 99.41.3 |
| AD / MCI | $100_{0}$ | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| MCI / NC | $94.2_{2.0}$ | 97.15.7 | 91.44.0 | 91.94.3 | 91.94. | 97.14.5 | 97.12.0 | 94.41.7 |

**Table 4.** Comparative performance (ACC, %) of the classifier vs. seven competitors on ADNI dataset (n/a – non-available).

| **Task-specific classification $mean_{STD}$ ,%** | | | | | | |
|---|---|---|---|---|---|---|
| Approach | **Modalities** | **AD/MCI/NC** | **AD+MCI/NC** | **AD/NC** | **AD/MCI** | **MCI/NC** |
| Gupta *et al.* (20) | MRI | 85.0.n/a | n/a | 94.7.n/a | 88.1.n/a | 86.3.n/a |
| Suk *et al.* (22) | PET+MRI+CSF | n/a | n/a | 95.9.1.1. | n/a | 85.0.1.2. |
| Suk *et al.* (23) | PET+MRI | n/a | n/a | 95.4.5.2. | n/a | 85.7.5.2. |
| Zhu *et al.* (25) | PET+MRI+CSF | n/a | n/a | 95.9.n/a | n/a | 82.0.n/a |
| Zu *et al.* (27) | PET+MRI | n/a | n/a | 96.0.n/a | n/a | 80.3.n/a |
| Liu *et al.* (24) | PET+MRI | 53.8.4.8. | n/a | 91.4.5.6. | n/a | 82.1.4.9. |
| Payan *et al.* (28) | MRI | 89.4.n/a | n/a | 95.3.9n/a | 86.8.n/a | 92.1.n/a |
| Liu *et al.* (19) | MRI | n/a | n/a | 93.8.n/a | n/a | 89.1.n/a |
| Li *et al.* (26) | PET+MRI+CSF | n/a | n/a | 91.4.1.8. | 70.1.2.3. | 77.4.1.7. |
| 3D-ACNN (29) | MRI | 89.1.1.7. | 90.3.1.4. | 97.6.0.6. | 951.8. | 90.8.1.1. |
| 3D-DSA-CNN | MRI | 94.8.2.6. | 95.7.3.1. | 99.3.1.4. | 1000 | 94.2.2.0. |

outperforms (28) in accuracy due to better pre-training of 3D-CAE layers, by exploiting the whole 3D image compared to random patch selection, and improves 3D-ACNN model (29) by better fine-tuning using deep supervision technique.

## 5. DISCUSSION

Several popular non-invasive neuroimaging tools, such as structural MRI (sMRI), functional MRI (fMRI), and positron emission tomography (PET) have been investigated for developing such a system (8), (9). The latter extracts features from the available images, and a classifier is trained to distinguish between different groups of subjects, e.g., AD, Mild Cognitive Impairment (MCI), and Normal Control (NC) groups (3), (10)–(12). The sMRI has been recognized as a promising indicator of the AD progression (3), (13). Comparing to the known diagnostic systems outlines below in Section 2, the proposed system employs a deep 3D Convolutional Neural Network (3D-CNN) pre-trained by 3D Convolutional Autoencoder (3D-CAE) to learn generic discriminative AD features in the lower layers. This captures characteristic AD biomarkers and can be easily adapted to datasets collected in different domains. To increase the specificity of features in upper layers of 3D-CNN, the discriminative loss function is enforced on upper layers (deep supervision). (14).

Voxel-wise, cortical thickness, and hippocampus shape volume features of the sMRI are used to detect the AD (3). The voxel-wise features are extracted after co-aligning (registering) all the brain image data to associate each brain voxel with a vector (signature) of multiple scalar measurements. Kloppel *et al.* (15) used the Gray Matter (GM) voxels as features and trained an SVM to discriminate between the AD and NC subjects. The brain volume in (16) is segmented to GM, White Matter (WM), and Cerebrospinal Fluid (CSF) parts, followed by calculating their voxel-wise densities and associating each voxel with a vector of GM, WM, and CSF densities for classification. For extracting cortical thickness features, Lerch *et al.* (17) segmented the registered brain MRI into the GM, WM, and CSF; fitted the GM and WM surfaces using deformable models; deformed and expanded the WM surface to the GM-CSF intersection; calculated distances between corresponding points at the WM and GM surfaces to measure the cortical thickness, and used these features for classification. To quantify the hippocampus shape for feature extraction, Gerardin *et al.* (18) segmented and spatially aligned the hippocampus regions for various subjects and modeled their shape with a series of spherical harmonics. Coefficients of the series were then normalized to eliminate rotation–translation effects and used as features for training an SVM based classifier.

