ALZHEIMER'S DISEASE DIAGNOSTICS BY ADAPTATION OF 3D CONVOLUTIONAL NETWORK

Ehsan Hosseini-Asl[1], Robert Keynton[2], Ayman El-Baz[2]*

[1]Electrical and Computer Engineering Department, University of Louisville, Louisville, KY, USA.
[2]Bioengineering Department, University of Louisville, Louisville, KY, USA.

ABSTRACT

Early diagnosis, playing an important role in preventing progress and treating the Alzheimer's disease (AD), is based on classification of features extracted from brain images. The features have to accurately capture main AD-related variations of anatomical brain structures, such as, e.g., ventricles size, hippocampus shape, cortical thickness, and brain volume. This paper proposed to predict the AD with a deep 3D convolutional neural network (3D-CNN), which can learn generic features capturing AD biomarkers and adapt to different domain datasets. The 3D-CNN is built upon a 3D convolutional autoencoder, which is pre-trained to capture anatomical shape variations in structural brain MRI scans. Fully connected upper layers of the 3D-CNN are then fine-tuned for each task-specific AD classification. Experiments on the *CADDementia* MRI dataset with no skull-stripping preprocessing have shown our 3D-CNN outperforms several conventional classifiers by accuracy. Abilities of the 3D-CNN to generalize the features learnt and adapt to other domains have been validated on the *ADNI* dataset.

Index Terms— Alzheimer's disease, deep learning, 3D convolutional neural network, autoencoder, brain MRI

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive brain disorder and the most common case of dementia in the late life. AD leads to the death of nerve cells and tissue loss throughout the brain, thus reducing the brain volume in size dramatically through time and affecting most of its functions [1]. The estimated number of affected people will double for the next two decades, so that one out of 85 persons will have the AD by 2050 [2]. Because the cost of caring 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].

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 [4,5]. 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, 6–8]. The sMRI has been recognized as a promising indicator of the AD progression [3,9].

Various machine learning techniques were employed to leverage multi-view MRI, PET, and CSM data to predict the AD. Liu et al. [10] 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 machine (SVM) was used to classify the subject. Deep networks were also used for diagnozing the AD with different image modalities and clinical data. Suk et al. [11] used a stacked autoencoder to separately extract features from MRI, PET, and cerebrospinal fluid (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. Later on, a multi-modal deep Boltzmann machine (BM) was used [12] 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. [13] extracted 83 regions-of-interest (ROI) from the MRI and PET scans and used multimodal fusion to create a set of features to train stacked layers of denoising autoencoders. Li et al. [14] developed a multi-task deep learning for both AD classification and MMSE and ADAS-cog scoring by multimodal fusion of MRI and PET features into a deep restricted BM, which was pre-trained by leveraging the available MMSE and ADAS-cog scores.

Voxel-wise, cortical thickness, and hippocampus shape-volume features of the sMRI are used to diagnose 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. Klöppel 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 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; deform and expand the WM surface to the GM-CSF intersection; calculate distances between corresponding points at the WM and GM surfaces to measure the cortical thickness, and use 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.

Comparative evaluations [3, 6–8] revealed several limitations of the above feature extraction techniques for classifying the AD. 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 vari-

