2018 IEEE 31st International Symposium on Computer-Based Medical Systems

Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs using Cross-Modal Transfer Learning.

Karim Aderghal*†, Alexander Khvostikov*§, Andrei Krylov§, Jenny Benois-Pineau*, Karim Afdel†, Gwenaelle Catheline‡
* *LaBRI UMR 5800, Univ. Bordeaux, Bordeaux, France*
Email: karim.aderghal@u-bordeaux.fr , jenny.benois-pineau@u-bordeaux.fr
† *LabSIV, Departement of Computer Science, University Ibn Zohr, Agadir, Morocco*
Email: k.afdel@uiz.ac.ma
‡ *INCIA, Univ. Bordeaux, Bordeaux, France*
Email: gwenaelle.catheline@u-bordeaux.fr
§ *Lomonosov Moscow State University, Department of Computational Mathematics and Cybernetics, Moscow, Russia*
Email: xubiker@gmail.com , kryl@cs.msu.ru

***Abstract*—A recent imaging modality Diffusion Tensor Imaging completes information used from Structural MRI in studies of Alzheimer disease. A large number of recent studies has explored pathologic staging of Alzheimer disease using the Mean Diffusivity maps extracted from the Diffusion Tensor Imaging modality. The Deep Neural Networks are seducing tools for classification of subjects' imaging data in computer-aided diagnosis of Alzheimer's disease. The major problem here is the lack of a publicly available large amount of training data in both modalities. The lack number of training data yields over-fitting phenomena. We propose a method of a cross-modal transfer learning: from Structural MRI to Diffusion Tensor Imaging modality. Models pre-trained on a structural MRI dataset with domain-depended data augmentation are used as initialization of network parameters to train on Mean Diffusivity data. The method shows a reduction of the over-fitting phenomena, improves learning performance, and thus increases the accuracy of prediction. Classifiers are then fused by a majority vote resulting in augmented scores of classification between Normal Control, Alzheimer Patients and Mild Cognitive Impairment subjects on a subset of ADNI dataset.**

***Keywords*-Multi-Modal ; Alzheimer's Disease ; Hippocampus ; Mild Cognitive Impairment ; Convolutional Neural Networks ; Transfer Learning ; Deep Learning ; Medical Imaging.**

I. Introduction

Dementia is a set of symptoms of impairment in memory affecting the brain in the elderly individuals. Accordingly to the recent statistics, there are more than 46.8 million people now living with dementia, and the number will be increased to 131.5 million in 2050 [1]. Alzheimer's Disease (AD) is the most common neurodegenerative disease form of dementia, It is characterized by the brain atrophy and the neuronal dysfunction. It begins with short-term memory loss, and in advanced stage includes the cognitive impairments, behavioral disturbances, and distortion of protein nerve cells in the brain. Mild cognitive Impairment (MCI) is the condition associated with the risk to convert to Alzheimer's disease. For a long time, the Structural Magnetic resonance imaging (sMRI) has been the most used technique for diagnosis and analysis of progressive deterioration. It allows for tracking structural changes in the brain and the inevitable atrophy caused by the neurodegeneration aspect of the AD pathology [2]. Most studies use the sMRI modality to explore the deterioration process of some parts of the brain known to be affected by AD, such as the hippocampus. The latter is one of the brain structures to be affected first. Amongst other affected structures there are temporal and cingulate gyri, and precuneus [3][4]. Diffusion tensor imaging (DTI) is a recent MRI modality based on motion of water molecules by quantifying the random water diffusion in cerebral gray and white matters. The interpretation of diffusion tractography of water infers the tissue architecture continuity of the fiber bundles in brain from voxel to voxel [5][6]. For diagnostic tasks, Mean Diffusivity (MD) and Fractional Anisotropy (FA) are the main density maps derived from DTI data. The MD describes the magnitude of self-diffusion of water regardless of direction, while FA is used to characterize the degree of anisotropy of the diffusion [6]. In the AD patients studies, the computed DTI maps have shown abnormally increased MD and reduced FA values in gray matter and white matter, including some regions as the Hippocampus. Quantifying the neurodegeneration in the Hippocampus area leads to measure the pathological atrophy change, and discriminate AD subjects from those with Mild Cognitive Impairment (MCI) and Normal Control (NC) [7].

