bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI

Saman Sarraf[a,b,*], Ghassem Tofighi[c], for the Alzheimer's Disease Neuroimaging Initiative ⋆⋆

[a] *Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada*
[b] *Rotman Research Institute at Baycrest, Toronto, Ontario, Canada*
[c] *Electrical and Computer Engineering Department, Ryerson University, Toronto, Ontario, Canada*

Abstract

To extract patterns from neuroimaging data, various techniques, including statistical methods and machine learning algorithms, have been explored to ultimately aid in Alzheimer's disease diagnosis of older adults in both clinical and research applications. However, identifying the distinctions between Alzheimer's brain data and healthy brain data in older adults (age > 75) is challenging due to highly similar brain patterns and image intensities. Recently, cutting-edge deep learning technologies have been rapidly expanding into numerous fields, including medical image analysis. This work outlines state-of-the-art deep learning-based pipelines employed to distinguish Alzheimer's magnetic resonance imaging (MRI) and functional MRI data from normal healthy control data for the same age group. Using these pipelines, which were executed on a GPU-based high performance computing platform, the data were strictly and carefully preprocessed. Next, scale and shift invariant low- to high-level features were obtained from a high volume of training images using convolutional neural network (CNN) architecture. In this study, functional MRI data were used for the first time in deep learning applications for the purposes of medical image analysis and Alzheimer's disease prediction. These proposed and implemented pipelines, which demonstrate a significant improvement in classification output when compared to other studies, resulted in high and reproducible accuracy rates of 99.9% and 98.84% for the fMRI and MRI pipelines, respectively.

Keywords: Deep Learning, Alzheimer’s Disease, MRI, FMRI

1. Introduction

1.1. Alzheimer's Disease

Alzheimer's disease (AD) is an irreversible, progressive neurological brain disorder and multifaceted disease that slowly destroys brain cells, causing memory and thinking skill losses, and ultimately loss of the ability to carry out even the simplest tasks. The cognitive decline caused by this disorder ultimately leads to dementia. For instance, the disease begins with mild deterioration and grows progressively worse as a neurodegenerative type of dementia. Diagnosing Alzheimer's disease requires very careful medical assessment, including patient history, a mini mental state examination (MMSE), and physical and neurobiological exams (Vemuri et al., 2012)(He et al., 2007). In addition to these evaluations, structural magnetic resonance imaging and resting state functional magnetic resonance imaging (rs-fMRI) offer non-invasive methods of studying the structure of the brain, functional brain activ-

*Corresponding author: samansarraf@ieee.org

⋆⋆Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigator within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ity, and changes in the brain. During scanning using both structural (anatomical) and rs-fMRI techniques, patients remain prone on the MRI table and do not perform any tasks. This allows data acquisition to occur without any effects from a particular task on functional activity in the brain (Sarraf and Sun, 2016)(Grady et al., 2016)(Saverino et al., 2016). Alzheimer's disease causes shrinkage of the hippocampus and cerebral cortex and enlargement of ventricles in the brain. The level of these effects is dependent upon the stage of disease progression. In the advanced stage of AD, severe shrinkage of the hippocampus and cerebral cortex, as well as significantly enlarged ventricles, can easily be recognized in MR images. This damage affects those brain regions and networks related to thinking, remembering (especially short-term memory), planning and judgment. Since brain cells in the damaged regions have degenerated, MR image (or signal) intensities are low in both MRI and rs-fMRI techniques (Warsi, 2012)(Grady et al., 2003)(Grady et al., 2001). However, some of the signs found in the AD imaging data are also identified in normal aging imaging data. Identifying the visual distinction between AD data and images of older subjects with normal aging effects requires extensive knowledge and experience, which must then be combined with additional clinical results in order to accurately classify the data (i.e., MMSE) (Vemuri et al., 2012). Development of an assistive tool or algorithm to classify MR-based imaging data, such as structural MRI and rs-fMRI data, and, more importantly, to distinguish brain disorder data from healthy subjects, has always been of interest to clinicians (Tripoliti et al., 2008). A robust machine learning algorithm such as Deep Learning, which is able to classify Alzheimer's disease, will assist scientists and clinicians in diagnosing this brain disorder and will also aid in the accurate and timely diagnosis of Alzheimer's patients (Raventós and Zaidi).

1.2. Deep Learning

Hierarchical or structured deep learning is a modern branch of machine learning that was inspired by the human brain. This technique has been developed based upon complicated algorithms that model high-level features and extract those abstractions from data by using similar neural network architecture that is actually much more complicated. Neuroscientists have discovered that the neocortex, which is a part of the cerebral cortex concerned with sight and hearing in mammals, processes sensory signals by propagating them through a complex hierarchy over time. This served as the primary motivation for the development of deep machine learning that focuses on computational models for information representation which exhibits characteristics similar to those of the neocortex (Jia et al., 2014) (Ngiam et al., 2011). Convolutional neural networks (CNNs) that are inspired by the human visual system are similar to classic neural networks. This architecture has been specifically designed based on the explicit assumption that raw data are comprised of two-dimensional images that enable certain properties to be encoded while also reducing the amount of hyper parameters. The topology of CNNs utilizes spatial relationships to reduce the number of parameters that must be learned, thus improving upon general feed-forward backpropagation training (Erhan et al., 2010) (Schmidhuber, 2015). Equation 1 demonstrates how the gradient component for a given weight is calculated in the backpropagation step, where E is error function, y is the neuron $N_{i,j}$, x is the input, l represents layer numbers, w is filter weight with a and b indices, N is the number of neurons in a given layer, and m is the filter size.

$$\frac{\partial E}{\partial \omega_{ab}} = \sum_{i=0}^{N-m}\sum_{j=0}^{N-m}\frac{\partial E}{\partial x_{ij}^{\ell}}\frac{\partial x_{ij}^{\ell}}{\partial \omega_{ab}} = \sum_{i=0}^{N-m}\sum_{j=0}^{N-m}\frac{\partial E}{\partial x_{ij}^{\ell}} y_{(i+a)(j+b)}^{\ell-1} \quad (1)$$

As shown in Equation 1: In convolutional layers, the gradient component of a given weight is calculated by applying the chain rule. Partial derivatives of the error for the cost function with respect to the weight are calculated and used to update the weight.