---
*Corresponding Author:— Tel: (502) 852 3165, Fax: (502) 852 3940, E-mail: ehsan.hosseiniasl@louisville.edu
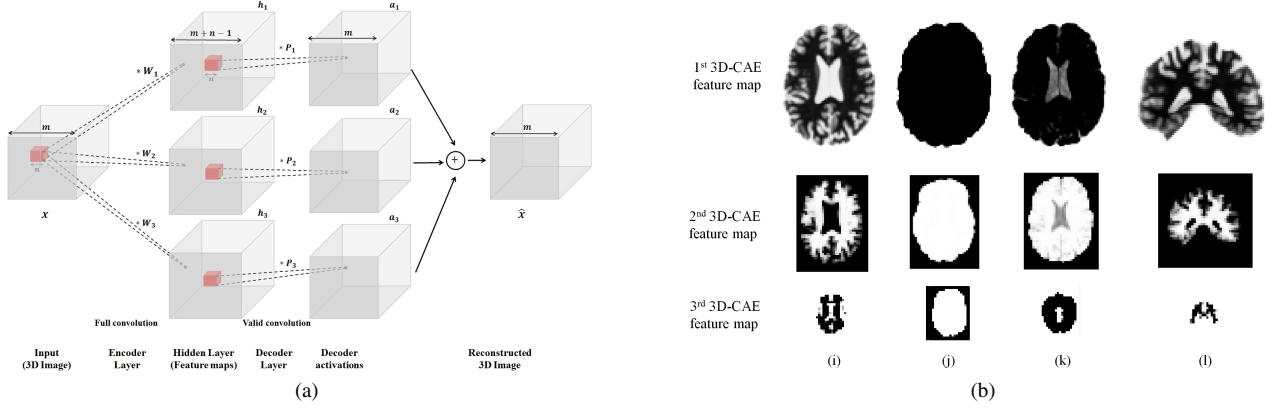
Fig. 1. (a) Schematic diagram of 3D Convolutional Autoencoder for feature extraction of a 3D Image based on reconstructing the input. Note that the image dimension increases in the encoding layer due to full convolution, and decreases to original dimension by valid convolution. (b) Selected slices of hierarchical 3D feature maps (trained on 30 subjects of *CADDementia*) in (i,j,k) axial and (l) sagittal view extracted at 3 layers of stacked 3D-CAE indicating (i) cortex thickness and volume, (j) brain size, (k) ventricle size, and (l) hippocampus model, extracted from the brain structural MRI. The feature maps are downsampled at each layer using max-pooling to reduce the size and detect higher level features.

ations of the whole brain structure affected by the AD in other ROIs, e.g., the ventricle's volume. Appropriateness of the extracted feature vectors highly depend on image preprocessing due to registration errors and noise, so that feature engineering requires the domain expert knowledge. Most of the trainable classifiers are biased toward a particular dataset, which was used for training and testing (i.e., the classification features extracted at the learning stage are dataset-specific).

We propose a new deep 3D convolutional neural network (3D-CNN), for unsupervised generic and transferable feature extraction based on 3D extension of convolutional autoencoder (3D-CAE) [19] to overcome the aforementioned limitations in feature extraction from brain sMRI for AD classification. The proposed network combines a pretrained 3D-CAE in the source domain, e.g. *CAD-Dementia*, with upper task-specific layers to be fine-tuned in the target domain, e.g. *ADNI* dataset, [20, 21]. Such adaptation of pre-learned generic features allows for calling the proposed classifier a 3D Adaptable CNN (3D-ACNN).

2. MODEL

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 target-domain-adaptable 3D-CNN. The 3D-CAE architecture and the AD diagnosis framework using the 3D-ACNN are detailed in Sections 2.1 and 2.2, respectively.

2.1. Feature Extraction using 3D Convolutional Autoencoder:

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, due to growing fast numbers of parameters to be evaluated in the input (encoding) and output (decoding) layers [22]. Moreover, local features are more suitable for extracting patterns from high-dimensional images, although autoencoders with full connections between all nodes of the layers try to learn global features. 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 [19, 23, 24]. Each input image is reduced hierarchically, using the hidden feature (activation) map of each CAE for training the next-layer CAE.

Our 3D extension of a hierarchical CAE proposed in [19] is sketched in Fig. 1(a). To capture characteristic variations of an input 3D image, $\mathbf{x}$, 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, $\mathbf{x}_{i:\mathrm{neib}}$, of this node with a linear encoding filter, specified by its weights, $\mathbf{W}_k = [\mathbf{W}_{j:k} : j = 1, \ldots, J]$ for each relative neighboring location with respect to the node i and each voxel-wise signal component j, followed by feature-specific biases, $\mathbf{b}_k = [b_{j,k} : j = 1, \ldots, J]$ and non-linear transformations with a certain activation function, $f(\cdot)$:

$$h_{i:j:k} = f\left(\mathbf{W}_k * \mathbf{x}_{i:\mathrm{neib}} + b_{j:k}\right) \tag{1}$$

