

Received September 23, 2019, accepted October 17, 2019, date of publication October 25, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949577

MCADNet: Recognizing Stages of Cognitive Impairment Through Efficient Convolutional fMRI and MRI Neural Network Topology Models

SAMAN SARRAF¹, (Senior Member, IEEE), DANIELLE D. DESOUZA², JOHN A. E. ANDERSON^{3,4}, AND CRISTINA SAVERINO⁵

¹Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8P 1A2, Canada

²Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305, USA

³Kimel Family Translational Imaging Genetics Research Laboratory, The Centre for Addiction and Mental Health, University of Toronto, Toronto, ON M5S 1L8, Canada

⁴Campbell Family Mental Health Research Institute, The Centre for Addiction and Mental 5 Health, University of Toronto, Toronto, ON M5T 1L8, Canada

⁵Toronto Western Hospital, University Health Network, Toronto, ON M5T 2S8, Canada

Corresponding author: Saman Sarraf (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/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

ABSTRACT Mild cognitive impairment (MCI) represents the intermediate stage between normal cerebral aging and dementia associated with Alzheimer's disease (AD). Early diagnosis of MCI and AD through artificial intelligence has captured considerable scholarly interest; researchers hope to develop therapies capable of slowing or halting these processes. We developed a state-of-the-art deep learning algorithm based on an optimized convolutional neural network (CNN) topology called MCADNet that simultaneously recognizes MCI, AD, and normally aging brains in adults over the age of 75 years, using structural and functional magnetic resonance imaging (fMRI) data. Following highly detailed preprocessing, four-dimensional (4D) fMRI and 3D MRI were decomposed to create 2D images using a lossless transformation, which enables maximum preservation of data details. The samples were shuffled and subject-level training and testing datasets were completely independent. The optimized MCADNet was trained and extracted invariant and hierarchical features through convolutional layers followed by multi-classification in the last layer using a softmax layer. A decision-making algorithm was also designed to stabilize the outcome of the trained models. To measure the performance of classification, the accuracy rates for various pipelines were calculated before and after applying the decision-making algorithm. Accuracy rates of $99.77\% \pm 0.36\%$ and $97.5\% \pm 1.16\%$ were achieved for MRI and fMRI pipelines, respectively, after applying the decision-making algorithm. In conclusion, a cutting-edge and optimized topology called MCADNet was designed and preceded a preprocessing pipeline; this was followed by a decision-making step that yielded the highest performance achieved for simultaneous classification of the three cohorts examined.

INDEX TERMS Deep learning, classification, structural and functional magnetic resonance imaging, brain, Alzheimer's disease, MCI.

I. INTRODUCTION

A. COGNITIVE IMPAIRMENT

Cognitive impairment is a general term referring to impairments in cognition among the domains of memory, learning,

The associate editor coordinating the review of this manuscript and approving it for publication was Mohan Venkateshkumar¹.

concentration and decision-making. Cognitive impairment ranges from mild to severe and the symptoms can worsen over time and ultimately prevent a patient from performing daily tasks. Mild Cognitive impairment (MCI) was first utilized by Reisberg *et al.* [1] and is currently defined as a decline in cognitive ability that is detectable however lacking in terms of the severity to alter one's functioning of daily living.

The National Institute on Aging Alzheimer's Association (NIA-AA) has provided criteria to diagnose dementia and MCI when there occurs a significant cognitive deterioration from an individual's previous level [2], [3]. Additionally, research demonstrates that elderly adults with a diagnosis of MCI have a higher risk of developing dementia and age-related cognitive decline [4]. Petersen et al.'s research demonstrates that although there is still a scoring threshold in determining MCI, the memory decline of individuals with MCI is approximately 1.5 standard deviations below normative data of same age and educationally matched peers [5], [6]. Gallagher et al. indicate depression and anxiety have been reported in almost 50% of individuals with MCI, and a link between depression and anxiety with cognitive decline has been found [7]–[9]. Despite research demonstrating the elevated risk of dementia among those with a MCI diagnosis, it is unclear which factors and profiles of MCI are at greatest risk of progression and therefore most likely to benefit from early intervention. Researchers have investigated the effectiveness of MCI treatment with medication [10]. Morris et al. identified the major challenge in MCI research is distinguishing which memory deficits inevitably progress to Alzheimer's. An additional barrier to research within this population is that the diagnosis of MCI is established through various assessments such as Clinical Dementia Rating (CDR), Short Blessed Test (SBT) and Mini-Mental State Examination (MMSE) that are insensitive to early-stage AD. For example, researchers have shown that the MMSE scores are not good at predicting risk of future dementia [11], [12]. Grundman et al. explained the details of recruiting normal subjects in MCI studies. The normal subjects must be in the same age range and maintain a CDR of 0 and an MMSE score above 26. The subjects should also have a similar level of education [13]. Structural Magnetic Resonance Imaging (MRI) that captures the structure of the brain is the most popular imaging modality to recognize MCI [6], [14]–[17].

B. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

The human visual system consists of cells and synapses that capture visual information from the environment and transfer it to the human brain through a visual portal called the lateral geniculate nucleus (LGN) located in the thalamus. Interestingly, a set of pathways operate largely in parallel to transceiver visual information. Convolutional neural networks are inspired by the human visual system and perform hierarchical learning based on complicated algorithms that model low-high level features and extract abstractions from data. This architecture has been specifically designed based on the explicit assumption that raw datum are comprised of two-dimensional images that enable certain properties to be encoded while also reducing the amount of hyper parameters. 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. The network 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 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 [18]–[20]. 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 [19], [21]–[23]. Equation 1 demonstrates how the gradient component for a given weight is calculated in the backpropagation step, where E is the 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.

$$\begin{aligned} \frac{\partial E}{\partial \omega_{ab}} &= \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial x_{ij}^{\ell}} \frac{\partial x_{ij}^{\ell}}{\partial \omega_{ab}} \\ &= \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial x_{ij}^{\ell}} y^{l-1}_{(i+a)(j+b)} \end{aligned} \quad (1)$$

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 that backpropagation results in convolution while the weights are rotated. The rotation of the weights derives from a delta error in the convolutional neural network.

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

Several successfully developed deep CNN architectures have already been introduced for various computer vision tasks such as object recognition and classification, object classification and localization, object detection and image segmentation. LeNet-5 [18] is considered a fundamental architecture designed for handwritten and machine-printed character recognition. AlexNet [24] is also one of first CNN architectures designed for image classification. VGGNet [25] was developed at the University of Oxford for large-scale image recognition. GoogleNet [23] was introduced by the research team at Google by which the inception module was added to the network architecture. ResNet [26], one

of the monster architectures that defines a deep learning network and the residual module, was utilized for the first time. Next, ResNext [27] was designed and consisted of the concepts of inception and residual modules. This architecture has a vast application in image recognition. You Only Look Once or YOLO architecture [28] was designed to solve complicated image detection problems. The high computation costs in the CNN-based architectures encouraged researchers to develop optimized architectures to be functional on mobile devices and SqueezeNet [29] was introduced for low bandwidth scenarios. Image segmentation is one of the crucial tasks in computer vision and has been of interest to scientists in the field. SegNet [30] applied the deep learning concepts to solve image segmentation problems that included a set of encoders and decoders where the high frequency details are retained. Additionally, Generative Adversarial Networks (GANs) [31] were introduced to generate entirely new images not used in training datasets.