Since recently, Computer-Aided Diagnosis (CAD) systems have benefited of the performance of machine learning algorithms. Due to the great success achieved in various applications, Deep Learning (DL) and specifically Convolutional Neural Networks (CNN) are promising methods in the domain of medical imaging, such as medical image classification, detection and segmentation problems. However, they require a large number of data samples to train a

2372-9198/18/$31.00 ©2018 IEEE
DOI 10.1109/CBMS.2018.00067

network. If the training dataset is not sufficiently large, then the so-called over-fitting phenomena occur [8]. This means an increase of cost function to minimize along training iterations. Specific data augmentation and regularization methods such as Dropout can prevent and solve the problem of insufficiency of samples [8][9]. Nevertheless, they remain limited in case of small datasets as this is the case in AD studies: only few AD scans from both sMRI and DTI modalities are available for numerous cohorts in a public ADNI dataset [1] or in particular cohorts such as "Bordeaux-3City cohort" [10]. In this case the transfer learning, that is re-training models based on large dataset (source dataset) achieves good performance results compared to training only with an arbitrary initialization of network parameters ("from scratch") [11]. In this work, we propose a method that combines the sMRI and MD imaging modalities, we apply it on the Hippocampal region which is one of AD biomarkers. The fusion of multiple sources is a way to increase the performance of classification accuracy. Furthermore, with the limited amount of MD dataset we employ the transfer learning of models that have been trained on the sMRI dataset to the target MD dataset.

II. Related work

In the literature, there are few works focused on applying the machine learning on the DTI data. Due to limited number of samples several studies have proposed approaches to expand available datasets. In [12] the authors demonstrated the utility of using pretrained models on a large-scale dataset as an initialization of other network. In their case instead of training the network from scratch they used GoogleNet and Inception-ResNet networks trained on non-medical dataset, and then fine-tunined the final fully connected (FC) layer. Another study [13] was employing the transfer learning and data augmentation to deal with the limited data to recognize MCI on MRI images. Here two different datasets "OASIS" [2] "LIDC"[3] have been used for the pre-training stage, they achieved best performance with accuracy of 90.6% for MCI vs. NC. The DTI image maps are often seen as good modality for the detection of Alzheimer's disease. Thus the authors in [14] have compared the NC, AD and MC using MD and FA maps. Their results showed that MD was a better indicator of change than FA. [15] integrates a method called Multi-auxiliary domain, they transfer the learned models from a trained model. They use MCI vs NC and progressive MCI (pMCI) vs. stable MCI (sMCI) as Auxiliary domains for AD vs. NC classification. Achieved classification accuracies are of 95.2% and 82.1% for AD vs. NC, and MCI vs. NC respectively. Another work, using the ImageNet as the source domain, [16] uses full brain with joint classification (Axial, Coronal, and Sagittal projections), all layers being initialized with pretrainded model except the last FC layer. Fine-tuning was performed here on images of the three modalities (sMRI, PET-FDG and PET-AV45). AD vs. NC the accuracy of classification is of 83.57%. Finally, [17] used FA and MD with the SVM classifier, using Plant's approach and an information gain criterion to classify 137 subjects as AD or NC. They selected voxels to improve classification performance, and they achieved an accuracy of 80% for classification based on FA, and 83% for MD.

[1] http://adni.loni.usc.edu/

[2] http://www.oasis-brains.org/

[3] https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI/

III. Methodology and approach

A. Data Acquisition and Preprocessing

In this study, data were acquired from the Alzheimer's Disease Neuroimagine Initiative (ADNI) database (http://adni.loni.usc.edu). Subjects with only sMRI images were selected from the ADNI 1 phase, the same dataset as used in [9]. The second subset of images that included the group of subjects with both sMRI and the DTI modalities, was acquired from the ADNI Go and ADNI 2. The table I presents the description of the demographic characteristics of subjects, including age, gender, and the Mini Mental State Examination (MMSE) score.

Classes	Subjects	Age	Gender (F/M)	MMSE
AD	188	75.4 ± 7.52	99/89	23.3 ± 2.03
	*48	75.6 ± 8.63	28/20	23.0 ± 2.42
MCI	399	74.9 ± 7.30	256/143	27.0 ± 1.78
	*108	73.4 ± 7.47	66/42	27.4 ± 1.99
NC	228	76.0 ± 5.02	118/110	29.1 ± 1.00
	*58	73.4 ± 5.90	28/30	28.9 ± 1.18

Table I: Demographic description of the ADNI dataset group. Values are reported as mean and standard deviation (* Subjects with both modalities).