Equation 2 describes the backpropagation error for the previous layer using the chain rule. This equation is similar to the convolution definition, where $x_{(i+a)(j+b)}$ is replaced by $x_{(i-a)(j-b)}$. It demonstrates the backpropagation results in convolution while the weights are rotated. The rotation of the weights derives from a delta error in the convolutional neural network.

$$\frac{\partial E}{\partial y_{ij}^{\ell-1}} = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1}\frac{\partial E}{\partial x_{(i-a)(j-b)}^{\ell}}\frac{\partial x_{(i-a)(j-b)}^{\ell}}{\partial y_{ij}^{\ell-1}} = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1}\frac{\partial E}{\partial x_{(i-a)(j-b)}^{\ell}}\omega_{ab} \quad (2)$$

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The error of backpropagation for the previous layer (Equation 2) is calculated using the chain rule. This equation is similar to the definition of convolution, but the weights are rotated.

In CNNs, small portions of the image (called local receptive fields) are treated as inputs to the lowest layer of the hierarchical structure. One of the most important features of CNNs is that their complex architecture provides a level of invariance to shift, scale and rotation, as the local receptive field allows the neurons or processing units access to elementary features, such as oriented edges or corners. This network is primarily comprised of neurons having learnable weights and biases, forming the convolutional layer. It also includes other network structures, such as a pooling layer, a normalization layer and a fully connected layer. As briefly mentioned above, the convolutional layer, or conv layer, computes the output of neurons that are connected to local regions in the input, each computing a dot product between its weight and the region it is connected to in the input volume. The pooling layer, also known as the pool layer, performs a downsampling operation along the spatial dimensions. The normalization layer, also known as the rectified linear units (ReLU) layer, applies an elementwise activation function, such as max (0, x) thresholding at zero. This layer does not change the size of the image volume (LeCun et al., 1998) (Jia et al., 2014) (Arel et al., 2010). The fully connected (FC) layer computes the class scores, resulting in the volume of the number of classes. As with ordinary neural networks, and as the name implies, each neuron in this layer is connected to all of the numbers in the previous volume (Jia et al., 2014) (Szegedy et al., 2015). The convolutional layer plays an important role in CNN architecture and is the core building block in this network. The conv layer's parameters consist of a set of learnable filters. Every filter is spatially small but extends through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, producing a 2D activation map of that filter. During this convolving, the network learns of filters that activate when they see some specific type of feature at some spatial position in the input. Next, these activation maps are stacked for all filters along the depth dimension, which forms the full output volume. Every entry in the output volume can thus also be interpreted as an output from a neuron that only examines a small region in the input and shares parameters with neurons in the same activation map (Jia et al., 2014) (Wang et al., 2015). A pooling layer is usually inserted between successive conv layers in CNN architecture. Its function is to reduce (down sample) the spatial size of the representation in order to minimize network hyper parameters, and hence also to control overfitting. The pooling layer operates independently on every depth slice of the input and resizes it spatially using the max operation (LeCun et al., 1998) (Jia et al., 2014) (Arel et al., 2010) (Szegedy et al., 2015) (Wang et al., 2015). In convolutional neural network architecture, the conv layer can accept any image (volume) of size $W_1 \times H_1 \times D_1$ that also requires four hyper parameters, which are K, number of filters; F, their spatial extent; S, the size of stride; and P, the amount of zero padding. The conv layer outputs the new image, whose dimensions are $W_2 \times H_2 \times D_2$, calculated as Equation 3. An understanding of how the conv layer produces new output images is important to realize the effect of filters and other operators, such as stride (S), on input images.

$$\begin{aligned} W_2 &= (W_1 - F)/S + 1 \\ H_2 &= (H_1 - F)/S + 1 \\ D_2 &= D_1 \end{aligned} \tag{3}$$

LeNet-5 was first designed by Y. LeCun et al. (LeCun et al., 1998). This architecture successfully classified digits and was applied to hand-written check numbers. The application of this fundamental but deep network architecture expanded into more complicated problems by adjusting the network hyper parameters. LeNet-5 architecture, which extracts low- to mid-level features, includes two conv layers, two pooling layers, and two fully connected layers, as shown in Figure 1. More complex CNN architecture was developed to recognize numerous objects derived from high volume data, including AlexNet (ImageNet) (Krizhevsky et al., 2012), ZF Net (Lowe, 2004), GoogleNet (Szegedy et al., 2015) and ResNet (He et al., 2015). GoogleNet, which was developed by Szegedy et al. (Szegedy et al., 2015), is a successful network that is broadly used for object recognition and classification. This architecture is comprised of a deep, 22-layer network based on a modern design module called Incep-

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

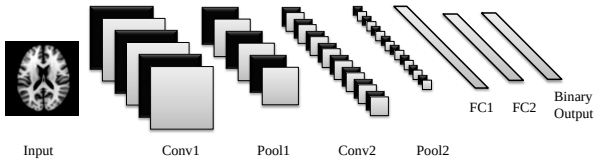

Figure 1: LeNet-5 includes two conv, two pool and two FC layers. The original version of this network classified 10 digits. In this work, the architecture was optimized for binary output, which were Alzheimer's disease (AD) and normal control (NC), respectively.

tion. One of the fundamental approaches to improving the accuracy of CNN architecture is to increase the size of layers. However, this straightforward solution causes two major issues. First, a large number of hyper parameters requires more training data and may also result in overfitting, especially in the case of limited training data. On the other hand, uniform increases in network size dramatically increase interactions with computational resources, which affect the timing performance and the cost of providing infrastructure. One of the optimized solutions for both problems would be the development of a sparsely connected architecture rather than a fully connected network. Strict mathematical proofs demonstrate that the well-known Hebbian principle of neurons firing and wiring together created the Inception architecture of GoogleNet (Szegedy et al., 2015). The Inception module of GoogleNet, as shown in Figure 2, is developed by discovering the optimal local sparse structure to construct convolutional blocks. Inception architecture allows for a significant increase in the number of units at each layer, while computational complexity remains under control at later stages, which is achieved through global dimensionality reduction prior to costly convolutions with larger patch sizes.