C. RELATED WORKS

A deep learning architecture including stacked auto-encoders and a softmax layer was designed by Siqi Liu to classify AD and MCI through a unique setting. The advantage of the design was to use less samples to train the model [32]. Suk *et al.* developed a deep learning-based extraction and classification method to classify AD/MCI where PET and MRI data were utilized. The accuracy rates of 95.9% and 85% were reported for AD and MCI, respectively [33]. Another work from Suk *et al.* was to classify AD, NC and MCI through a multimodal fusion system in which CNN models were utilized. The maximum accuracy rates of 93.52%, 85.19% for AD vs NC and MCI vs NC were obtained [34]. Changes in brain structure and function caused by Alzheimer's disease have proved of great interest to numerous scientists and research groups. In diagnostic imaging in particular, classification and predictive modeling of the stages of Alzheimer's have been broadly investigated. Suk *et al.* [33], [35], [36] developed a deep learning-based method to classify AD magnetic current imaging (MCI) and MCI-converter structural MRI and PET data. 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 [34]. Randomized denoising auto-encoder marker (rDAm) was used to design a multimodal imaging system against PET and structural MRI data to classify MCI and AD [37]. An automatic classification system was developed to recognize AD and MCI data who converted to AD using deep neural networks. The best accuracy rates achieved in this work were up to 86% for all AD and MCI samples vs healthy control, and MCI converters vs MCI stable with accuracy up to 75% [38]. Senanayake *et al.* implemented an approach for classification of MCI subtypes using deep

learning and stacked auto-encoder architectures. They classified 5 subtypes of MCI and the accuracy rates between 84% up to 97% were obtained [39]. Payan and Montana [40] of Imperial College London designed 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 with a reported accuracy rate nearly identical in terms of value. Additionally, a multimodal neuroimaging feature extraction pipeline for multiclass AD diagnosis was developed by Liu *et al.* [41]. 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.* [42], Liu and Shen [43], Liu *et al.* [44], Brosch *et al.* [45], Rampasek and Goldenberg [46], De Brebisson and Montana [47] and Ijjina and mohan [48]. Also, Qui *et al.* investigated the improvement in the accuracy of diagnosing MCI using MMSE scores and logical memory (LM) by adding MRI data and their fusion model could achieve up to 90% [49]. Another study showed a deep learning approach based on convolutional neural networks to accurately predict MCI-to-AD using structural MRI data with an accuracy of 79.9% and an area under the receiver operating characteristic curve (AUC) of 86.1% in leave-one-out cross validations [50]. Mazrina *et al.* [51], Wen *et al.* [52] and Srinivasan *et al.* [53] developed similar methodologies to predict MCI and AD brains. Nicola *et al.* developed a new method for early prediction of Alzheimer's disease that involves extracting random forest features from the data of an international challenge and classifying them via deep neural networks. In the classification, the authors considered four stages of the disease, including two stages of MCI. Their methodology produced higher accuracy rates, they found, than other machine learning strategies in that challenge [59]. Shi *et al.* employed a new strategy for classifying Alzheimer's data through the use of MRI and PET data. They introduced an algorithm called multimodal stacked DPN (MM-SDPN), which involves two steps: 1) fusing and 2) learning features from the brain imaging data. In their binary classification task, they showed the capability of using such a design to improve the performance of classification through multimodal feature learning [60]. A deep learning-based architecture derived from GoogLeNet's "InceptionV3" was employed to classify the F-FDG PET brain images of 40 patients. The CNN-based algorithm produced high specificity and sensitivity rates with a confidence level of 95%. However, the population employed in this study seemed insufficient for significant claims [61]. An automatic classification method using deep neural networks was designed to categorize AD and MCI big data. Basaia *et al.* demonstrated the capability

TABLE 1. The demographic information of data utilized in this study is shown in this table. The same age range of MCI, AD and NC subjects was considered to develop and validate the deep learning models in order to ignore aging effects.

Modality	Total Subj.	Group	Subj.	Female	Mean Age	SD	Male	Mean Age	SD	MMSE \pm SD
rs-fMRI	275	Alzheimer	52	21	79.42	16.35	31	80.54	15.98	22.70 \pm 2.10
		Control	92	43	80.79	19.16	49	81.75	21.43	28.82 \pm 1.35
		MCI	131	66	79.15	9.57	65	79.72	23.45	26.53 \pm 2.51
MRI	1076	Alzheimer	211	85	80.98	21.6	126	81.27	16.66	23.07 \pm 2.06
		Control	91	43	79.37	12.52	48	80.81	19.51	28.81 \pm 1.35
		MCI	774	265	80.28	10.94	509	81.61	17.25	26.53 \pm 2.09

of DNN by recruiting a huge dataset for training and testing the developed models. The highest accuracy rate of 98% was achieved in this work for AD and HC classification [62]. Jabason et al. developed a novel semi-supervised learning approach based on an auto encoder to classify ADNI data. They claimed that their algorithm improved the performance of classification for different evaluation metrics [63]. Using longitudinal multi-domain data from ADNI, Lee et al. developed a framework for predicting the conversion MCI stages to Alzheimer's disease. For the performance evaluation, they concentrated on the area under the AUC curve. Their algorithm uses a recurrent neural network that integrates longitudinal information and demographic information to train the model [64]. Spasov et al. employed structural MRI data and developed a deep learning-based framework that considered optimization of the number for the network parameters. They found that their framework produced very competitive results when they reduced the number of parameters in the training phase. Reducing the number of parameters results in faster training and the need for fewer images to develop the machine learning models [65]. Recently, various techniques including 3D CNN and parameter-efficient deep learning models were utilized for MRI data classification [65]–[67].

II. METHODS

A. PARTICIPANTS

Alzheimer's disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>) was considered in this work by which two categories of subjects for MRI and fMRI modalities were selected to develop the deep learning models. The first category included 275 subjects who were scanned for resting-state functional fMRI (rs-fMRI) studies. This dataset included 52 Alzheimer's patients, 92 healthy control subjects and 131 MCI patients (age group > 75). In the second category, 1076 subjects were selected to participate in MRI data acquisition. This group included 211 Alzheimer's patients, 91 healthy control subjects and 774 MCI patients. In both categories, certain subjects were scanned at substantially different points in time, and their imaging data were separately considered in this work. Table 1 presents the demographic information for both categories, which also include mini mental state examination (MMSE) scores.

B. IMAGE ACQUISITION

MRI data acquisition was performed according to the ADNI acquisition protocol [23]. Scanning was performed on three different Tesla scanners, General Electric (GE) Healthcare, Philips Medical Systems, and Siemens Medical Solutions, and was based on the same scanning parameters. Anatomical scans were acquired with a 3D MPRAGE sequence (TR=2s, TE=2.63 ms, FOV=25.6 cm, 256 \times 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 \times 64 matrix, 30 axial slices of 5mm thickness without gap).

C. 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 Rorden et al. <http://www.sph.sc.edu/comd/rorden/micron/dcm2nii.html>. Next, non-brain regions, including skull and neck voxels, were removed from the structural T1-weighted image corresponding to each fMRI time course using FSL-BET [54]. Resting-state fMRI data, including 140 time series per subject, were corrected for motion artefact using FSL-MCFLIRT [55], as low frequency drifts and motion could adversely affect decomposition. The next necessary step was the regular slice timing correction, applied to each voxel's time series because of the assumption that later processing assumes all slices were acquired exactly half-way through the relevant volume's acquisition time (TR). In fact, each slice was 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 individual's 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 $45 \times 54 \times 45$ images per time course.

D. 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 [54] 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> [45]. 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, averaged and flipped along the x-axis, the two mirror images then re-averaged to obtain a first-pass, study-specific affine GM template. Second, the GM images were re-registered to this affine GM template using non-linear registration, concatenated into a 4D image, 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 $2 \times 2 \times 2$ 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, by which the 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 resulted in full width at half maximum (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 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. The copyright holder for this preprint is the author/funder. It is made available under a CC-BY-NC 4.0 International license.

E. DATA CONVERSION

Various data conversion and augmentation methods are available in the literature. However, it seems the algorithm developed by Sarraf *et al.* [56], [57] produces the highest classification performance in which MRI and fMRI data are decomposed along Z direction and converted from 3D and 4D data into 2D imaging samples. The content of imaging data

must be preserved during data conversion, therefore lossless data conversion was utilized. In a lossless data conversion, all original data are recovered and every single bit of data remains after conversion and the information is fully restored. In this work, Portable Network Graphics (PNG) lossless data conversion was used. 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 lossless PNG format using the Python OpenCV (opencv.org). The last 10 slices of each time course were removed since they included no functional information. Also, the sum of pixel intensities of each slice was calculated and any slices with zero sum of pixel intensities equal to zero were ignored to augment the data. Equation 3 describes the conversion of a given slice to a PNG sample which applies to every subject’s time course.

$$\begin{aligned}
 & \text{for } \forall z = 1 \text{ to } Z - 10 \\
 & \text{for } \forall t = 1 \text{ to } T \\
 & \text{if } I_{z,t}(S_{z,t}(x,y)) = \sum_{x=1}^X \sum_{y=1}^Y S(X,Y) \neq 0 : \\
 & \quad S_{z,t}(x,y) \longrightarrow \text{PNG}(S_{z,t}(x,y)) \\
 & \text{otherwise :} \\
 & \quad \text{Ignore } S_{z,t}(x,y) \tag{3}
 \end{aligned}$$

where x, y and z are spatial dimensions (from 1 to X, Y, Z , respectively), t is a time point of a given fMRI time course with T points, $S_{z,t}(x,y)$ is a given slice with a dimension of (x,y) for the position of (z,t) and $I_{z,t}$ represent the intensity function of $S_{z,t}(x,y)$. PNG represents the lossless PNG transformation function. The preprocessed MRI data were also loaded into memory using a similar approach to the fMRI pipeline and were converted from Nifti to lossless PNG format using Nibabel and OpenCV, which created three groups (MCI, AD and NC) of four preprocessed datasets (MRI samples with sigma = 0,2,3,4). Additionally, to augment the data, the slices with zero mean pixels were removed from the datasets. The conversion criteria are similar to Equation 4 but without removing any slice from the end of 3D image and represents the subject number in the stack of structural MRI data.