Leveraging multi-view MRI, PET, and CSM data for trainable feature extraction of AD prediction involves various techniques of machine learning techniques. In particular, Liu *et al.* (19) extracted multi-view features using several selected templates in the subjects' MRI dataset. Tissue density maps of each template were used then for clustering subjects within each class in order to extract an encoding feature of each subject. Finally, an ensemble of Support Vector Machines (SVM) was used to classify the subject. Deep networks were also used for diagnosing the AD with different image modalities and clinical data. Gupta *et al.* (20) employed 2D CNN for slice-wise feature extraction of MRI scans. To boost the classification performance, CNN was pre-trained using Sparse Autoencoder (SAE) (21) trained on random patches of natural images. Suk *et al.* (22) used a stacked autoencoder to separately extract features from MRI, PET, and CSF images; compared combinations of these features with due account of their clinical mini-mental state examination (MMSE) and AD assessment scale-cognitive (ADAS-cog) scores, and classified the AD on the basis of three selected MRI, PET, and CSF features with a multi-kernel SVM. Subsequently, a multimodal deep Boltzmann machine (BM) was used (23) to extract one feature from each selected patch of the MRI and PET scans and predict the AD with an ensemble of SVMs. Liu *et al.* (24) extracted 83 regions of interest (ROI) from the MRI and PET scans and used multi-modal fusion to create a set of features to train stacked layers of denoising autoencoders. Zhu *et al.* (25) proposed a joint regression and prediction model for clinical score and disease group. A linear combination of features are used with imposing group lasso constraint to sparsify the feature selection in regression or classification. Li *et al.* (26) developed a multi-task deep learning for both AD classification and MMSE and ADAS-cog scoring by multi-modal fusion of MRI and PET features into a deep restricted BM, which was pre-trained by leveraging the available MMSE and ADAS-cog scores. Zu *et al.* (27) developed a multi-modal classification model, by proposing a multi-task feature selection method. The feature learning method was based on several regression models of different modalities, where label information is used as regularization parameter to decrease the discrepancy of similar subjects across different modalities, in the new feature space. Then a multi-kernel SVM is used to fuse modality based extracted features for classification. Payan *et al.* (28) proposed a 3D CNN for AD diagnosis based on pre-training by SAE. Randomly selected small 3D patches of MRI scans are used to pre-train SAE, where the trained weights of SAE are later used for pre-training of convolutional filters of 3D CNN. Finally, the fully-connected layers of 3D-CNN are fine tuned for classification. Hosseini-Asl *et al.* (29) proposed a 3DCNN model based on pre-training on a skull-stripped sMRI images to capture anatomical

features, and fine-tuning the fully-connected layers on raw sMRI images for AD diagnosis.

Comparative evaluations of the sMRI-based feature extraction techniques reveal several limitations for classifying the AD (3), (10)–(12). The voxel-wise feature vectors obtained from the brain sMRI are very noisy and can be used for classification only after smoothing and clustering to reduce their dimensionality (16). The cortical thickness and hippocampus model features neglect correlated shape variations of the whole brain structure affected by the AD in other ROIs, e.g., the ventricle's volume. Extracted feature vectors highly depend on image preprocessing due to registration errors and noise, so that feature engineering requires the domain expert knowledge. Moreover, the developed trainable feature extraction and/or classifiers models (19), (20), (22)–(28), (30) are either dependent on using multi-modal data for feature extraction or biased toward a particular dataset which used for training and testing (i.e., the classification features extracted at the learning stage are dataset-specific).

In contrast to all these solutions, Hosseini-Asl *et al.* (29) proposed a 3D Adaptive CNN (3D-ACNN) for learning generic and transferable features across different domains which is able to detect and extract the characteristic AD biomarkers in one (source) domain and perform task-specific classification in another (target) domain. The proposed network combined a generic feature-extracting stacked 3D-CAE, pre-trained in the source domain, as lower layers with the upper task-specific fully-connected layers, which are fine tuned in the target domain (6), (7).