1.3. Data Acquisition

In this study, two subsets of the ADNI database (http://adni.loni.usc.edu/) were used to train and validate convolutional neural network classifiers. The first subset included 144 subjects who were scanned for resting-state functional magnetic resonance imaging (rs-fMRI) studies. In this dataset, 52 Alzheimer's patients and 92 healthy control subjects were recruited (age group > 75). The second dataset included 302 subjects whose structural magnetic resonance imaging data (MRI) were acquired (age group > 75). This group included 211 Alzheimer's patients and 91 healthy control subjects. Certain subjects were scanned at different points in time, and their imaging data were separately considered in this work. Table 1 presents the demographic information for both subsets, including mini mental state examination (MMSE) scores. MRI data acquisition was performed according to the ADNI acquisition protocol (Jack et al., 2008). Scanning was performed on three different Tesla scanners, General Electric (GE) Healthcare, Philips Medical Systems, and Siemens Medical Solutions, and was based on identical scanning parameters. Anatomical scans were acquired with a 3D MPRAGE sequence (TR=2s, TE=2.63 ms, FOV=25.6 cm, 256 × 256 matrix, 160 slices of 1mm thickness). Functional scans were acquired using an EPI sequence (150 volumes, TR=2 s, TE=30 ms, flip angle=70, FOV=20 cm, 64 × 64 matrix, 30 axial slices of 5mm thickness without gap).

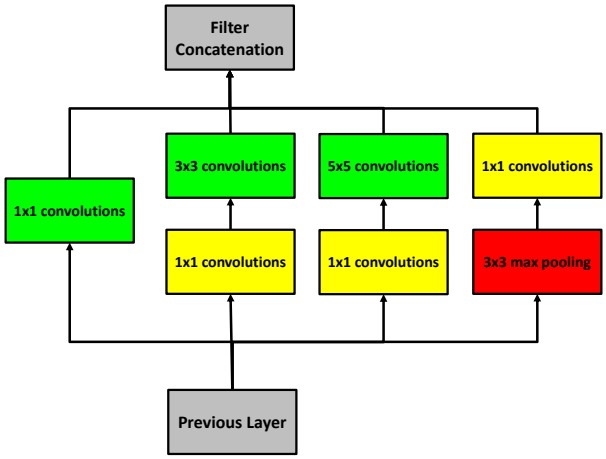

Figure 2: Inception module with dimensionality reduction in GoogleNet architecture

2. Related Work

Changes in brain structure and function caused by Alzheimer's disease have proved of great interest to numerous scientists and research groups. In diagnostic

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1: Two subsets of the ADNI database were used in this study, including 144 subjects with fMRI data and 302 subjects with MRI data. The mean and standard deviation (SD) of age and total MMSE scores per group are delineated in the table below.

Modality	Total Subj.	Group	Subj.	Female	Mean of Age	SD	Male	Mean of Age	SD	MMSE	SD
rs-fMRI	144	Alzheimer	52	21	79.42	16.35	31	80.54	15.98	22.70	2.10
		Control	92	43	80.79	19.16	49	81.75	21.43	28.82	1.35
MRI	302	Alzheimer	211	85	80.98	21.6	126	81.27	16.66	23.07	2.06
		Control	91	43	79.37	12.52	48	80.81	19.51	28.81	1.35

imaging in particular, classification and predictive modeling of the stages of Alzheimer's have been broadly investigated. Suk et al. (Suk and Shen, 2013) (Suk et al., 2015a) (Suk et al., 2015b) developed a deep learning-based method to classify AD magnetic current imaging (MCI) and MCI-converter structural MRI and PET data, achieving accuracy rates of 95.9%, 85.0% and 75.8% for the mentioned classes, respectively. In their approach, Suk et al. developed an auto-encoder network to extract low- to mid-level features from images. Next, classification was performed using multi-task and multi-kernel Support Vector Machine (SVM) learning methods. This pipeline was improved by using more complicated SVM kernels and multimodal MRI/PET data. However, the best accuracy rate for Suk et al. remained unchanged (Suk et al., 2014). Payan et al. of Imperial College London designed (Payan and Montana, 2015) a predictive algorithm to distinguish AD MCI from normal healthy control subjects'imaging. In this study, an auto-encoder with 3D convolutional neural network architecture was utilized. Payan et al. obtained an accuracy rate of 95.39% in distinguishing AD from NC subjects. The research group also tested a 2D CNN architecture, and the reported accuracy rate was nearly identical in terms of value. Additionally, a multimodal neuroimaging feature extraction pipeline for multiclass AD diagnosis was developed by Liu et al. (Liu et al., 2015). This deep-learning framework was developed using a zero-masking strategy to preserve all possible information encoded in imaging data. High-level features were extracted using stacked auto-encoder (SAE) networks, and classification was performed using SVM against multimodal and multiclass MR/PET data. The highest accuracy rate achieved in that study was 86.86%. Aversen et al. (Arvesen, 2015), Liu et al. (Liu and Shen, 2014), Siqi et al. (Liu et al., 2014), Brosch et al. (Brosch et al., 2013), Rampasek et al. (Rampasek and Goldenberg, 2016), De Brebisson et al. (de Brebisson and Montana, 2015) and Ijjina et al. (Ijjina and Mohan, 2015) also demonstrated the application of deep learning in automatic classification of Alzheimer's disease from structural MRI, where AD, MCI and NC data were classified.

3. Methods

Classification of Alzheimer's disease images and normal, healthy images required several steps, from preprocessing to recognition, which resulted in the development of an end-to-end pipeline. Three major modules formed this recognition pipeline: a) preprocessing b) data conversion; and c) classification, respectively. Two different approaches were used in the preprocessing module, as preprocessing of 4D rs-fMRI and 3D structural MRI data required different methodologies, which will be explained later in this paper. After the preprocessing steps, the data were converted from medical imaging to a Portable Network Graphics (PNG) format to input into the deep learning-based classifier. Finally, the CNN-based architecture receiving images in its input layer was trained and tested (validated) using 75% and 25% of the dataset, respectively. In practice, two different pipelines were developed, each of which was different in terms of preprocessing but similar in terms of data conversion and classification steps, as demonstrated in Figure 3.