In this work, Hippocampal region is in focus. In order to select it using a Brain atlas, a preprocessing is necessary. Thus (see Fig. 1), the alignment of anatomical scans onto the MNI template standard space [18] was performed using affine linear registration with re-sampling to the new resolution ($121\times145\times121$) with $1.5mm^3$ voxel size [9][19]. Then the voxel intensity was normalized in order for similar structures to have similar intensities. After the alignment of sMRI scans all of them are registered in the same coordinate system, that one of MNI template. The corresponding atlas AAL [20] can be used to select the hippocampal region-of-interest (ROI). Now the MD image has to be "aligned" with corresponding sMRI in order to select the same ROI in it. For each subject, we affinely co-registered MD image to the coordinate system of its corresponding sMRI scan. The affine transformation between sMRI and MD was estimated using mutual information criterion [21], implemented with Statistical Parametric Mapping software (SPM8, Wellcome Trust Centre for Neuroimaging at UCL, London, UK)[4]

[4] http://www.fil.ion.ucl.ac.uk/

running on MATLAB (Math-Works, Sherborn, MA, USA). We use the skull stripping segmentation on the sMRI to eliminate "bright" voxels of skull which would disturb co-registration process. This segmentation was performed with SPM8 toolbox. It allows for segmenting the brain into grey matter, white matter and Cerebro-Spinal Fluid (CSF). Merging these three maps we can subtract skull region from original sMRI scan. By selecting voxels labeled as Hippocampal region in AAL and computing their 3D bounding boxes we get a sub-volume of the whole 3D scan which encircles Hippocampus. Hippocampus is a symmetrical anatomical structure in the brain consisting of two regions. In order to get only one ROI for a given patient we calculate the mean of the two regions when flipping the left hippocampal region on the right. Note that the resolution of normalized sMRI and MD volumes is quite low (121×145×121, see above) thus the hippocampal ROI occupies a small amount of voxels (28×28×28).

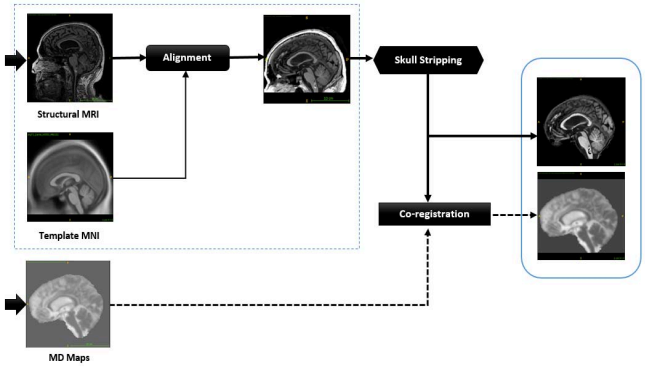

Figure 1: Diagram of the Preprocessing dataset.

B. Data Augmentation and Preparation

In order to reduce overfitting caused by the sparse dataset, this work uses specific data augmentation method to increase the number of samples both for training and validation sets. For general purpose image datasets the method of augmentation consists in applying a number of geometric and colour transformations to the original image data. It mainly includes flip, zoom, shift, scale, contrast, noise, and color modification. In this paper, we increase the number of samples only by shift translation and a weak Gaussian blurring with randomly selected, in a limited range, parameters applied to the whole brain [22]. Then the ROI is extracted. We divide the data into Training, Validation, and Test subsets. For the Test set, we select only subjects with both modalities from ADNI Go and ADNI 2. The remaining scans from ADNI Go and ADNI 2 are added to ADNI 1 and split into training and validation subsets as reported in the Table II.

	AD	**MCI**	**NC**	**AD** [a]	**MCI** [a]	**NC** [a]
Train	196	443	241	44300	44300	44300
	*20	*80	*30	8000	8000	8000
Valid	20	42	24	4200	4200	4200
	*8	*8	*8	800	800	800
Test	20	20	20	20	20	20
	*20	*20	*20	20	20	20

Table II: Number of subjects for each class, with its corresponding augmentation, ([a] after augmentation, * Both modalities).