To overcome the aforementioned feature extraction limitations of the conventional approaches, the 3D-CAE learns and automatically extracts discriminative AD features capturing anatomical variations due to the AD. The pre-trained convolutional filters of the 3DCAE are further adapted to another domain dataset, e.g., to the ADNI after pre-training on the CADDementia. Then the entire 3D-CNN is built by stacking the pre-trained 3D-CAE encoding layers followed by the fully-connected layers. In this study, the deep superivision cost function (14) is employed, to improve the task-specific classification performance of 3DACNN, resulting in a 3D Deeply Supervised Adaptable CNN (3D-DSA-CNN) model.

In summary, this paper proposed a 3D-DSA-CNN classifier which can predicts the AD on structural brain MRI scans, more accurately than several other state-of-the-art predictors. The transfer learning concept is used to enhance generalization of the features, in capturing the AD biomarkers, with three stacked 3D CAE network which pre-trained on CADDementia dataset. Subsequently, the features are extracted and used as AD biomarkers detection in upper layers of a 3D-CNN network. Then, three fully-connected layers are stacked on top of the lower layers, to perform AD classification on 210 subjects of ADNI dataset. To boost the classification performance, discriminative loss function was imposed on each fully-connected layers, in addition to the output classification layers, to improve the discrimination between subjects. The results demonstrated that hierarchical feature extraction was improved in hidden layers of 3D-CNN, by better discrimination between AD, MCI, and NC subjects. Seven classification metrics were measured using ten-fold cross-validation and were compared to the state-of-the-art models. The results have demonstrated the out-performance of the proposed 3D-DSA-CNN.

## 6. ACKNOWLEDGEMENT

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH- 12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research& Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the North-ern California Institute for Research and Education, and the study is coordinated by the Alzheimer's disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## 7. REFERENCES

1. G McKhann, D Drachman, M Folstein, R Katzman, D Price, E Stadlan: Clinical diagnosis of Alzheimer's disease: Report of the NINCDSADRDA Work Group under

the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34(7), 939 (1984)

2. Alzheimer's Association: 2014 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 10(2), e47–e92 (2014)

3. E Bron, M Smits, W van der Flier, H Vrenken, F Barkhof, P Scheltens, J Papma, R Steketee, C Orellana, R Meijboom, M Pinto: Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *NeuroImage*, (2015)

4. S Plis, D Hjelm, R Salakhutdinov, E Allen, H Bockholt, J Long, H Johnson, J Paulsen, J Turner, V Calhoun: Deep learning for neuroimaging: a validation study. *Frontiers in Neuroscience* 8, (2014)

5. T Chen, I Goodfellow, J Shlens: Net2Net: Accelerating learning via knowledge transfer. arXiv:1511.0.5641 (cs.LG), (2015)

6. M Long, J Wang: Learning transferable features with deep adaptation networks. arXiv:1502.0.2791 (cs.LG), (2015)

7. J Yosinski, J Clune, Y Bengio, H Lipson: How transferable are features in deep neural networks?. *in Advances in Neural Information Processing Systems*, 3320–3328 (2014)

8. C Jack, M Albert, D Knopman, G McKhann, R Sperling, M Carrillo, B Thies, C Phelps: Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* 7(3), 257–262 (2011)
DOI: 10.1016/j.jalz.2011.03.004
PMid:21514247 PMCid:PMC3096735

9. G McKhann, D Knopman, H Chertkow, B Hyman, C Jack, C Kawas, W Klunk, W. Koroshetz, J. Manly, R. Mayeux, R Mohs: The diagnosis of dementia due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* 7(3), 263–269 (2011)
DOI: 10.1016/j.jalz.2011.03.005
PMid:21514250 PMCid:PMC3312024

10. R Cuingnet, E Gerardin, J Tessieras, G Auzias, S Lehericy, M Habert, M Chupin, H Benali, O Colliot, Alzheimer's Disease Neuroimaging Initiative: 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 (2011)
DOI: 10.1016/j.neuroimage.2010.06.013
PMid:20542124

11. F Falahati, E Westman, A Simmons: Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. *Journal of Alzheimer's Disease* 41(3), 685–708 (2014)

12. M Sabuncu, E Konukoglu: Clinical Prediction from Structural Brain MRI Scans: A Large-Scale Empirical Study. *Neuroinformatics* 13(1), 31–46 (2015)
DOI: 10.1007/s12021-014-9238-1
PMid:25048627 PMCid:PMC4303550