3.1. rs-fMRI Data Preprocessing

The raw data in DICOM format for both the Alzheimer's (AD) group and the normal control (NC) group were converted to NII format (Neuroimaging Informatics Technology Initiative - NIfTI) using the dcm2nii software package developed by Chris Roden et al. http://www.sph.sc.edu/comd/rorden/mricron/dcm2nii.html. Next, non-brain regions, including skull and neck voxels,

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

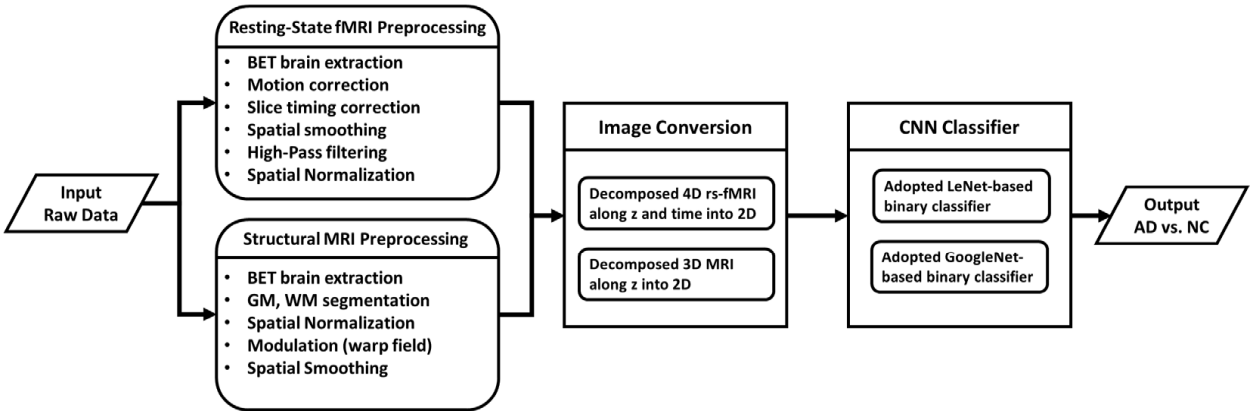

Figure 3: End-to-end recognition based on deep learning CNN classification methods is comprised of three major components: preprocessing, image conversion and classification modules. In the preprocessing step, two different submodules were developed for rs-fMRI and structural data. Next, the image conversion module created PNG images from medical imaging data using the algorithm described in the following section of this paper. The final step was to recognize AD from NC samples using CNN models, which was performed by training and testing models using 75% and 25% of the samples, respectively.

were removed from the structural T1-weighted image corresponding to each fMRI time course using FSL-BET (Smith, 2002). Resting-state fMRI data, including 140 time series per subject, were corrected for motion artefact using FSL-MCFLIRT (Jenkinson et al., 2002), as low frequency drifts and motion could adversely affect decomposition. The next necessary step was the regular slice timing correction, which was applied to each voxels time series because of the assumption that later processing assumes all slices were acquired exactly half-way through the relevant volumes acquisition time (TR). In fact, each slice is taken at slightly different times. Slice timing correction works by using Hanning-windowed Sinc interpolation to shift each time series by an appropriate fraction of a TR relative to the middle of the TR period. Spatial smoothing of each functional time course was then performed using a Gaussian kernel of 5 mm full width at half maximum. Additionally, low-level noise was removed from the data by a temporal high-pass filter with a cut-off frequency of 0.01 HZ (sigma = 90 seconds) in order to control the longest allowed temporal period. The functional images were registered to the individuals high-resolution (structural T1) scan using affine linear transformation with seven degrees of freedom (7 DOF). Subsequently, the registered images were aligned to the MNI152 standard space (average T1 brain image constructed from 152 normal subjects at the Montreal Neurological Institute) using affine linear registration with 12 DOF followed by 4 mm resampling, which resulted in $45x54x45$ images per time course.

3.2. *Structural MRI Data Preprocessing*

The raw data of structural MRI scans for both the AD and the NC groups were provided in NII format in the ADNI database. First, all non-brain tissues were removed from images using Brain Extraction Tool FSL-BET (Smith, 2002) by optimizing the fractional intensity threshold and reducing image bias and residual neck voxels. A study-specific grey matter template was then created using the FSL-VBM library and relevant protocol, found at http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLVBM (Douaud et al., 2007). In this step, all brain-extracted images were segmented to grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). GM images were selected and registered to the GM ICBM-152 standard template using linear affine transformation. The registered images were concatenated and averaged and were then flipped along the x-axis, and the two mirror images were then re-averaged to obtain a first-pass, study-specific affine GM template. Second, the GM images were re-registered to

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

this affine GM template using non-linear registration, concatenated into a 4D image which was then averaged and flipped along the x-axis. Both mirror images were then averaged to create the final symmetric, study-specific non-linear GM template at $2x2x2$ mm3 resolution in standard space. Following this, all concatenated and averaged 3D GM images (one 3D image per subject) were concatenated into a stack (4D image = 3D images across subjects). Additionally, the FSL-VBM protocol introduced a compensation or modulation for the contraction/enlargement due to the non-linear component of the transformation, where each voxel of each registered grey matter image was multiplied by the Jacobian of the warp field. The modulated 4D image was then smoothed by a range of Gaussian kernels, sigma = 2, 3, 4 mm (standard sigma values in the field of MRI data analysis), which approximately resulted in full width at half maximums (FWHM) of 4.6, 7 and 9.3 mm. The various spatial smoothing kernels enabled us to explore whether classification accuracy would improve by varying the spatial smoothing kernels. The MRI preprocessing module was applied to AD and NC data and produced two sets of four 4D images, which were called Structural MRI 0 fully preprocessed without smoothing as well as three fully preprocessed and smoothed datasets called Structural MRI 2, 3, 4, which were used in subsequent classification steps.