C. Network Architecture and the 2D+ϵ Approach

The choice of the depth of the CNN architecture depends on the input patches. In [9] we proposed a so-called 2D+ϵ approach fusing the data from different projections classified with a relatively shallow CNN. Indeed, due to the low resolution (28×28) of the ROI, the optimal architecture consisted of two convolutional layers followed by max pooling layer for each one, and a fully connected layer. After the extraction process of the 3D bounding box as presented in section III-A, the median slice and its closest neighbors were selected from the ROI volume and considered as a three channel image. Moreover, we get three networks associated with each modality, the Sagittal, Axial, and Coronal projections. Different tested fusion schemes resulted in application of the (best) late fusion with majority vote fusion operator [23] on the six binary classification tasks AD vs NC, NC vs MCI, and MCI vs AD on three projections. We use the same fusion scheme in the present work.

D. Transfer learning

From the domain knowledge of medical research, we retain that the shrinkage of hippocampal ROI which accompanies the development of AD is observable on both modalities: sMRI and MD-DTI. This phenomenon is illustrated in Fig. 2. Therefore, classification models trained in source domain of sMRI can be adapted for classification tasks on MD. Hence, the transfer learning strategy can be designed from one modality to another. In the framework of learning of CNN parameters we can formally define the transfer learning strategy from the source modality (sMRI) to target MD as follows:

$$\begin{cases} \boldsymbol{W}_0 & \leftarrow \boldsymbol{W}'_\phi \\ \boldsymbol{W}_{i+1} & \leftarrow F(\boldsymbol{W_i}) \end{cases} \tag{1}$$

where $\boldsymbol{W}'_\phi$ is the best trained model on the large sMRI dataset, we initialize the training with the parameters of $\boldsymbol{W}'_\phi$, and fine-tune all the layers of the architecture. F is the optimization scheme which will be detailed in IV-A2. The transfer of the parameters in our CNN architecture is illustrated in Fig. 3. The arrows depict initialization

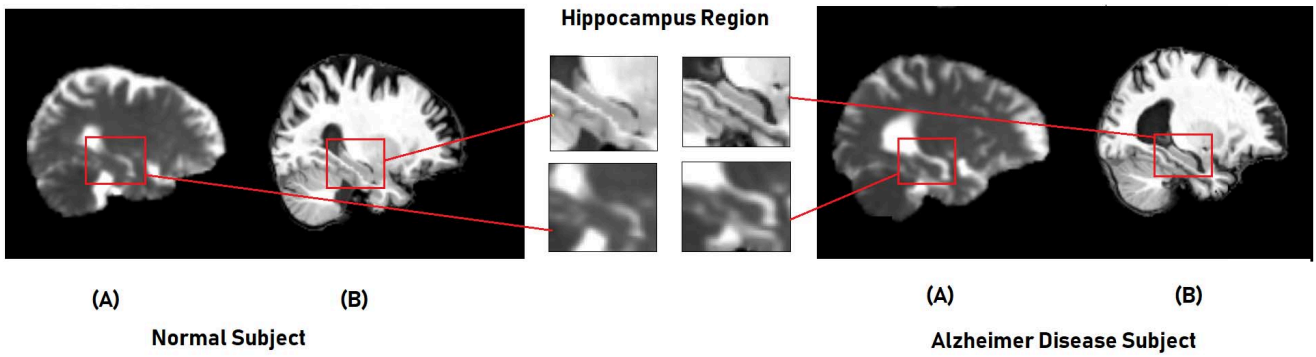

Figure 2: Example: Mean Diffusivity map (A) and Structured MRI (B) of: Left a Normal Subject, and Right: AD Patient. Image examples from ADNI dataset.

of optimization process for each convolutional and fully connected layer.

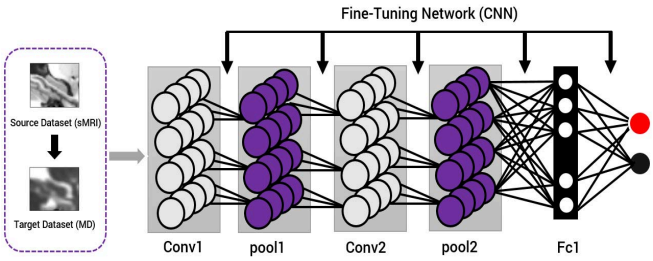

Figure 3: The scheme of Transfer Learning parameters from sMRI to MD imaging.