F. MCADNNET TOPOLOGY

To recognize MCI, Alzheimer’s disease and Normal control brains through a unique network, an efficient CNN-based topology called MCADNNet was designed and trained from scratch. As discussed previously, various CNN-based architectures including LeNet-5 [18], DeepAD [57] with two and four layers [56]–[58], GoogleNet [23] and ResNet [26] were utilized to classify the dementia data. Although deep learning pipeline design requires massive testing, grid search as well as applying various techniques for hyper parameters optimization, a simultaneous understanding of the machine learning models and data usually leads to an efficient

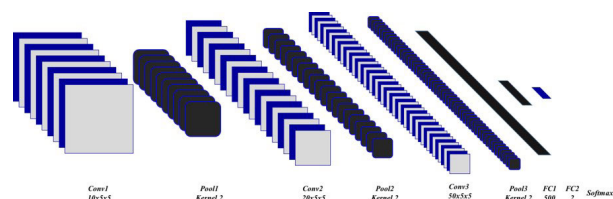


FIGURE 1. MCADNNet is a unique and optimized topology that simultaneously recognizes MCI, AD, NC participants in MRI and fMRI data. The architecture includes three layers of convolution (Grey-Blue layers) including 10, 20 and 50 filters of 5×5 , as well as three Max Pooling layers (Black-Blue layers). MCADNNet ends with two fully connected layers of 500 and two hidden neurons (Black layers) followed by a softmax function (Blue layer). In both cases of MRI and fMRI, the first layer receives 56×56 images, the closest dimension to preprocessed images enabling the architecture to extract more details from the images. Also, compared to the DeepAD model, the 3-layer MCADNNet extracts more hierarchical features that result in high performance of multi-class recognition.

topology. The experiments demonstrated that more complicated networks including several convolutional layers do not necessarily produce higher accuracy rates and that trade-off between network complexity and performance of classification must be achieved through a valid hypothesis, for example, the input dimensions. Three layers of convolution with three pooling layers followed by two fully-connected layers were utilized in MCADNNet topology (<https://github.com/samansarraf/MCADNNet>). Finally, a softmax layer to classify three classes was added to the end of the network. Three convolutional layers were designed to extract deep but hierarchical features from data. Figure 1 images the MCADNNet architecture. In this topology design, functional and structural MRI samples were upsampled to 56×56 pixels, the closest dimension to the original image size after data conversion. The upsampled images were fed into the first convolution layer that contains 10 filters of 5×5 . In the second layer, the first max pooling layer downsampled the data by a factor of two. Next, in the third layer that is the second convolution layer, the features were passed through 20 filters of 5×5 . As we will see later, the first Conv. layer extracted high-level features. After, the second max pooling layer downsampled the outcome of the second conv. layer which were mid-level features by a factor of two. The final convolution layer (the 5th layer) generated the low-level features to feed the last pooling layer. Two consecutive fully-connected layers were learned from the hierarchical features and transferred the output to the softmax layer for multiclass classification. Increasing the number of convolutional layers as well as the number of filters generated a higher number of network parameters. To avoid any potential overfitting or extraction of various features from the data, the pooling layers were utilized, which also accelerated the training process. As described earlier, DeepAD and MCADNNet were trained from scratch using ADNI data, so we considered no fine-tuning of network parameters. Fine-tuning of parameters occurs when a pre-trained network is employed; however, both of our architectures were freshly trained.

III. RESULTS

A. RESTING-STATE FMRI PIPELINE

The 4D preprocessed fMRI time series were randomly shuffled in subject-level and five training datasets including 75% of subjects for three classes (MCI, AD and NC) and five validation datasets including 25% of subjects were generated. Based on subject-level data selection, a given training and validation dataset has no samples from the same subject in common. This approach enables aggressive testing and validation of the trained CNN models by examining the robustness of the models against unseen data. Next, the 4D time series were passed through the data conversion module, producing a total of 1433880 2D PNG samples, including 640640 MCI, 270900 AD and 522340 NC images. DeepAD and MCADNNet were trained and validated for various classifications as shown in Table 3. The DeepAD architecture input layer received the resized samples to 28×28 , while MCADNNet was fed by the 56×56 images. The PNG samples were then converted to the Lightning Memory-Mapped Database (LMDB) for high throughput for the Caffe Deep Learning platform [19] used for this classification experiment. Both CNNs modes were adjusted for 30 epochs and initialized for Stochastic Gradient Descent with $\gamma = 0.1$, momentum = 0.9, learning rate = 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 validation of Caffe models were performed and 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. During the training and testing processes, the loss of training, testing and accuracy of testing data were monitored. To confirm the reproducibility of the results, the entire process described above was repeated five times on the same server using NVIDIA DIGITS Caffe (the Deep Learning GPU Training System) and the identical results were replicated. The accuracy rates, loss values of testing datasets and loss values for training datasets were monitored during the training process. Figure 2 demonstrates the performance of training and testing during 30 epochs in DeepAD and MCADNNet architectures using fMRI data for coincident classification of MCI/AD/NC. In the first epoch, the loss values were slightly above the unit value (one) derived from a random initialization. The convergence rapidly occurred in the first epochs. Although DeepAD model against fMRI data converged in the first iterations, the accuracy rate for testing dataset was lower than the 3-layer MCADNNet model.

B. STRUCTURAL MRI PIPELINE

Using a similar methodology described above, the 3D MRI subjects were five randomly-shuffled training and validation datasets by which we divided the data into 75% and 25%. The data conversion module produced a total of 110002

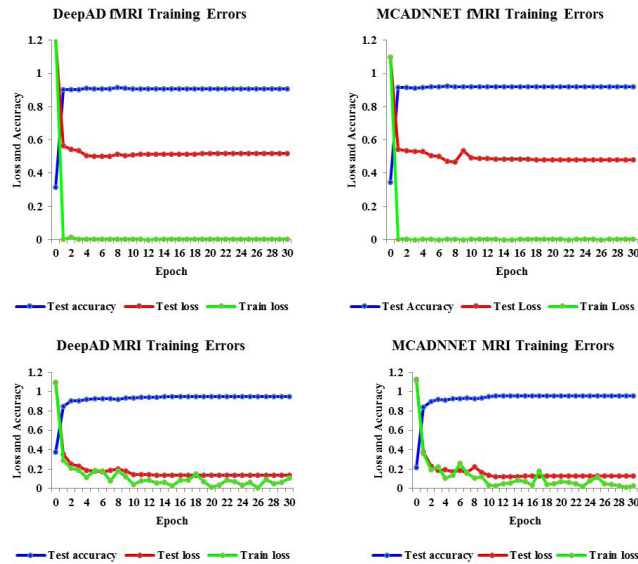


FIGURE 2. Top: Training process in one of the fMRI experiments for DeepAD (left) and MCADNNet (right) architectures indicate the models that produce a very high performance of classification. The random initiation at the zeroth epoch was above one, which dropped dramatically after the first iterations (can occur due to Caffe DIGITS implementation). The model converged in the first iterations due to utilizing high-volume, aggressively-preprocessed fMRI data. The final accuracy rate achieved from this MCADNNet model (utilized in this visualization) was 92.35% in the subject-level experience before decision making, while the accuracy rate was 90.76% from the DeepAD model trained and tested by the same dataset. **Bottom:** Training processes for MCADNNet against structural MRI data (right) and DeepAD (left) architecture were figured. The random initialization of loss values began above one and dramatically dropped after the first epoch. Additionally, the slight fluctuation in the loss of training datasets explains the impact of SGD as a randomly-selected block of data was injected into the network in order to train the model. The highest accuracy rate achieved was 95.44% from MCADNNet and 94.80% from DeepAD against a given MRI testing dataset.