The latter function is selected from a rich set of constraining differentiable one, including, in particular, the sigmoid, $f(u) = (1 + \exp(-u))^{-1}$ and rectified linear unit (ReLU), $f(u) = \max(0, u)$ [25]. Since the 3D image $\mathbf{x}$ in Eq. (1) has the J-vectorial voxel-wise signals, actually, 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 notation, let $\mathbf{h}_k = \mathbb{T}\left(\mathbf{x} : \mathbf{W}_k, \mathbf{b}_k, f(\cdot)\right)$ denote the entire encoding of the input 3D image with J-vectorial voxel-wise signals with the k-th 3D feature map, $\mathbf{h}_k$, 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 like inverse transformation, $\mathbf{T}_{\mathrm{inv}}(\ldots)$, with the same voxel neighborhood, but generally with the different convolutional weights, $\mathbf{P}_k$, biases, $\mathbf{b}_{\mathrm{inv}:k}$,
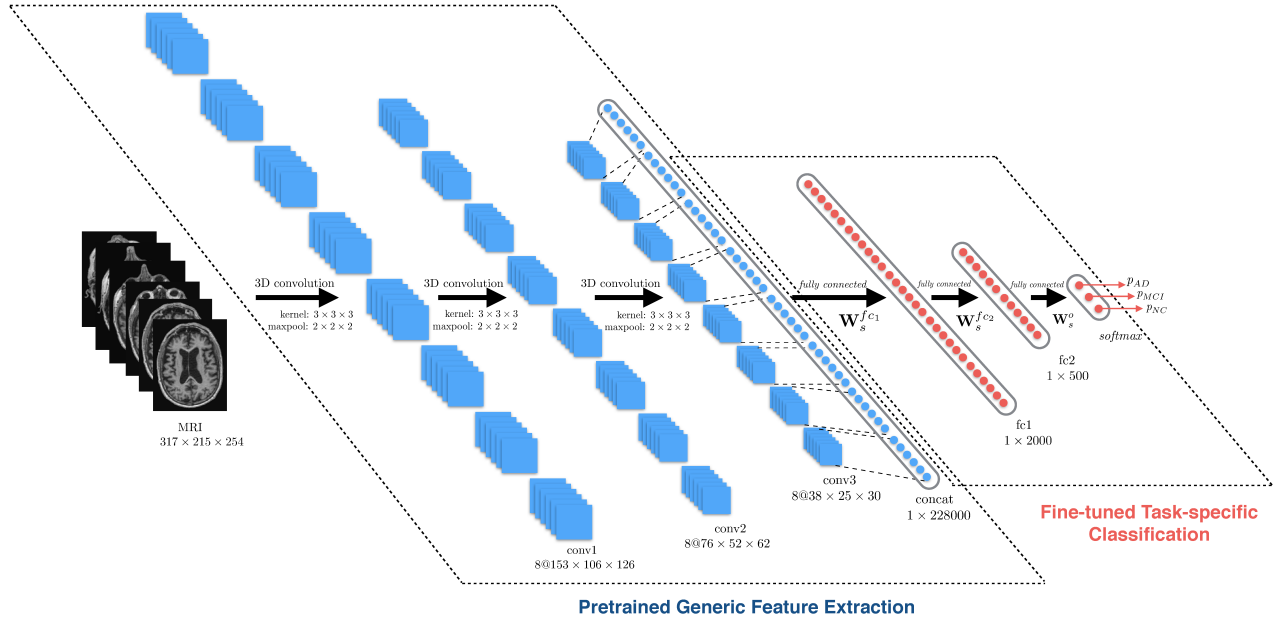
Fig. 2. Our 3D adaptable CNN (3D-ACNN) for Alzheimer's disease diagnosis.

and, possibly, activation function, $g(\cdot)$, decodes, or reconstructs the initial 3D image:

$$\widehat{\mathbf{x}} = \sum_{k=1}^{K} \underbrace{\mathbf{T}_{\text{inv}}\left(\mathbf{h}_k : \mathbf{P}_k, \mathbf{b}_{\text{inv}:k}, g(\cdot)\right)}_{\mathbf{a}_k} \quad (2)$$

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

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, $\mathbf{x}^{[t]}$; $t = 1, \ldots, T$,

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

where $\boldsymbol{\theta} = [\mathbf{W}_k; \mathbf{P}_k; \mathbf{b}_k; \mathbf{b}_{\text{inv}:k} : k = 1, \ldots, K]$, 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 the free parameters, the decoding, $\mathbf{P}_k$, and encoding, $\mathbf{W}_k$, weights were tied by flipping over all their dimensions as proposed in [19]. The cost of Eq. (3) was minimized in the parameter space by using the stochastic gradient descent search, combined with error back-propagation.