4. Results and Discussion

4.1. rs-fMRI Pipeline

The preprocessed rs-fMRI time series data were first loaded into memory using neuroimaging package Nibabel (http://nipy.org/nibabel/) and were then decomposed into 2D (x,y) matrices along z and time (t) axes. Next, the 2D matrices were converted to PNG format using the Python OpenCV (opencv.org). The last 10 slices of each time course were removed since they included no functional information. During the data conversion process, a total of 793,800 images were produced, including 270,900 Alzheimer's and 522,900 normal control PNG samples. In the data conversion step, the 4D time courses of subjects were randomly shuffled, and five random datasets were created in order to repeat training and testing of the CNN classifier (fivefold cross-validation against all of the data). The random datasets were labeled for binary classification, and 75% of the images were assigned to the training dataset, while the remaining 25% were used for testing purposes. The training and testing images were resized to $28x28$ pixels and were then converted to the Lightning Memory-Mapped Database (LMDB) for high throughput for the Caffe Deep Learning platform (Jia et al., 2014) used for this classification experiment. The adopted LeNet architecture was adjusted for 30 epochs and initialized for Stochastic Gradient Descent with $gamma = 0.1, momentum = 0.9, learningrate = 0.01, weight_decay = 0.005$, and the step learning rate policy dropped the learning rate in steps by a factor of gamma every stepsize iteration. The mean of images was calculated and subtracted from each image. Training and testing of Caffe models were performed and were repeated five times on the Amazon AWS Linux G2.8xlarge, including four high-performance NVIDIA GPUs, each with 1,536 CUDA cores and 4GB of video memory and 32 High Frequency Intel Xeon E5-2670 (Sandy Bridge) vCPUs with 60 GB memory overall. An average accuracy rate of 99.9986% was obtained for five randomly shuffled datasets using the adopted LeNet architecture shown in Table 2. Alternatively, the first set of five randomly shuffled datasets was resized to $256x256$ and was then converted to LMDB format. The adopted GoogleNet was adjusted for 30 epochs and initialized with the same parameters mentioned above, and the experiment was performed on the same GPU server. An accuracy testing rate of 100%, as reported in the Caffe log file, was achieved (in practice, Caffe rounded the accuracy up after the seventh decimal), as shown in Table 2. A very high level of accuracy of testing rs-fMRI data was obtained from both of the adopted LeNet and GoogleNet models. During the training and testing processes, the loss of training, loss of testing and accuracy of testing data were monitored. In Figures 4 and 5, the accuracy of testing and the loss of testing of the first randomly shuffled dataset are presented for the adopted LeNet and GoogleNet models, respectively.

4.2. Structural MRI Pipeline

The preprocessed MRI data were then loaded into memory using a similar approach to the fMRI pipeline and were converted from NII to PNG format using Nibabel and OpenCV, which created two groups (AD and NC) four preprocessed datasets (MRI 0,2,3,4). Additionally, the last 10 slices of subjects, as well as slices with zero mean pixels, were removed from the data. This

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

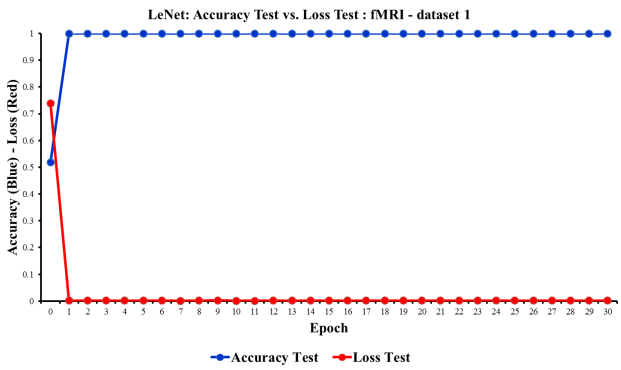

Figure 4: The accuracy and loss of the first testing dataset are shown over 30 epochs. As seen, the accuracy of testing data reached almost 99.99%, and the loss of testing data dropped down to zero in the LeNet classifier.

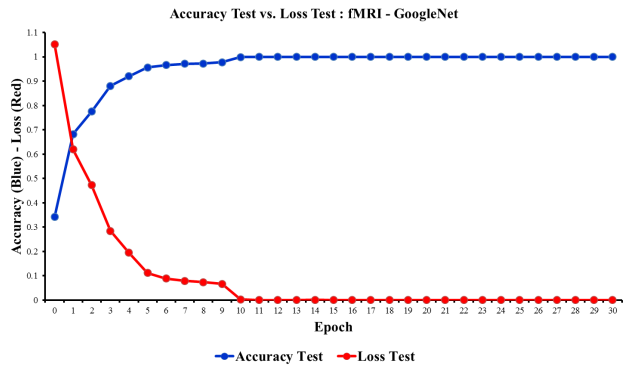

Figure 5: Adopted GoogleNet training and testing resulted in a very high level of accuracy of almost 100%. As seen, the loss of testing approached zero in the 10th epoch. The accuracy rates of both the LeNet and GoogleNet networks were close. However, the final accuracy of GoogleNet was slightly better than the LeNet model.