2D samples including 58067 MCI, 43743 AD and 8192 images. In DeepAD, the effect of imbalanced data proved no impact on the performance of classification in this case. To train and validate, both DeepAD and MCADNNet 82419 and 27583 samples were utilized, respectively. As mentioned in the MRI preprocessing section, to explore the effect of spatial smoothing on the model development, four sets of samples per datasets were generated and a total of 20 training and validation datasets were utilized. Additionally, the slices with zero mean pixels were removed from the data, which was then converted to the LMDB format and resized to 28×28 pixels for DeepAD and 56×56 pixels for MCADNNet. The DeepAD model was set for 30 epochs and initiated for Stochastic Gradient Descent with $\gamma = 0.1$, momentum = 0.9, base learning rate = 0.01, weight-decay = 0.0005, and a step learning rate policy dropping the learning rate in steps by a factor of gamma every stepsize iteration. 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 results are shown in Table 3 ('Before Decision Making' section) for various models and spatially smoothed samples. The training behaviors of two models against structural MRI data were shown in Figure 2. The impact of utilizing

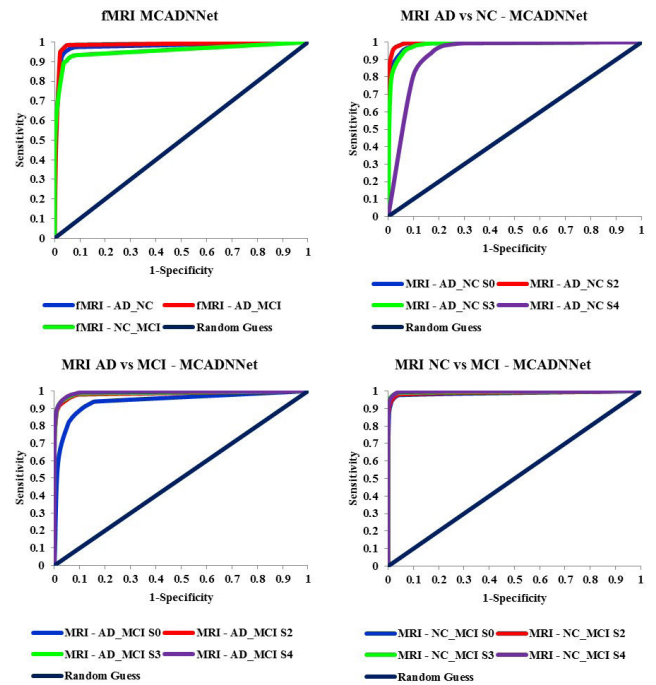


FIGURE 3. ROC Curves show the performance of classification for the trained models. In this study, the ROC and AUC indicate classification was accurately performed and each curve representing an experiment was very far from the random guess in the binary classification tasks where the structural and functional MRI data were used for training both DeepAD and MCADNNet architectures. The figure top-left shows the performance of MCADNNet for binary classification including fMRI testing data (AD/NC, AD/MCI, NC/MCI). As mentioned in the discussion of the structural MRI preprocessing pipeline, four Gaussian kernels including $\sigma = 0, 2, 3,$ and 4 mm were utilized. The performance of binary classification for MRI testing data is shown in Figure 3 (top-right, bottom-left, and bottom-right) for AD/NC, AD/MCI and NC/MCI (and different Gaussian kernels used for smoothing). The ROC curves show that all the binary classifications provide high performance as they are extremely close to the upper left corner of the plot.

stochastic gradient descent (SGC) in the training process was remarkable in loss graph of the training and testing datasets. As MCADNNet included a higher number of parameters representing a more complex architecture converged in the later iterations compared to DeepAD. Higher volume of fMRI and higher pattern complexity in structural MRI data result in a better performance of classification for fMRI data.

C. PERFORMANCE OF CLASSIFICATION

The performance of MCADNNet trained models was qualitatively evaluated by calculating the area under curve (AUC) for receiver operating characteristic (ROC) curves and the accuracy rate per class to generate confusion matrices (CM). For the sake of performance analysis, two approaches were considered. First, the ROC curves shown in Figure 3 were extracted for binary classification tasks such as AD vs MCI or NC vs MCI where sensitivity and specificity of each experience were calculated by obtaining the number of true and false positive and negative results. In the second approach, the confusion matrices were calculated for the 3-class MCADNNet models of both functional and Structural

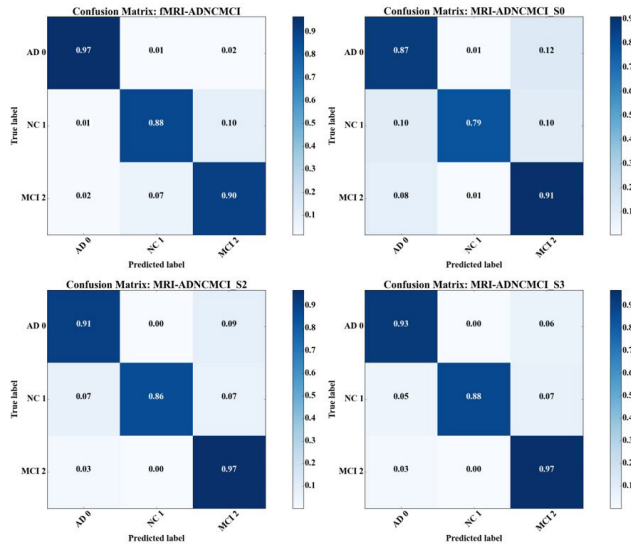


FIGURE 4. Confusion matrices for simultaneously classifying MCI, AD and NC classes through MCADNNet architecture were extracted indicating the quality of prediction for the three classes. As seen, in most cases including both fMRI and MRI, the performance of classification (prediction) was very high for classes specifying that the trained models were unbiased to any of three classes. However, in the fMRI experiment, the highest score belonged to MCI class and the lowest prediction scores were obtained from the normal brains, confirming that MCI is the early stage of the dementia. In structural MRI, the highest accuracy rate belonged to the MCI group and interestingly the highest error rate also occurred in recognizing AD from MCI, revealing the fact of structural similarity between certain brain regions in AD and MCI groups. However, in the MRI classification experiment, the number of samples utilized in training and validation was significantly smaller than fMRI methods (3D vs 4D data) that also played an important role in the model convergence in the early epochs as well as in the performance of classification. This figure shows the normalized confusion matrices for fMRI and MRI experiments for AD vs NC vs MCI multi-class classification. The top-left figure shows the performance of classification for fMRI testing data. All accuracy rates for three classes are located in the diagonal of the confusion matrix; the rates are 97%, 88%, and 90%, respectively, for AD, NC, and MCI.

MRI data displayed in Figure 4 where the items on the diagonal were cases that the models' prediction were correct. As shown in the figures, the performance of binary classification was close to a perfect ROC curve in the most trained CNN models. The curves validated that the classification was not impacted by the number of samples in each class and training process was successfully completed. Additionally, the confusion matrices demonstrated that the multi-class MCADNNet trained models could recognize the AD samples slightly better than two other classes, although the AD class had a smaller contribution in training process in terms of number of samples. MCI and NC samples showed higher similarity, therefore the prediction rates closely competed with each other, proving this clinical fact that MCI is an early stage of the disease and the brain has a similar function and structure of a normal aging brain. In the structural MRI pipeline, the spatial smoothing affected the output where the data $\sigma=0$ mm (without smoothing) provided the lower accuracy rates while the samples smoothed by $\sigma=2$ or 3 mm demonstrated a higher performance of classification. The accuracy rate before decision making was measured

by dividing the amount of correctly predicted "slices" by the number of all slices in a given experiment according to the standard definition of accuracy. This was called "slice-level" prediction, as described in the DeepAD paper [57]. One of the particular strengths of CNN architectures is its extraction of hierarchical features through several layers containing filters. Research showed paradoxical results by visualizing the weights of filters that sometimes represent a pattern or random shapes. Therefore, the interpretation of kernel weights is still challenging. However, the researchers showed that the visualization of the features extracted by a given filter is often meaningful and helps to better understand what features levels are represented by a given CNN layer and its kernels. Also, some research works have recently shown the potential benefit of using the hierarchical features among or instead of preprocessed data in the brain studies. Figure 6 demonstrates the hierarchical features extracted from three different CNN layers in MCADNNet for randomly selected one structural MRI subject per three groups.