13. C Jack, D Knopman, W Jagust, R Petersen, M Weiner, P Aisen, L Shaw, P Vemuri, H Wiste, S Weigand, T Lesnick: Tracking pathophysio-logical processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* 12(2), 207– 216 (2013)
DOI: 10.1016/S1474-4422(12)70291-0

14. C Lee, S Xie, P Gallagher, Z Zhang, Z Tu: Deeply-supervised nets. *arXiv:1409.5.185* (2014)

15. S Kloppel, C Stonnington, C Chu, B Draganski, R Scahill, J Rohrer, N Fox, C Jack, J Ashburner, R Frackowiak: Automatic classifi-cation of MR scans in Alzheimer's disease. *Brain* 131(3), 681–689 (2008)
DOI: 10.1093/brain/awm319
PMid:18202106 PMCid:PMC2579744

16. Y Fan, D Shen, R Gur, R. Gur, C Davatzikos: COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imag.* 26(1), 93–105 (2007)

17. J Lerch, J Pruessner, A Zijdenbos, D Collins, S Teipel, H Ham-pel, A Evans: Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of Aging* 29(1), 23–30 (2008)
DOI: 10.1016/j.neurobiolaging.2006.09.013
PMid:17097767

18. E Gerardin, G Chetelat, M Chupin, R Cuingnet, B Desgranges, H. Kim, M Niethammer, B Dubois, S Lehericy, L Garnero, F Eustache: Multidimensional classification of hippocampal shape features discrim-inates alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage* 47(4), 1476–1486 (2009) DOI: 10.1016/j.neuroimage.2009.05.036 PMid:19463957 PMCid:PMC3001345

19. M Liu, D Zhang, E Adeli-Mosabbeb, D Shen: Inherent structure based multi-view learning with multi-template feature representation for Alzheimer's disease diagnosis. *IEEE Trans. Biomed. Eng*. 63(7), 1473–1482 (2016)

20. A Gupta, M Ayhan, A Maida: Natural image bases to represent neuroimaging data. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13), *PMLR* 28(3), 987–994 (2013)

21. A Ng: Sparse autoencoder. In: CS294A Lecture notes. URL https://web.stanford.edu/class/cs294a/sparseAutoencoder2011new.pdf: Stanford University (2011)

22. H Suk, D Shen: Deep learning-based feature representation for ad/mci classification. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, 583–590 (2013)

23. H Suk, S Lee, D Shen, Alzheimer's Disease Neuroimaging Initiative: Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* 101, 569–582 (2014) DOI: 10.1016/j.neuroimage.2014.06.077 PMid:25042445 PMCid:PMC4165842

24. S Liu, S Liu, W Cai, H Che, S Pujol, R Kikinis, D Feng, M Fulham: Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* 62(4), 1132–1140 (2015)

25. X Zhu, H Suk, D Shen: A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage* 100, 91–105 (2014) DOI: 10.1016/j.neuroimage.2014.05.078 PMid:24911377 PMCid:PMC4138265

26. F Li, L Tran, K Thung, S Ji, D Shen, J Li: A robust deep model for improved classification of AD/MCI patients. *IEEE J. Biomed. Health Inform.* 19(5), 1610–1616 (2015)

27. C Zu, B Jie, M Liu, S Chen, D Shen, D Zhang, Alzheimer's Disease Neuroimaging Initiative: Label-aligned multi-task feature learning for multimodal classification of Alzheimers disease and mild cognitive impairment. *Brain Imaging and Behavior* 10(4), 1148–1159 (2015) DOI: 10.1007/s11682-015-9480-7 PMid:26572145

28. A Payan, G Montana: Predicting alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *arXiv:1502.0.2506 (cs.CV)*, (2015)

29. E Hosseini-Asl, R Keynton, A El-Baz: Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 126–130 (2016)

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

31. A Alansary, M Ismail, A Soliman, F Khalifa, M Nitzken, A Elnakib, M Mostapha, A Black, K Stinebruner, M Casanova, J Zurada: Infant brain extraction in t1-weighted mr images using bet and refinement using lcdg and mgrf models. *IEEE journal of biomedical and healthinformatics* 20(3), 925–935 (2016) DOI: 10.1109/JBHI.2015.2415477 PMid:25823048