2.2. AD Classification by 3D Adaptive CNN (3D-ACNN)

While the lower layers of a goal predictive 3D-CNN extract generalized features, the upper layers have to facilitate task-specific classification using those features [21]. The proposed classifier extracts the generalized features by using a stack of locally connected bottom convolutional layers, while performing task-specific fine-tuning of parameters of the fully connected upper layers. Training the proposed hierarchical 3D-CNN consists of pre-training, initial training of the lower convolutional layers, and final 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 pre-trained in the source domain. Then these layers are initialized by encoding the 3D-CAE weights [26], and, finally, the deep-supervision-based [27] fine-tuning of the upper fully-connected layers is performed for each task-specific binary or multi-class sMRI classification.

Due to 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 Fig. 1(b). We use the Net2Net initialization [26], which allows for different convolutional kernel and pool sizes of the 3D-CNN layers, comparing to those in the 3D-CAE, based on the target-domain image size and imaging specifications, and facilitates adapting the 3D-CNN across the different domains. To classify the extracted features in a task-specific way, weights of the upper fully-connected 3D-CNN layers 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 negated log-likelihood of the true output classes, given the input features extracted from the training target-domain images by the pre-trained bottom part of the network.

Our implementation of the 3D-CNN uses the ReLU activation functions at each inner layer and the fully connected upper layers with a softmax top-most output layer (Fig. 2), predicting the probability of belonging an input brain sMRI to the AD, MCI, or NC group. The Adadelta gradient descent [28] was used to update the pre-trained 3D-CAE and fine-tune the entire 3D-ACNN.

Table 1. Demographic data for 210 subjects from the 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$_{\pm \text{STD}}$)	$75.0_{\pm 7.9}$	$75.9_{\pm 7.7}$	$74.6_{\pm 6.1}$

Table 2. Comparative performance (ACC,%) of our classifier vs. seven competitors (n/a – non-available).

Approach	Modalities	Task-specific classification [mean_{STD},%].				
		AD/MCI/NC	AD+MCI/NC	AD/NC	AD/MCI	MCI/NC
Suk et al. [11]	PET+MRI+CSF	n/a	n/a	$95.9_{1.1}$	n/a	$85.0_{1.2}$
Suk et al. [12]	PET+MRI	n/a	n/a	$95.4_{5.2}$	n/a	$85.7_{5.2}$
Zhu et al. [29]	PET+MRI+CSF	n/a	n/a	$95.9_{\text{n/a}}$	n/a	$82.0_{\text{n/a}}$
Zu et al. [30]	PET+MRI	n/a	n/a	$96.0_{\text{n/a}}$	n/a	$80.3_{\text{n/a}}$
Liu et al. [13]	PET+MRI	$53.8_{4.8}$	n/a	$91.4_{5.6}$	n/a	$82.1_{4.9}$
Liu et al. [10]	MRI	n/a	n/a	$93.8_{\text{n/a}}$	n/a	$89.1_{\text{n/a}}$
Li et al. [14]	PET+MRI+CSF	n/a	n/a	$91.4_{1.8}$	$70.1_{2.3}$	$77.4_{1.7}$
Sarraf et al. [31]	fMRI	n/a	n/a	$96.8_{\text{n/a}}$	n/a	n/a
Our 3D-ACNN	MRI	$\mathbf{89.1_{1.7}}$	$\mathbf{90.3_{1.4}}$	$\mathbf{97.6_{0.6}}$	$\mathbf{95_{1.8}}$	$\mathbf{90.8_{1.1}}$

3. EXPERIMENTAL RESULTS

To performance of the proposed 3D-CAES is evaluated on *CADDementia* [1] as source domain, for generalized feature extraction. The 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]. To pretrain 3D-CAES on *CADDementia*, sMRI are preproceesed by spatially normalizing using rigid registration approach. Then skull is removed and image intensities are normalized to $[0, 1]$, resulting in sMRI of size $(200 \times 150 \times 150)$.