D. FURTHER PERFORMANCE EVALUATION

To further evaluate the performance of classification for the MCADNNet model, we employed three other metrics (precision, recall, and F1-score) through three methods of calculation called micro, macro, and weighted average. Therefore, we measured nine metrics for each experiment using structural and functional MRI. A macro-metric will compute the metric independently for each class and then take the average, thereby treating all classes equally. A micro-metric will aggregate the contributions of all classes to compute the average metric, a weighted-metric for each class, and find their average weighted by support of the number of true instances for each class. This approach alters the macro to account for class imbalance. In a multi-class classification setup, the micro-average is preferable if you suspect a class imbalance. As known, precision represents the total true positive over summation of the true positive and false positive, whereas recall is calculated by the total true positive over summation of the true positive and false negative. Additionally, the F1-score is a function of precision and recall and is calculated as in Equation 4:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The F1-score produces balance precision and recalls where a false positive and negative might have significant cost, which is not considered in an accuracy rate. Using those nine additional metrics allowed us to fully evaluate the performance of classification independently from the accuracy rates. Tables 5 and 6 demonstrate the results from MCADNNet and DeepAD using functional MRI data, respectively. Also, Tables 7 and 8 show the results of the two architectures using structural MRI data for various σ values and classifications. The analysis using the nine metrics showed a very high correlation with the accuracy metrics and therefore validates our finding and discussion in the previous section.

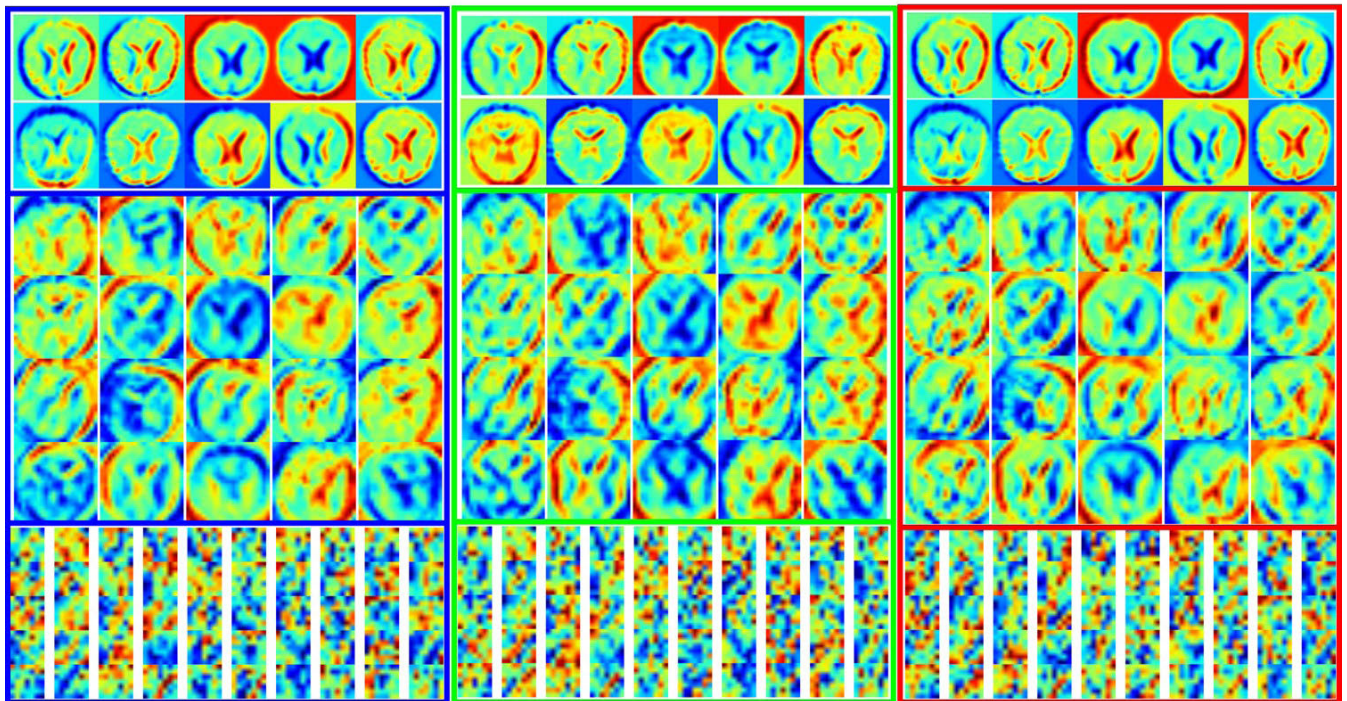


FIGURE 5. Structural MRI AD, MCI, and NC Features are visualized through MCADNNet Conv. layer 1, 2 and 3. The hierarchical features extracted by convolutional layers have opened new avenues to investigate the brain structure and function. We aimed to classify three classes of brains, so called NC, which is the blue framed group at left. The MCI is green framed group in the middle and Alzheimer's brains are the red framed group at right. When the performance of classification in CNN-based experiment is fairly high, the features extracted by various CNN layers can be utilized for other analyses. One randomly selected sample from each of three classes was used and passed through the prediction module of the final trained version of MCADNNet. The features extracted from each CNN layer were visualized and shown above.

E. DECISION MAKING: VOTE FOR MAJORITY

A decision-maker algorithm is a means for discovering the best choice among a list of alternatives based on the preferred single criterion or multi-criteria and values. In a majority rule-based system, a multi-class decision is made by voting for the class or group that has the highest number of candidates. The deep learning-based pipeline implemented in this study uses 2D images from subjects in all three classes that are separated in the subject-level for training and testing. The performance of classification is measured by counting the number of slices correctly recognized. In order to recognize a given scan that includes the slices of a given subject whether it belongs to MCI, AD or NC group, the decision making algorithm is required. In the given scan, the number of slices for each class was calculated, then the number of slices for each class within a subject was compared, and the class with most slices was presented as the candidate. Furthermore, the decision-maker system stabilized both fMRI and MRI pipeline by significantly improving the accuracy rates. The algorithm description is as follows:

The decision-maker algorithm as a post-classification method was applied to all DeepAD and MCADNNet models for both structural MRI and rs-fMRI. The final results shown in Table 2 and Table 3 indicate a significant improvement in the accuracy rate of subject-level recognition. After applying the rapid with low complexity decision making algorithm to

the output of classification, most of the subject-level accuracies reached a rate of 100%, demonstrating a superior confidence level of the pipeline. MCADNNet topology is a CNN-based mode which offers an optimal solution to recognize three major stages of Alzheimer's disease. The performance of this optimal topology has been obtained because of massive and precise preprocessing steps, and an optimal CNN-based model followed a decision-making algorithm which improved and stabilized the outcome of the network. Unlike other works that emphasize the classification part of the entire pipeline, MCADNNet aggressively preprocess the data and decomposes the data into 2D samples to develop a three-layer CNN model. The advantages of using MCADNNet over other architectures are aggressive preprocessing, an optimal deep learning model, and decision making in which a trade-off between network complexity and performance of classification exists.

IV. CONCLUSION

The number samples extracted from fMRI data through the converting algorithm sufficed to train both MCADNNet and DeepAD in the early stages of the training process. In addition, aggressively preprocessing the fMRI data removed noise, allowing the samples to be distinguished more readily, which enabled the models to converge in the very first epochs. The slight improvement in the accuracy rate using

TABLE 2. The decision-making algorithm as a post-classification step was applied to the results from both DeepAD and MCADNNet in order to stabilize the classification outputs and provide the new accuracy rates for all individuals. In this experiment, four classifications tasks through four sigma values were completed. As shown below, the decision-making algorithm improved the performance of classification, which resulted in an accurate recognition in most experiments. Conditionally formatted tables indicate how the accuracy rates were improved (from blue range to white range). For instance, the accuracy rate of 96.6% before decision making was improved to 100%, which shows all of the subjects in the classification were correctly recognized.