IV. Experiments and Results

A. Implementation

1) Data augmentation parameters: For the methods of data augmentation we used, see section III-B, augmentation factor was set to 100. Max shift translation was set to 2 (two slices of Sagittal axis represent about 7.4% of the Hippocamus 3D Bounding Box), and maximum scale parameter of smoothing Gaussian Blur was chosen 1.2. Indeed, the original signal on both modalities is blurred and stronger blurring destroy structure of the ROI. Table II describes the split of samples before and after the augmentation process. Note that test subset is not augmented and one sample represents 2.5% of the total samples for a binary classification task.

2) Training of Network parameters : The network was implemented with "Caffe" Deep learning Framework [24]. We used the Stochastic Gradient Descent with Nesterov momentum [25] as optimization method (solver) to minimize the cost function $\mathcal{L}(\boldsymbol{W}_i)$. The weights update formulas are as follows:

$$\begin{aligned} \boldsymbol{V}_{i+1} &\leftarrow \mu \boldsymbol{V}_i - \alpha \nabla \mathcal{L}(\boldsymbol{W}_i + \mu \boldsymbol{V}_i) \\ \boldsymbol{W}_{i+1} &\leftarrow \boldsymbol{W}_i + \boldsymbol{V}_{i+1} \end{aligned} \tag{2}$$

where $\boldsymbol{W}_i$ are the parameters of each layer at iteration i, α is the learning rate, μ the momentum ($\mu = 0.9$) and $\boldsymbol{V}_i$ is the velocity.

We use the exponential learning rate decay policy, $\alpha = \alpha_0 \cdot \gamma^i$, where α_0 is the initial value of learning rate, and $\gamma \in [0, 1]$. In our case we set γ to 0.95, and α_0 = 0.0001. We set the batch size as 128. After several iterations stabilization of the training is observed around the 30th epoch.

The experiments were conducted on a GPU-based high-performance computing platform comoposed of Intel(R) i7-6950X CPU@3.00GHz, equipped with three Nvidia GeForce GPU TITAN-X 12GB, and 128GB of RAM. The computational time for one epoch at the training step with batch-size of 128 samples was 5.86 seconds in average.

B. Evaluation of transfer learning

For a comparison purpose, we demonstrate the effects of pre-training the model on the classification of AD vs. NC as an example on the MD dataset. Two experiments were conducted. In the first experiment, when parameter values of the network were randomly initialized, we obtain the over-fitting phenomena. Fig. 4 depicts the curve of validation loss. After the second epoch, the values of the cost function increase, and therefore the generalization gap increases very quickly. In the second experiment, we tackle the over-fitting problem, by reusing the best trained parameters in the same binary classification on the sMRI dataset for initialization of parameters on MD. The Fig. 5, illustrates the improved behavior. Here the loss value at the first epoch is lower (0.4441 vs 0.4831). Furthermore the generalization gap between temporal mean of loss on training and on validation datasets also decreases. After 10 epochs, it is of : 0.1975 and 0.3453 respectively. Furthermore it increases more slowly.

Without transfer learning after 30 epochs loss values are of 0.4733 and with transfer learning 0.3197.

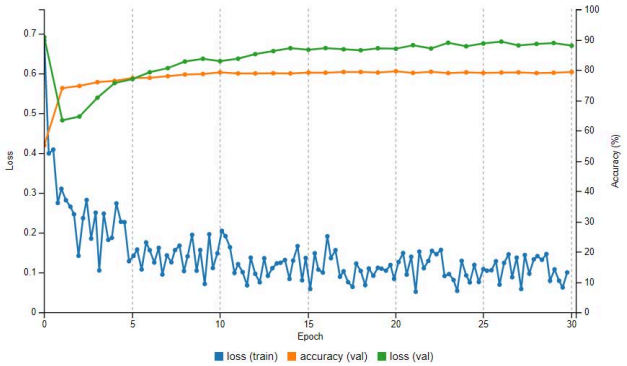

Figure 4: AD vs. NC : Over-fitting graph.

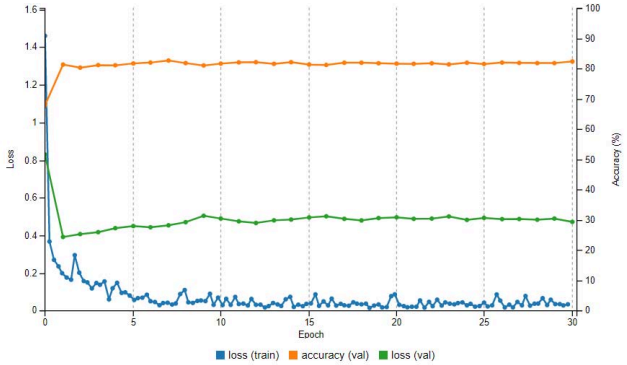

Figure 5: AD vs. NC : Normal fitting graph.