The classification performance of proposed 3D-ACNN is evaluated on Alzheimers Disease Neuroimaging Initiative (ADNI) database, as target domain, for five classification tasks: four binary ones (AD vs. NC, AD+MCI vs. NC, AD vs. MCI, MCI vs NC) and the ternary classification (AD vs. MCI vs. NC). Classification accuracy was evaluated for each test by ten-fold cross-validation. ADNI dataset without using any preprocessing and skull stripping is used AD classification, compared to preprocessed *CADDementia* dataset. The Theano library [32] was used to develop the deep CNN implemented for our experiments on the Amazon EC2 g2.8xlarge instances with GPU GRID K520.

To pretrain 3D-CAE for feature extraction of brain sMRI, we use ReLu nonlinear function as the encoder and decoder layer's activation. The 3D-CAE contains eight encoding and decoding filters of size $(3 \times 3 \times 3)$. Three hierarchical 3D-CAE's are trained to extract the low-dimensional feature maps. The extracted feature maps are of dimension $(102 \times 76 \times 76)$, $(52 \times 40 \times 40)$, and $(28 \times 22 \times 22)$, subsequently. Each 3D-CAE extracts eight feature maps, according to the number of their encoding filters (Fig. 1(b)). Selected slices of the three feature maps from each layer of our stacked 3D-CAE (abbreviated 3D-CAES below) in Fig. 1(b), show that the learnt generic convolutional filters can really capture features related to the 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 CADementia database. According to these projections, the first layer of the 3D-CAES extracts the cortex thickness as a discriminative AD feature of AD, whereas the brain size (related to the patient gender), size of ventricles, and hippocampus model are represented by the subsequent layers. Each 3D-CAES layer combines the extracted lower-layer feature maps in order to train the higher level for describing more in detail the anatomical variations of the brain sMRI . Both the ventricle size and cortex thickness features are combined to extract conceptually higher-level features at the next layers.

Performance of the proposed 3D-ACNN classifier, in terms of accuracy (ACC), for each task-specific classification was evaluated and compared to competing approaches [10–14, 29–31]. Table 2 presents the average results of ten-fold cross-validation of our classifier. According to these experiments, the proposed 3D-ACNN outperforms the other approaches in all five task-specific cases, in spite of employing only a single imaging modality (sMRI) and performing no prior skull-stripping.

4. CONCLUSION AND FUTURE WORK

This paper proposed a 3D-ACNN classifier, which can more accurately predict the AD on structural brain MRI scans, than several other state-of-the-art predictors. The pretraining and layer freezing were used to enhances generality of features in capturing the AD biomarkers. Three stacked 3D CAE network were pretrained on *CADDementia* Dataset. Then the learnt features are extracted and used as AD biomarkers detection in bottom layers of a 3D CNN network. Then three fully connected layers are stacked on top of the bottom layers to perform AD classification on 210 subjects of ADNI dataset. Classification performance were measured using ten-fold crossvalidation, and were compared to the state-of-the-art models, demonstrated the out-performance of the proposed 3D CNN. The future application of proposed 3D-ACNN include detection of lung cancer [33–36], heart failure [37], and autism detection [38].

5. REFERENCES

[1] G. McKhann et al., "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*, vol. 34, no. 7, pp. 939, 1984.

[2] Alzheimer's Association et al., "2014 alzheimer's disease facts and figures," *Alzheimers Dement*, vol. 10, no. 2, pp. e47–e92, 2014.

[3] E.E. Bron et al., "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge," *NeuroImage*, 2015.

[4] C.R. Jack et al., "Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dement*, vol. 7, no. 3, pp. 257–262, 2011.

[5] G.M. McKhann et al., "The diagnosis of dementia due to Alzheimers disease: Recommendations from the National In-

stitute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dement*, vol. 7, no. 3, pp. 263–269, 2011.

[6] R. Cuingnet et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, 2011.

[7] F. Falahati et al., "Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging," *J. of Alzheimer Dis*, vol. 41, no. 3, pp. 685–708, 2014.

[8] M.R. Sabuncu et al., "Clinical Prediction from Structural Brain MRI Scans: A Large-Scale Empirical Study," *Neuroinformatics*, vol. 13, no. 1, pp. 31–46, 2015.