DeepAD - Structural MRI											
Subject Level		Before Decision Making					After Decision Making				
Experiment	σ	1	2	3	4	5	1	2	3	4	5
AD_NC	S0	0.9661	0.9583	0.9653	0.949	0.9593	1	0.9709	0.9942	0.9651	0.9826
	S2	0.9797	0.9787	0.9728	0.9745	0.9771	1	1	1	1	1
	S3	0.9836	0.9723	0.9776	0.9842	0.9763	1	1	1	1	1
	S4	0.9802	0.9761	0.9804	0.9821	0.9753	1	1	1	1	1
AD_NC_MCI	S0	0.87227	0.88459	0.87468	0.8796	0.87381	0.9918	0.99454	0.99454	0.98907	0.99454
	S2	0.93203	0.93378	0.93432	0.9379	0.9354	0.99727	1	1	1	1
	S3	0.94805	0.94366	0.94299	0.94067	0.94466	1	0.99727	1	1	1
	S4	0.94769	0.94408	0.94297	0.94516	0.94454	1	0.99727	1	1	1
AD_MCI	S0	0.89394	0.8867	0.8804	0.89162	0.88791	0.98516	0.98521	0.97879	0.99405	0.97923
	S2	0.93953	0.93934	0.94094	0.94017	0.94801	0.99703	1	0.99703	1	1
	S3	0.94776	0.95005	0.95055	0.95347	0.95053	1	0.99407	0.99408	1	0.99703
	S4	0.95081	0.95356	0.95363	0.95066	0.95281	0.99701	1	1	0.99703	1
NC_MCI	S0	0.96797	0.96995	0.97288	0.97925	0.97019	0.99078	0.98643	0.99091	1	0.99548
	S2	0.98619	0.98386	0.9845	0.98474	0.98419	1	1	1	1	1
	S3	0.98432	0.98867	0.98426	0.98538	0.98443	1	1	1	1	1
	S4	0.98478	0.98583	0.98765	0.98597	0.98568	1	1	1	1	1

MCADNNet - Structural MRI											
Subject Level		Before Decision Making					After Decision Making				
Experiment	σ	1	2	3	4	5	1	2	3	4	5
AD_NC	S0	0.96989	0.96118	0.96822	0.95014	0.96588	1	0.97674	0.99419	0.96512	0.98256
	S2	0.98261	0.98361	0.97769	0.97909	0.98154	1	1	1	0.99419	1
	S3	0.98133	0.97776	0.97824	0.98547	0.98374	1	1	1	1	1
	S4	0.98394	0.98331	0.9864	0.98672	0.98034	1	1	1	1	1
AD_NC_MCI	S0	0.87658	0.88905	0.87266	0.87773	0.87594	0.98907	0.99454	0.9918	0.9918	0.99454
	S2	0.9416	0.9396	0.94379	0.94069	0.94142	0.99727	1	1	1	1
	S3	0.95371	0.94812	0.95168	0.94295	0.94994	1	0.99727	1	1	1
	S4	0.95147	0.95363	0.94997	0.94948	0.95032	1	1	1	1	0.99727
AD_MCI	S0	0.89596	0.89237	0.88448	0.8973	0.89578	0.97626	0.98225	0.98182	0.99107	0.97923
	S2	0.94622	0.94608	0.94888	0.94432	0.9533	1	1	1	1	1
	S3	0.95461	0.95532	0.95463	0.95867	0.9561	1	0.99407	0.99408	1	0.99703
	S4	0.9563	0.95969	0.95909	0.95329	0.96025	0.99701	1	1	0.99703	1
NC_MCI	S0	0.97182	0.97175	0.97468	0.98201	0.97276	0.99078	0.98643	0.99091	1	0.99548
	S2	0.98812	0.98627	0.98625	0.98504	0.98389	1	1	1	1	1
	S3	0.98509	0.99051	0.98728	0.98728	0.98532	1	1	1	1	1
	S4	0.98895	0.98856	0.98852	0.98626	0.98725	1	1	1	1	1

TABLE 3. Functional MRI data used to train and validate both DeepAD and MCADNNet architectures. Improvement in the performance of classification is discovered by applying the decision-making algorithm in the post classification step. This improvement has been displayed from blue range (lower values) to white range (higher values). As mentioned above, the voting method enabled the pipeline to produce highly robust and reproducible outcomes.

DeepAD - Functional MRI											
Subject Level		Before Decision Making					After Decision Making				
Experiment		1	2	3	4	5	1	2	3	4	5
AD_NC		0.97	0.9327	0.9468	0.9032	0.966	1	1	0.9706	0.9211	1
AD_NC_MCI		0.91	0.854	0.9068	0.8968	0.8893	0.9444	0.9444	0.9861	0.9583	0.9861
AD_MCI		0.92	0.9345	0.9683	0.9327	0.9336	0.9574	0.9787	1	0.9787	1
NC_MCI		0.92	0.9319	0.9199	0.9054	0.9223	0.9828	0.9655	0.9828	0.9828	0.9828

MCADNNet - Functional MRI											
Subject Level		Before Decision Making					After Decision Making				
Experiment		1	2	3	4	5	1	2	3	4	5
AD_NC		0.95	0.9797	0.9711	0.8985	0.9663	0.9744	0.9744	1	0.9231	1
AD_NC_MCI		0.92	0.8614	0.8967	0.9097	0.8993	0.9861	0.9722	0.9722	0.9583	0.9861
AD_MCI		0.92	0.942	0.9666	0.9387	0.9484	0.9787	0.9787	1	0.9574	1
NC_MCI		0.92	0.9326	0.9196	0.9109	0.9294	0.9828	0.9828	0.9655	0.9828	0.9655

MCADNNet compared to DeepAD revealed that high-level features were extracted from the highly correlated fMRI samples. However, the number of structural MRI samples

used for training the models was significantly less than fMRI experiments due to 3D vs 4D data decomposition into 2D images, which explains the early convergence of the

TABLE 4. The comparison below shows the accuracy rates of testing datasets obtained from various machine learning algorithms to distinguish between three major categories of adult brains.

Reference	Modality	Method	AD/MCI/NC	AD+MCI/NC	AD/NC	AD/MCI	NC/MCI
Liu et al. [32]	MRI,PET	AE+SVM	-	-	87.76	-	76.92
Suk et al. [33]	MRI,PET	LLF+SAEF+SVM	-	-	95.9	-	85
Suk et al. [34]	MRI,PET	Patch+DBM	-	-	95.35	-	85.67
Suk et al. [35]	MRI,PET	SAE+SVM	-	-	85.7	64.5	70.6
Basaia et al. [38]	MRI	DNN	-	86	-	-	98
Senanayake et al. [39]	MRI	SAE	-	-	-	-	88.72
Payan et al. [40]	MRI	CNN	85.53	-	95.39	82.24	90.13
Liu et al. [43]	MRI	CNN	88.37	-	95.01	91.82	88.73
Qiu et al. [49]	MRI	MMSE+CNN	-	-	-	-	90.9
Lin et al. [50]	MRI	CNN	-	-	88.79	-	-
Srinivasan et al. [53]	MRI	SYMLET+SVM	-	-	89.7	-	-
Hosseini et al. [58]	MRI	CNN	89.1	90.3	97.6	95.18	90.81
Sarraf et al. [56]	fMRI	CNN	-	-	96.8	-	-
MCADNet	fMRI	CNN	97.43	-	97.5	98.3	97.59
MCADNet	MRI	CNN	100	-	99.9	99.7	100

TABLE 5. The performance of MCADNet for fMRI data is analyzed using three major metrics: Precision, Recall, and F1-Score. We calculated those metrics using three approaches called micro, macro, and weighted average, explained earlier in the manuscript. The main idea behind employing those techniques is to ensure that the machine learning models developed work properly in the case of imbalanced data and the accuracy rates obtained are valid for further analyses. As shown in this table, all the scores from the nine measurements demonstrate high performance of classification for MCADNet and are highly correlated to the accuracy rates in Table 3.

Subject Level		MCADNet - Functional MRI								
Experiment	Repetition	Precision			Recall			F1-Score		
		Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.
AD_NC	1	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	2	0.94	0.93	0.94	0.94	0.93	0.94	0.94	0.93	0.94
	3	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	4	0.9	0.9	0.9	0.9	0.87	0.9	0.9	0.88	0.9
	5	0.97	0.96	0.97	0.97	0.96	0.97	0.97	0.96	0.97
AD_NC_MCI	1	0.92	0.92	0.92	0.92	0.9	0.92	0.92	0.91	0.92
	2	0.86	0.86	0.86	0.86	0.85	0.86	0.86	0.85	0.86
	3	0.9	0.9	0.9	0.9	0.91	0.9	0.9	0.9	0.9
	4	0.91	0.92	0.91	0.91	0.89	0.91	0.91	0.9	0.91
	5	0.9	0.91	0.9	0.9	0.88	0.9	0.9	0.89	0.9
AD_MCI	1	0.92	0.91	0.92	0.92	0.91	0.92	0.92	0.91	0.92
	2	0.94	0.94	0.94	0.94	0.93	0.94	0.94	0.93	0.94
	3	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.97
	4	0.94	0.93	0.94	0.94	0.92	0.94	0.94	0.93	0.94
	5	0.95	0.93	0.95	0.95	0.95	0.95	0.95	0.94	0.95
NC_MCI	1	0.92	0.92	0.92	0.92	0.91	0.92	0.92	0.92	0.92
	2	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
	3	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	4	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
	5	0.93	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.93

TABLE 6. The same concept in the previous analysis -Table 5- is applied to DeepAD architecture using fMRI data. The qualitative analysis reveals that DeepAD performance is very high; however, the new MCADNet architecture produces better performance.