C. Results

The performance of the proposed method was evaluated on single projection classification tasks: AD vs. NC, AD vs. MCI, and MCI vs. NC. We train the networks for each projection: the Sagittal, Axial, and Coronal on the sMRI dataset, then the transfer learning on the MD dataset is employed. In the table III, are given the accuracy values on validation set obtained at 20th epoch, separately for each projection of the two modalities. MD modality gives slightly better results in AD vs NC and AD vs. MCI. The Axial projection still shows a weakness in the prediction comparing to other projections. We realize that the most difficult classification tasks are AD vs. MCI, and MCI vs. NC. Indeed MCI class encompasses the early MCI and the late MCI subjects. Some of them can be considered at a high risk of conversion to AD, it is the case of LMCI. They are hardly distinguishable even for medical experts. These results are confirmed in Table IV with reported figures on test dataset. We are comparing here the result of single-modality classification and of fusion. Sensitivity (SEN), specificity (SPE) and balanced accuracy (BAC) metrics of the binary classification tasks in case of MCI class are also lower than for AD vs NC, which is expectable. Fusion increases performance.

Tasks	Projection	AD vs. NC	AD vs. MCI	MCI vs. NC
SMRI	Sagittal	83.66%	67.95%	64.54%
	Axial	80.90%	68.30%	67.73%
	Coronal	82.39%	70.57%	**70.19%**
MD	Sagittal	**84.20%**	66.04%	60.30%
	Axial	69.71%	63.04%	66.40%
	Coronal	80.46%	**72.17%**	69.53%

Table III: The Accuracy for each single projection at the validation step.

Tasks	Method	ACC (%)	SEN (%)	SPE (%)	BAC (%)
AD vs. NC	SMRI*	90.0%	94.4%	86.3%	90.3%
	MD*	85.0%	88.8%	81.8%	85.3%
	Fusion	**92.5%**	94.7%	90.4%	92.5%
AD vs. MCI	SMRI*	82.5%	88.2%	78.2%	83.2%
	MD*	82.5%	93.3%	76.0%	84.6%
	Fusion	**85.0%**	93.7%	79.1%	86.4%
MCI vs. NC	SMRI*	72.5%	80.0%	68.0%	74.0%
	MD*	62.5%	66.6%	60.0%	63.3%
	Fusion	**80.0%**	92.8%	73.0%	82.9%

Table IV: Results on test dataset (* results of the best single projection).

V. Conclusion and Perspective

In this paper, we have proposed a classification approach based on cross-modal transfer learning for AD diagnosis using DTI-MD and sMRI brain images. The combination of multi CNNs by fusion two modalities demonstrates the effectiveness of the proposed method. Our method achieved classification accuracies of 92.5%, 85.0%, and 80.0% respectively for AD vs. NC, Ad vs. MCI, and MCI vs. NC. The obtained results are encouraging and open interesting perspectives. In the future work, we will proceed with a longitudinal datasets "3C Bordeaux Dataset", and develop a method based on the spatial optimization of the ROI.

Acknowledgment

This research was supported by TOUBKAL French-Morocco research grant AlClass and Ostrogradsky scholarship grant 2017. We thank Dr. Pierrick Coupé from LABRI UMR 5800 University of Bordeaux/CNRS/Bordeaux-INP who provided insights and expertise that greatly assisted the research. Data used in the preparation of this article were obtained from the Alzheimers disease Database Initiative (ADNI) database. A complete listing of ADNI investigators can be found at: http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Autorship_list.pdf

REFERENCES

[1] "Alzheimers Disease International (AZ) World Alzheimer Report 2016, London UK : Alzheimers Disease International," https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf/, 2016, [Accessed December 12, 2017].

[2] G. B. Frisoni, N. C. Fox, C. R. Jack Jr, P. Scheltens, and P. M. Thompson, "The clinical use of structural mri in alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, p. 67, 2010.

[3] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, pp. 239–259, 1991.

[4] C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. OBrien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of ad with mri-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1397, 1999.

[5] D. Le Bihan and H. Johansen-Berg, "Diffusion mri at 25: exploring brain tissue structure and function," *Neuroimage*, vol. 61, no. 2, pp. 324–341, 2012.