[9] C.R. Jack et al., "Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers," *Lancet Neurol*, vol. 12, no. 2, pp. 207–216, 2013.

[10] M. Liu et al., "Inherent structure based multi-view learning with multi-template feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. PP, no. 99, pp. 1–1, 2015.

[11] Heung-Il Suk and Dinggang Shen, "Deep learning-based feature representation for ad/mci classification," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 583–590. Springer, 2013.

[12] Heung-Il Suk et al., "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.

[13] Siqi Liu et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 4, pp. 1132–1140, 2015.

[14] Feng Li et al., "A robust deep model for improved classification of AD/MCI patients," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1610–1616, 2015.

[15] S. Klöppel et al., "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, 2008.

[16] Y. Fan et al., "COMPARE: classification of morphological patterns using adaptive regional elements," *IEEE-TBME*, vol. 26, no. 1, pp. 93–105, 2007.

[17] J.P. Lerch et al., "Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls," *Neurobiol Aging*, vol. 29, no. 1, pp. 23–30, 2008.

[18] E. Gerardin et al., "Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage*, vol. 47, no. 4, pp. 1476–1486, 2009.

[19] J. Masci et al., "Stacked convolutional auto-encoders for hierarchical feature extraction," in *ICANN 2011*, pp. 52–59. Springer, 2011.

[20] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, 2014, pp. 3320–3328.

[21] Mingsheng Long and Jianmin Wang, "Learning transferable features with deep adaptation networks," *arXiv:1502.02791 [cs.LG]*, 2015.

[22] Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[23] A. Makhzani and B. Frey, "A winner-take-all method for training sparse convolutional autoencoders," *arXiv preprint arXiv:1409.2752*, 2014.

[24] B. Leng et al., "3D object retrieval with stacked local convolutional autoencoder," *Signal Processing*, vol. 112, no. 0, pp. 119 – 128, 2015.

[25] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *AISTATS*, 2010, pp. 249–256.

[26] Tianqi Chen, Ian Goodfellow, and Jonathon Shlens, "Net2Net: Accelerating learning via knowledge transfer," *arXiv:1511.05641 [cs.LG]*, 2015.

[27] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu, "Deeply-supervised nets," *arXiv:1409.5185 [stat.ML, cs.LG, cs.NE]*, 2014.

[28] Matthew D Zeiler, "ADADELTA: An adaptive learning rate method," *arXiv:1212.5701 [cs.LG]*, 2012.

[29] Xiaofeng Zhu et al., "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *NeuroImage*, vol. 100, pp. 91–105, 2014.

[30] Chen Zu et al., "Label-aligned multi-task feature learning for multimodal classification of Alzheimers disease and mild cognitive impairment," *Brain Imaging and Behavior*, pp. 1–12, 2015.

[31] Saman Sarraf and Ghassem Tofighi, "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks," *arXiv:1603.08631 [cs.CV]*, 2016.

[32] Frédéric Bastien et al., "Theano: new features and speed improvements," *arXiv:1211.5590 [cs.SC]*, 2012.

[33] Ayman El-Baz et al., "A unified approach for detection, visualization, and identification of lung abnormalities in chest spiral ct scans," in *International Congress Series*. Elsevier, 2003, vol. 1256, pp. 998–1004.

[34] Ayman El-Baz et al., "Automatic analysis of 3d low dose ct images for early diagnosis of lung cancer," *Pattern Recognition*, vol. 42, no. 6, pp. 1041–1051, 2009.

[35] Ayman El-Baz et al., "3d shape analysis for early diagnosis of malignant lung nodules," in *Information Processing in Medical Imaging*. Springer, 2011, pp. 772–783.

[36] Aly A Farag et al., "Appearance models for robust segmentation of pulmonary nodules in 3d ldct chest images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 662–670. Springer, 2006.

[37] Fahmi Khalifa et al., "Accurate automatic analysis of cardiac cine images," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 2, pp. 445–455, 2012.

[38] Manuel F Casanova et al., "Quantitative analysis of the shape of the corpus callosum in patients with autism and comparison individuals," *Autism*, vol. 15, no. 2, pp. 223–238, 2011.