Subject Level		DeepAD - Functional MRI								
Experiment	Repetition	Precision			Recall			F1-Score		
		Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.
AD_NC	1	0.95	0.93	0.95	0.95	0.95	0.95	0.95	0.94	0.95
	2	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	3	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.97
	4	0.91	0.92	0.91	0.91	0.88	0.91	0.91	0.9	0.91
	5	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
AD_NC_MCI	1	0.91	0.91	0.91	0.91	0.89	0.91	0.91	0.9	0.91
	2	0.85	0.85	0.86	0.85	0.83	0.85	0.85	0.84	0.85
	3	0.91	0.91	0.91	0.91	0.92	0.91	0.91	0.91	0.91
	4	0.9	0.9	0.9	0.9	0.87	0.9	0.9	0.88	0.9
	5	0.89	0.89	0.89	0.89	0.87	0.89	0.89	0.88	0.89
AD_MCI	1	0.92	0.91	0.92	0.92	0.91	0.92	0.92	0.91	0.92
	2	0.93	0.93	0.93	0.93	0.92	0.93	0.93	0.92	0.93
	3	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.97
	4	0.93	0.92	0.93	0.93	0.92	0.93	0.93	0.92	0.93
	5	0.93	0.91	0.94	0.93	0.94	0.93	0.93	0.92	0.93
NC_MCI	1	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	2	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
	3	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	4	0.91	0.91	0.91	0.91	0.9	0.91	0.91	0.9	0.91
	5	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92

models in fMRI data and the convergence of the MRI models in later epochs. Decomposing data to 2D images and thus adding an extra step to the pipeline provided more samples

during the training processes. Furthermore, MCADNet showed an improvement in the accuracy rates in recognizing three classes compared to DeepAD because a sufficiently

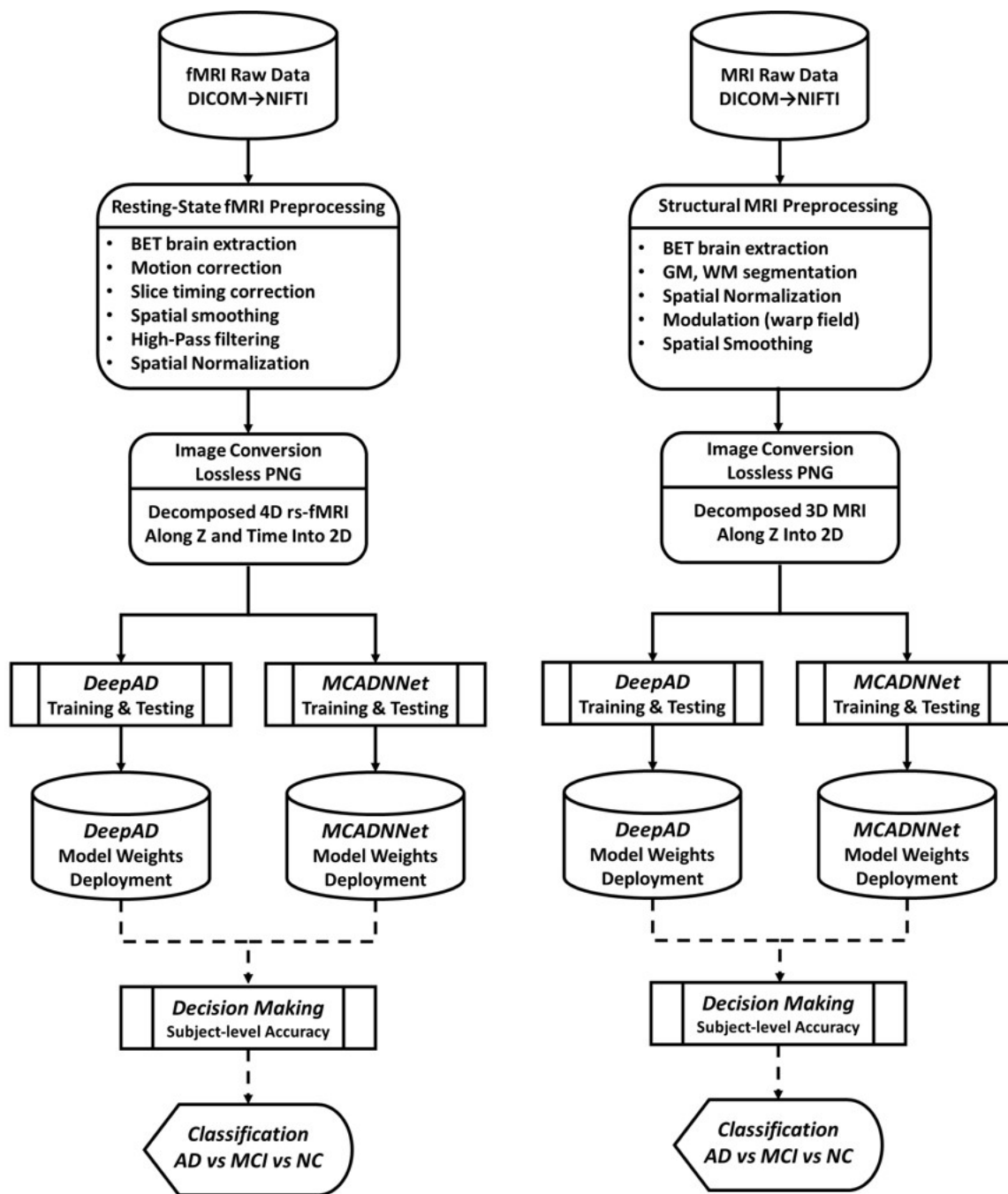


FIGURE 6. fMRI and MRI Pipelines. In this work, we designed a new CNN-based topology to predict NC, MCI and AD using functional and structural MRI data. The new topology contains three layers of CNNs in which the efficient parameters were utilized. Furthermore, a decision-making algorithm was developed to stabilize the results from the deep learning engine.

deeper set of features was extracted. The decision-making algorithm provided subject-level accuracy rates for stabilizing the output of the classification. For future work, a simultaneous classification of MCI subcategories and a potential pipeline with lower sensitivity to preprocessing steps should be considered. Also, a less dependent framework of preprocessing steps for classifying Alzheimer’s stages could be designed as a future project in which the pipeline would ideally receive raw data from users and perform classification.

ACKNOWLEDGMENT

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. Ho mann-La Roche Ltd and it’s a

TABLE 7. More parameters were explored to develop machine learning models using structural MRI data. Sigma representing the Gaussian kernels was considered, including four values. The idea behind employing nine metrics to evaluate the performance of classification of the models, as explained above, is applied to MCADNet models using structural MRI data. The qualitative analyses show that the measures are highly correlated with accuracy rates obtained from the previous analysis. As mentioned in the fMRI analysis, the findings allowed us to use the accuracy rates for model comparisons.