step produced a total number of 62,335 images, with 52,507 belonging to the AD group and the remaining 9,828 belonging to the NC group per dataset. The data were next converted to the LMDB format and resized to 28x28 pixels. The adopted LeNet model was set for 30 epochs and initiated for Stochastic Gradient Descent with a $gamma = 0.1, amomentum = 0.9, abaselearningrate = 0.01, aweight_decay = 0.0005$, and a step learning rate policy dropping the learning rate in steps by a factor of gamma every stepsize iteration. Next, the model was trained and tested by 75% and 25% of the data for four different datasets. The training and testing processes were repeated five times on Amazon AWS Linux G2.8xlarge to ensure the robustness of the network and achieved accuracy. The average of accuracies was obtained for each experiment separately, as shown in Table 2. The results demonstrate that a high level of accuracy was achieved in all of the experiments, with the highest accuracy rate of 98.79% achieved for the structural MRI dataset, which was spatially smoothed by $sigma = 3mm$. In the second run, the adopted GoogleNet model was selected for binary classification. In this experiment, the preprocessed datasets were converted to LMDB format and resized to $256x256$. The model was adjusted for 30 epochs using Stochastic Gradient Descent with a $gamma = 0.1, amomentum = 0.9, abaselearningrate = 0.01, aweight_decay = 0.0005$, and a step learning rate policy. The GoogleNet model resulted in a higher level of accuracy than the LeNet model, with the highest overall accuracy rate of 98.8431% achieved for MRI 3 (smoothed by sigma = 3mm). However, the accuracy rate of the unsmoothed dataset (MRI 0) reached 84.5043%, which was lower than the similar experiment with the LeNet model. This result may demonstrate the negative effect of interpolation on unsmoothed data, which may in turn strengthen the concept of spatial smoothing in MRI data analysis. In practice, most classification questions address imbalanced data, which refers to a classification problem in which the data are not represented equally and the ratio of data may exceed 4 to 1 in binary classification. In the MR analyses performed in this study, the ratio of AD to NC images used for training the CNN classifier was around 5 to 1. To validate the accuracy of the models developed, a new set of training and testing was performed by randomly selecting and decreasing the number of AD images to 10,722 for training, while the same number of images 9,828 was used for the NC group. In the balanced data experiment, the adopted LeNet model was adjusted for 30 epochs using the same parameters mentioned above and was trained for four MRI datasets. In Table 2, the new results are identified with labels beginning with the B. prefix (Balanced). The highest accuracy rate obtained from the balanced data experiment only decreased around 1% (B. Structural MRI 3 = 97.81%) compared to the same datasets in the original training. This comparison demonstrates that the new

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

results were highly correlated to the initial results, confirming that even a precipitous decrease in the data ratio from 5:1 to 1:1 had no impact on classification accuracy, which validated the robustness of the trained models in the original MRI classification.

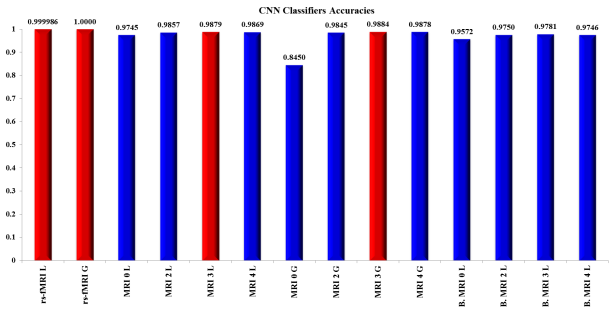

Figure 6: A total of 14 values, including five averaged accuracies and nine single accuracies from a total of 29 training CNN-based classifiers (adopted LeNet and GoogleNet), are demonstrated. Almost perfect accuracy was achieved using the ADNI fMRI data from both models. Additionally, the ADNI MRI data were successfully classified with an accuracy rate approaching 99%. These results demonstrate that CNN-based classifiers are highly capable of distinguishing between AD and NC samples by creating low- to high-level shift and scale invariant features. The results also demonstrate that in MRI classification, spatially smoothed data with sigma = 3 mm produced the highest accuracy rates.(L:LeNet, G:GoogleNet)

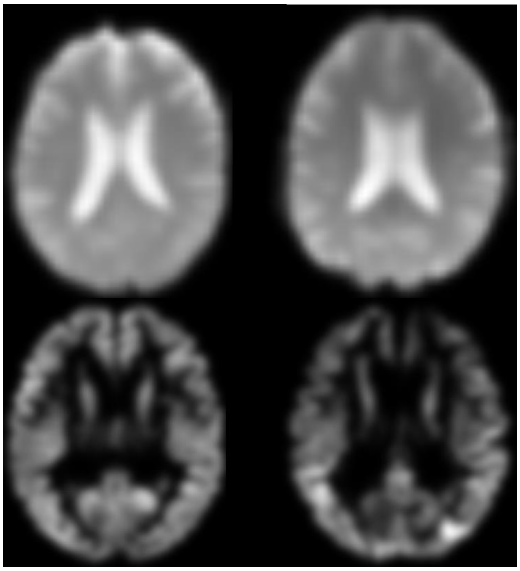

Figure 7: A middle cross-section of fMRI data (22, 27, 22) with thinness of 4 mm, representing a normal healthy brain (top-right), and an Alzheimer's brain are shown (top-left). A middle cross-section of structural MRI (45, 55, 45) with thickness of 2 mm, representing a normal brain (bottom-left), and an Alzheimer's subject (bottom-right) are also demonstrated. In both fMRI and MRI modalities, different brain patterns and signal intensities are identified.

Differentiation between subjects with Alzheimer's disease and normal healthy control subjects (older adults) requires solid preprocessing and feature learning, which reveal functional and structural dissimilarities between Alzheimer's damage and routine effects of age on the brain. In this study, two robust pipelines were designed and implemented that were capable of producing consistent and reproducible results. In the first block of the pipelines, extensive data preprocessing was performed against fMRI and MRI data, which removed potential noise and artefacts from the data. Next, a convolutional layer of CNN architecture consisting of a set of learnable filters, and which also serves as a shift and scale invariant operator, extracted low- to mid-level features (as well as high-level features in GoogleNet). In the fMRI pipeline, both adopted LeNet and GoogleNet architecture were trained and tested by a massive number of images created from 4D fMRI time series. Furthermore, removal of non-functional brain images from data improved the accuracy of recognition when compared to previous experience (Sarraf and Tofighi, 2016). In the MRI pipeline, four sets of images (smoothed with different kernels) were used to train and test the CNN classifier to ensure that the best preprocessed data were employed to achieve the most accurate trained model. The results demonstrate that spatial smoothing with an optimal kernel size improves classification accuracy (Figure 6). Certain differences in image intensity (Figure 7), brain size of AD and NC subjects, and lack of signals in brain regions of AD samples, such as the frontal lobe, are strong evidence in support of the success of the pipelines.