[6] D. Le Bihan, "Looking into the functional architecture of the brain with diffusion mri," *Nature Reviews Neuroscience*, vol. 4, no. 6, p. 469, 2003.

[7] M. M. Mielke, N. Kozauer, K. Chan, M. George, J. Toroney, M. Zerrate, K. Bandeen-Roche, M.-C. Wang, J. Pekar, S. Mori *et al.*, "Regionally-specific diffusion tensor imaging in mild cognitive impairment and alzheimer's disease," *Neuroimage*, vol. 46, no. 1, pp. 47–55, 2009.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[9] K. Aderghal, M. Boissenin, J. Benois-Pineau, G. Catheline, and K. Afdel, "Classification of smri for ad diagnosis with convolutional neuronal networks: A pilot 2d+ϵ study on adni," in *International Conference on Multimedia Modeling*. Springer, 2017, pp. 690–701.

[10] A. Pelletier, O. Periot, B. Dilharreguy, B. Hiba, M. Bordessoules, K. Pérès, H. Amieva, J.-F. Dartigues, M. Allard, and G. Catheline, "Structural hippocampal network alterations during healthy aging: a multi-modal mri study," *Frontiers in aging neuroscience*, vol. 5, p. 84, 2013.

[11] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[12] T. D. Phong, H. N. Duong, H. T. Nguyen, N. T. Trong, V. H. Nguyen, T. Van Hoa, and V. Snasel, "Brain hemorrhage diagnosis by using deep learning," in *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*. ACM, 2017, pp. 34–39.

[13] S. Wang, Y. Shen, W. Chen, T. Xiao, and J. Hu, "Automatic recognition of mild cognitive impairment from mri images using expedited convolutional neural networks," in *International Conference on Artificial Neural Networks*. Springer, 2017, pp. 373–380.

[14] M. A. Nowrangi, C. G. Lyketsos, J.-M. S. Leoutsakos, K. Oishi, M. Albert, S. Mori, and M. M. Mielke, "Longitudinal, region-specific course of diffusion tensor imaging measures in mild cognitive impairment and alzheimers disease," *Alzheimer's & dementia: the journal of the Alzheimer's Association*, vol. 9, no. 5, pp. 519–528, 2013.

[15] B. Cheng, M. Liu, D. Shen, Z. Li, D. Zhang, A. D. N. Initiative *et al.*, "Multi-domain transfer learning for early diagnosis of alzheimers disease," *Neuroinformatics*, vol. 15, no. 2, pp. 115–132, 2017.

[16] T. Glozman and O. Liba, "Hidden cues: Deep learning for alzheimers disease classification cs331b project final report."

[17] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, C. Plant, A. Oswald, T. Meindl, M. Pievani, A. L. Bokde, A. Fellgiebel *et al.*, "Robust automated detection of microstructural white matter degeneration in alzheimers disease using machine learning classification of multicenter dti data," *PloS one*, vol. 8, no. 5, p. e64925, 2013.

[18] G. Frisoni, C. Testa, F. Sabattoli, A. Beltramello, H. Soininen, and M. Laakso, "Structural correlates of early and late onset alzheimers disease: voxel based morphometric study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. 1, pp. 112–114, 2005.

[19] O. B. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, G. Catheline, A. D. N. Initiative *et al.*, "Classification of alzheimers disease subjects from mri using hippocampal visual features," *Multimedia Tools and Applications*, vol. 74, no. 4, pp. 1249–1266, 2015.

[20] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[21] A. Cherubini, P. Péran, I. Spoletini, M. Di Paola, F. Di Iulio, G. E. Hagberg, G. Sancesario, W. Gianni, P. Bossu, C. Caltagirone *et al.*, "Combined volumetry and dti in subcortical structures of mild cognitive impairment and alzheimer's disease patients," *Journal of Alzheimer's Disease*, vol. 19, no. 4, pp. 1273–1282, 2010.

[22] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3d cnn-based classification using smri and md-dti images for alzheimer disease studies," *arXiv preprint arXiv:1801.05968*, 2018.

[23] K. Aderghal, J. Benois-Pineau, K. Afdel, and G. Catheline, "Fuseme: Classification of smri images by fusion of deep cnns in 2d+ϵ projections," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*. ACM, 2017, p. 34.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[25] Y. Nesterov, "A method of solving a convex programming problem with convergence rate o (1/k2)," in *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983, pp. 372–376.