32. Y LeCun, L Bottou, Y Bengio, P Haffner: Gradient-based learning applied to document recognition. *Proc. IEEE* 86(11), 2278–2324 (1998) DOI: 10.1109/5.726791

33. E Hosseini-Asl, J Zurada, O Nasraoui: Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints. *Neural Networks and Learning Systems, IEEE Transactions on* 27(12), 2486 - 2498 (2016) DOI: 10.1109/TNNLS.2015.2479223 PMid:26529786

34. J Masci, U Meier, D Ciresan, J SchmidhuberL: Stacked convolutional auto-encoders for hierarchical feature extraction. In: Artificial

Neural Networks and Machine Learning–ICANN 2011. Springer, 52–59 (2011)

35. A Makhzani, B Frey: A winner-take-all method for training sparse convolutional autoencoders. *arXiv:1409.2.752 (cs.LG, cs.NE)*, (2014)

36. B Leng, S Guo, X Zhang, Z Xiong: 3D object retrieval with stacked local convolutional autoencoder. *Signal Processing* 112, 119–128 (2015)
DOI: 10.1016/j.sigpro.2014.09.005

37. X Glorot, Y Bengio: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS10). Society for Artificial Intelligence and Statistics, 249–256 (2010)

38. Y Bengio, P Lamblin, D Popovici, H Larochelle: Greedy layer-wise training of deep networks. *Advances in neural information processing systems* 19, 153–160 (2007)

39. S Thrun: Is learning the n-th thing any easier than learning the first?. *Advances in Neural Information Processing Systems*, 640–646 (1996)

40. R Caruana: Multitask learning. *Machine Learning* 28(1), 41–75 (1997)
DOI: 10.1023/A:1007379606734

41. R Raina, A Ng, D Koller: Constructing informative priors using transfer learning. In: Proceedings of the 23rd International Conference on Machine Learning. ACM, 713–720 (2006)

42. J Baxter: A Bayesian/information theoretic model of learning to learn via multiple task sampling. *Machine Learning* 28(1), 7–39 (1997)
DOI: 10.1023/A:1007327622663

43. J Bridle, S Cox: RecNorm: Simultaneous normalisation and classification applied to speech recognition. In: *Advances in Neural Information Processing Systems*, 234–240 (1990)

44. S Ben-David, J Blitzer, K Crammer, A Kulesza, F Pereira, J. Vaughan: A theory of learning from different domains. *Machine Learning*, 79(1), 151–175 (2009)

45. K Crammer, M Kearns, J Wortman: Learning from multiple sources. *The Journal of Machine Learning Research* 9, 1757–1774 (2008)

46. G Dauphin, X Glorot, S Rifai, Y Bengio, I Goodfellow, E Lavoie, X Muller, G Desjardins, D Warde-Farley, P Vincent, A Courville: Unsupervised and transfer learning challenge: a deep learning approach. ICML Unsupervised and Transfer Learning, 97–110 (2012)

47. X Glorot, A Bordes, Y Bengio: Domain adaptation for large scale sentiment classification: A deep learning approach. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), 513–520 (2011)

48. J Weston, F Ratle, H Mobahi, R Collobert: Deep learning via semi-supervised embedding. In Neural Networks: Tricks of the Trade. Springer, 639–655 (2012)

49. M Zeiler: ADADELTA: An adaptive learning rate method. *arXiv:1212.5.701 (cs.LG)*, (2012)

50. L Maaten, G Hinton: Visualizing Data using t-SNE. *Journal of Machine Learning Research,* 9(11), (2008)

51. F Bastien, P Lamblin, R Pascanu, J Bergstra, I Goodfellow, A Bergeron, N Bouchard, D Warde-Farley, Y Bengio: Theano: new features and speed improvements. *arXiv:1211.5.590 (cs.SC)*, (2012)

52. R Fletcher, S Fletcher, G Fletcher: Clinical epidemiology: the essentials. Lippincott Williams & Wilkins, (2012)

**Key Words:** Alzheimer's disease, deep learning, 3D convolutional network, Autoencoder, brain MRI

**Send correspondence to:** Mohammed Ghazal, Department of Electrical and Computer Engineering, Abu Dhabi University, UAE, Tel: 971-800-23968, Fax: 971-800-23970, E-mail: mohammed.ghazal@adu.ac.ae