A common strategy employed is to visualize the weights of filters to interpret the conv layer results. These are usually most interpretable on the first conv layer, which directly examines the raw pixel data, but it is also possible to find the filter weights deeper in the network. In a well-trained network, smooth filters without noisy patterns are usually discovered. A smooth pattern without noise is an indicator that the training process is suf-

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2: The accuracy of testing datasets is demonstrated below. As shown, a very high level of accuracy in testing datasets was achieved in both fMRI and MRI modalities in all of the runs. The experiment of the cells with asterisks * was not required in this study. Therefore, no value was assigned. The datasets used for testing balanced data begin with the prefix B. Abbreviation: MRI 0, the structural MRI dataset without spatial smoothing. MRI 2,3,4 are the datasets spatially smoothed by Gaussian kernel sigma = 2,3 and 4 mm.

		Accuracy of Testing per Experiment (out of 1)					
Dataset	Architecture	1	2	3	4	5	Average
resting-state fMRI	Adopted LeNet	0.99999	1	0.99998	0.99997	0.99999	0.999986
	Adopted GoogleNet	1	*	*	*	*	1
Structural MRI 0	Adopted LeNet	0.9755	0.9732	0.9746	0.9737	0.9753	0.97446
Structural MRI 2		0.9851	0.9874	0.9849	0.9848	0.9861	0.98566
Structural MRI 3		0.9862	0.9874	0.9885	0.9889	0.9885	0.9879
Structural MRI 4		0.9875	0.9864	0.9864	0.986	0.9873	0.98672
Structural MRI 0	Adopted GoogleNet	0.845043	*	*	*	*	0.845043
Structural MRI 2		0.98452	*	*	*	*	0.98452
Structural MRI 3		0.988431	*	*	*	*	0.988431
Structural MRI 4		0.987758	*	*	*	*	0.987758
B. Structural MRI 0	Adopted LeNet	0.9572	*	*	*	*	0.9572
B. Structural MRI 2		0.975	*	*	*	*	0.975
B. Structural MRI 3		0.9781	*	*	*	*	0.9781
B. Structural MRI 4		0.9746	*	*	*	*	0.9746

ficiently long, and likely no overfitting occurred. In addition, visualization of the activation of the networks features is a helpful technique to explore training progress. In deeper layers, the features become more sparse and localized, and visualization helps to explore any potential dead filters (all zero features for many inputs). Filters and features of the first layer for a given fMRI and MRI trained LeNet model were visualized using an Alzheimer's brain and a normal control brain.

5. Conclusion

In order to distinguish brains affected by Alzheimer's disease from normal healthy brains in older adults, this study presented two robust pipelines, including extensive preprocessing modules and deep learning-based classifiers, using structural and functional MRI data. Scale and shift invariant low- to high-level features were extracted from a massive volume of whole brain data using convolutional neural network architecture, resulting in a highly accurate and reproducible predictive model. In this study, the achieved accuracy rates for both MRI and fMRI modalities, as well as LeNet and GoogleNet state-of-the-art architecture, proved superior to all previous methods employed. Furthermore, fMRI data were used to train a deep learning-based pipeline for the first time. This successful and cutting-edge deep learning-based framework points to a number of applications in classifying brain disorders in both clinical trials and large-scale research studies. This study also demonstrated that the developed pipelines served as fruitful algorithms in characterizing multimodal MRI biomarkers. In conclusion, the proposed methods demonstrate strong potential for predicting the stages of the progression of Alzheimer's disease and classifying the effects of aging in the normal brain.

Figure 8 and Figure 9 demonstrate 20 filters of $5x5$ pixels for fMRI and MRI models, respectively. Additionally, 20 features of $24x24$ pixels in Figure 10 and Figure 11 reveal various regions of the brain that were activated in AD and NC samples.

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

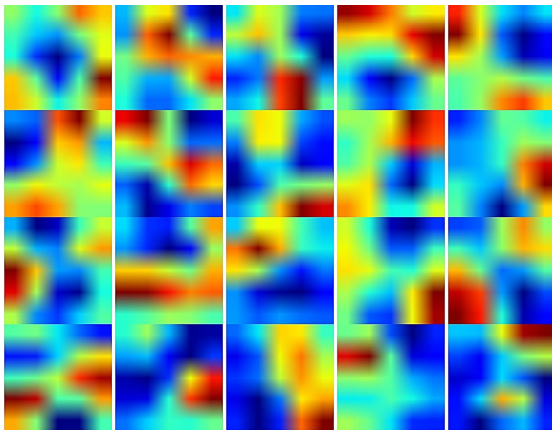

Figure 8: In the first layer of LeNet in a given trained fMRI model, 20 filters of 5x5 pixels were visualized. The weights shown were applied to the input data and produced activation, or features, of a given sample.

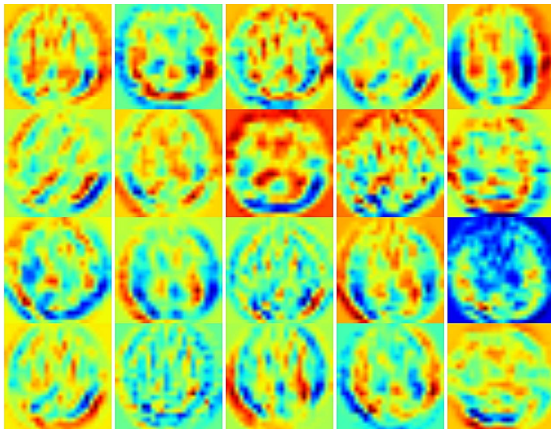

Figure 10: 20 activations (features) of the first layer of LeNet trained using MRI data were displayed for a given AD MRI sample (45, 55, 45). A smooth pattern without noise reveals that the model was successfully trained.

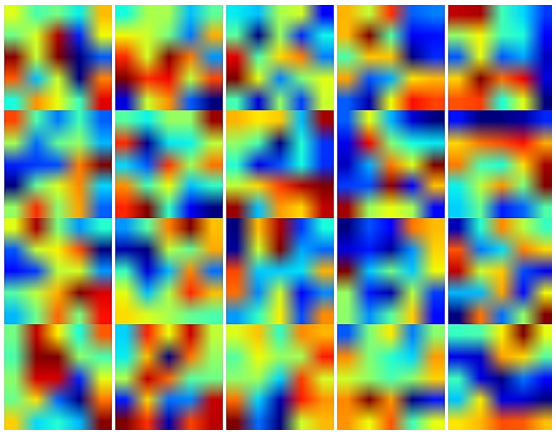

Figure 9: In a trained LeNet model, 20 filters with a kernel of 5x5 were visualized for the first layer. The filters shown were generated from a model in which MRI data smoothed by sigma = 3 mm were used for training.

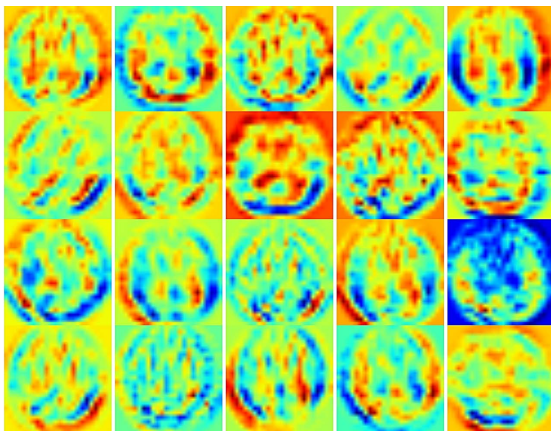

Figure 11: Features of the first layer of the same MRI trained model were displayed for a normal control (NC) brain slice (45, 55, 45). A basic visual comparison reveals significant differences between AD and NC samples.

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

6. Acknowledgments

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.; Cogstate; 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 Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Prashanthi Vemuri, David T Jones, and Clifford R Jack. Resting state functional mri in alzheimer's disease. *Alzheimer's research & therapy*, 4(1):1, 2012.

Yong He, Liang Wang, Yufeng Zang, Lixia Tian, Xinqing Zhang, Kuncheng Li, and Tianzi Jiang. Regional coherence changes in the early stages of alzheimers disease: a combined structural and resting-state functional mri study. *Neuroimage*, 35(2):488–500, 2007.

Saman Sarraf and Jian Sun. Functional brain imaging: A comprehensive survey. *arXiv preprint arXiv:1602.02225*, 2016.

Cheryl Grady, Saman Sarraf, Cristina Saverino, and Karen Campbell. Age differences in the functional interactions among the default, frontoparietal control, and dorsal attention networks. *Neurobiology of aging*, 41:159–172, 2016.

Cristina Saverino, Zainab Fatima, Saman Sarraf, Anita Oder, Stephen C Strother, and Cheryl L Grady. The associative memory deficit in aging is related to reduced selectivity of brain activity during encoding. *Journal of cognitive neuroscience*, 2016.

Mohammed A Warsi. The fractal nature and functional connectivity of brain function as measured by bold mri in alzheimers disease. 2012.

Cheryl L Grady, Anthony R McIntosh, Sania Beig, Michelle L Keightley, Hana Burian, and Sandra E Black. Evidence from functional neuroimaging of a compensatory prefrontal network in alzheimer's disease. *The Journal of Neuroscience*, 23(3):986–993, 2003.

Cheryl L Grady, Maura L Furey, Pietro Pietrini, Barry Horwitz, and Stanley I Rapoport. Altered brain functional connectivity and impaired short-term memory in alzheimer's disease. *Brain*, 124(4):739–756, 2001.

Evanthia E Tripoliti, Dimitrios I Fotiadis, and Maria Argyropoulou. A supervised method to assist the diagnosis and classification of the status of alzheimer's disease using data from an fmri experiment. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4419–4422. IEEE, 2008.

Allan Raventós and Moosa Zaidi. Automating neurological disease diagnosis using structural mr brain scan features.

Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, 11(Feb): 625–660, 2010.

Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Itamar Arel, Derek C Rose, and Thomas P Karnowski. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE Computational Intelligence Magazine*, 5(4):13–18, 2010.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Limin Wang, Zhe Wang, Wenbin Du, and Yu Qiao. Object-scene convolutional neural networks for event recognition in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 30–35, 2015.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for ad/mci classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 583–590. Springer, 2013.

Heung-Il Suk, Dinggang Shen, Alzheimers Disease Neuroimaging Initiative, et al. Deep learning in diagnosis of brain disorders. In *Recent Progress in Brain and Cognitive Engineering*, pages 203–213. Springer, 2015a.

Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimers Disease Neuroimaging Initiative, et al. Latent feature representation with stacked auto-encoder for ad/mci diagnosis. *Brain Structure and Function*, 220(2):841–859, 2015b.

Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.

Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.

Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J Fulham, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, 2015.

Eivind Arvesen. Automatic classification of alzheimers disease from structural mri. 2015.

Fayao Liu and Chunhua Shen. Learning deep convolutional features for mri based alzheimer's disease classification. *arXiv preprint arXiv:1404.3366*, 2014.

Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Michael Fulham, and Dagan Feng. High-level feature based pet image retrieval with deep learning architecture. *Journal of Nuclear Medicine*, 55 (supplement 1):2028–2028, 2014.

Tom Brosch, Roger Tam, Alzheimers Disease Neuroimaging Initiative, et al. Manifold learning of brain mris by deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 633–640. Springer, 2013.

Ladislav Rampasek and Anna Goldenberg. Tensorflow: Biologys gateway to deep learning? *Cell systems*, 2(1): 12–14, 2016.

Alexander de Brebisson and Giovanni Montana. Deep neural networks for anatomical brain segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–28, 2015.

Earnest Paul Ijjina and Chalavadi Krishna Mohan. Hybrid deep neural network model for human action recognition. *Applied Soft Computing*, 2015.

Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.

bioRxiv preprint doi: https://doi.org/10.1101/070441; this version posted August 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

Gwenaëlle Douaud, Stephen Smith, Mark Jenkinson, Timothy Behrens, Heidi Johansen-Berg, John Vickers, Susan James, Natalie Voets, Kate Watkins, Paul M Matthews, et al. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain*, 130(9):2375–2386, 2007.

Saman Sarraf and Ghassem Tofighi. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*, 2016.