Experiment	Repetition	σ	MCADNet - Structural MRI								
			Precision			Recall			F1-Score		
			Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.
AD_NC	1	S0	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.97	0.98
		S2	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.97	0.98
		S3	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.96	0.98
		S4	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.96	0.98
	2	S0	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.96	0.98
		S2	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.96	0.98
		S3	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.96	0.98
		S4	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.96	0.98
	3	S0	0.99	0.98	0.99	0.99	0.96	0.99	0.99	0.97	0.99
		S2	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.97	0.98
		S3	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.97	0.98
		S4	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.97	0.98
	4	S0	0.99	0.98	0.99	0.99	0.97	0.99	0.99	0.97	0.99
		S2	0.99	0.98	0.99	0.99	0.97	0.99	0.99	0.97	0.99
		S3	0.98	0.97	0.98	0.98	0.95	0.98	0.98	0.96	0.98
		S4	0.97	0.95	0.97	0.97	0.93	0.97	0.97	0.94	0.97
	5	S0	0.96	0.95	0.96	0.96	0.9	0.96	0.96	0.92	0.96
		S2	0.97	0.95	0.97	0.97	0.93	0.97	0.97	0.94	0.97
		S3	0.95	0.94	0.95	0.95	0.87	0.95	0.95	0.9	0.95
		S4	0.97	0.95	0.97	0.97	0.92	0.97	0.97	0.93	0.97
AD_NC_MCI	1	S0	0.94	0.95	0.94	0.94	0.93	0.94	0.94	0.93	0.94
		S2	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.93	0.94
		S3	0.94	0.95	0.94	0.94	0.93	0.94	0.94	0.94	0.94
		S4	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.93	0.94
	2	S0	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.93	0.94
		S2	0.95	0.96	0.95	0.95	0.94	0.95	0.95	0.95	0.95
		S3	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.95
		S4	0.95	0.96	0.95	0.95	0.94	0.95	0.95	0.95	0.95
	3	S0	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.93	0.94
		S2	0.95	0.96	0.95	0.95	0.94	0.95	0.95	0.95	0.95
		S3	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.95
		S4	0.95	0.96	0.95	0.95	0.93	0.95	0.95	0.94	0.95
	4	S0	0.95	0.96	0.95	0.95	0.94	0.95	0.95	0.95	0.95
		S2	0.95	0.95	0.95	0.95	0.93	0.95	0.95	0.94	0.95
		S3	0.95	0.96	0.95	0.95	0.93	0.95	0.95	0.94	0.95
		S4	0.88	0.88	0.88	0.88	0.84	0.88	0.88	0.86	0.88
	5	S0	0.89	0.89	0.89	0.89	0.87	0.89	0.89	0.88	0.89
		S2	0.87	0.88	0.87	0.87	0.85	0.87	0.87	0.86	0.87
		S3	0.88	0.88	0.88	0.88	0.86	0.88	0.88	0.87	0.88
		S4	0.88	0.88	0.88	0.88	0.85	0.88	0.88	0.86	0.88
AD_MCI	1	S0	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95
		S2	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.95
		S3	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S4	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	2	S0	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S2	0.95	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S3	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S4	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	3	S0	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S2	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S3	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S4	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	4	S0	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S2	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S3	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
		S4	0.9	0.89	0.9	0.9	0.89	0.9	0.9	0.89	0.9
	5	S0	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
		S2	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
		S3	0.9	0.9	0.9	0.9	0.89	0.9	0.9	0.9	0.9
		S4	0.9	0.89	0.9	0.9	0.89	0.9	0.9	0.89	0.9
NC_MCI	1	S0	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.97	0.99
		S2	0.99	0.99	0.99	0.99	0.95	0.99	0.99	0.97	0.99
		S3	0.99	0.98	0.99	0.99	0.95	0.99	0.99	0.97	0.99
		S4	0.99	0.98	0.98	0.99	0.95	0.99	0.99	0.96	0.98
	2	S0	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.96	0.98
		S2	0.99	0.98	0.98	0.99	0.95	0.99	0.99	0.96	0.98
		S3	0.99	0.99	0.99	0.99	0.97	0.99	0.99	0.98	0.99
		S4	0.99	0.98	0.99	0.99	0.96	0.99	0.99	0.97	0.99
	3	S0	0.99	0.99	0.99	0.99	0.95	0.99	0.99	0.97	0.99
		S2	0.99	0.98	0.99	0.99	0.95	0.99	0.99	0.97	0.99
		S3	0.99	0.98	0.99	0.99	0.96	0.99	0.99	0.97	0.99
		S4	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.97	0.99
	4	S0	0.99	0.98	0.99	0.99	0.96	0.99	0.99	0.97	0.99
		S2	0.99	0.98	0.99	0.99	0.95	0.99	0.99	0.97	0.99
		S3	0.99	0.98	0.99	0.99	0.96	0.99	0.99	0.97	0.99
		S4	0.97	0.95	0.97	0.97	0.91	0.97	0.97	0.93	0.97
	5	S0	0.97	0.96	0.97	0.97	0.9	0.97	0.97	0.93	0.97
		S2	0.97	0.96	0.97	0.97	0.92	0.97	0.97	0.94	0.97
		S3	0.98	0.96	0.98	0.98	0.95	0.98	0.98	0.96	0.98
		S4	0.97	0.95	0.97	0.97	0.92	0.97	0.97	0.94	0.97

TABLE 8. DeepAD performance is measured using the same concept utilized in Table 7, and although the results demonstrate very high performance, they also show that MCADNNet performs better for multi-class classification.

Experiment	Repetition	σ	DeepAD - Structural MRI									
			Precision			Recall			F1-Score			
			Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.	Micro Avg.	Macro Avg.	Weighted Avg.	
AD_NC	1	S0	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		S2	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		S3	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		S4	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	2	S0	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S2	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95
		S3	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S4	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	3	S0	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S2	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S3	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S4	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	4	S0	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S2	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S3	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		S4	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	5	S0	0.89	0.89	0.89	0.89	0.89	0.88	0.89	0.89	0.88	0.89
		S2	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
		S3	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
		S4	0.89	0.89	0.89	0.89	0.89	0.88	0.89	0.89	0.89	0.89
AD_NC_MCI	1	S0	0.93	0.94	0.93	0.93	0.91	0.93	0.93	0.93	0.93	0.93
		S2	0.93	0.94	0.93	0.93	0.9	0.93	0.93	0.93	0.92	0.93
		S3	0.93	0.94	0.93	0.93	0.92	0.93	0.93	0.93	0.93	0.93
		S4	0.94	0.95	0.94	0.94	0.91	0.94	0.94	0.94	0.93	0.94
	2	S0	0.94	0.95	0.94	0.94	0.91	0.94	0.94	0.94	0.93	0.94
		S2	0.95	0.96	0.95	0.95	0.93	0.95	0.95	0.95	0.94	0.95
		S3	0.94	0.95	0.94	0.94	0.93	0.94	0.94	0.94	0.94	0.94
		S4	0.94	0.95	0.94	0.94	0.93	0.94	0.94	0.94	0.94	0.94
	3	S0	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.94	0.93	0.94
		S2	0.94	0.95	0.95	0.94	0.93	0.94	0.94	0.94	0.94	0.94
		S3	0.95	0.95	0.95	0.95	0.93	0.95	0.95	0.95	0.94	0.95
		S4	0.94	0.95	0.94	0.94	0.92	0.94	0.94	0.94	0.93	0.94
	4	S0	0.94	0.95	0.94	0.94	0.93	0.94	0.94	0.94	0.94	0.94
		S2	0.95	0.95	0.95	0.95	0.92	0.95	0.95	0.95	0.94	0.94
		S3	0.94	0.95	0.95	0.94	0.93	0.94	0.94	0.94	0.94	0.94
		S4	0.87	0.88	0.87	0.87	0.83	0.87	0.87	0.87	0.85	0.87
	5	S0	0.88	0.89	0.88	0.88	0.86	0.88	0.88	0.88	0.87	0.88
		S2	0.87	0.88	0.87	0.87	0.85	0.87	0.87	0.87	0.87	0.87
		S3	0.88	0.89	0.88	0.88	0.86	0.88	0.88	0.88	0.87	0.88
		S4	0.87	0.88	0.87	0.87	0.85	0.87	0.87	0.87	0.86	0.87
AD_MCI	1	S0	0.98	0.97	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
		S2	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.96	0.98
		S3	0.97	0.97	0.97	0.97	0.92	0.97	0.97	0.97	0.95	0.97
		S4	0.97	0.97	0.97	0.97	0.93	0.97	0.97	0.97	0.95	0.97
	2	S0	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.95	0.98
		S2	0.98	0.98	0.98	0.98	0.93	0.98	0.98	0.98	0.95	0.98
		S3	0.97	0.97	0.97	0.97	0.92	0.97	0.97	0.97	0.94	0.97
		S4	0.98	0.98	0.98	0.98	0.93	0.98	0.98	0.98	0.96	0.98
	3	S0	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0.97	0.98
		S2	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0.97	0.98
		S3	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
		S4	0.98	0.97	0.98	0.98	0.94	0.98	0.98	0.98	0.95	0.98
	4	S0	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
		S2	0.98	0.97	0.98	0.98	0.96	0.98	0.98	0.98	0.97	0.98
		S3	0.98	0.97	0.98	0.98	0.93	0.98	0.98	0.98	0.95	0.97
		S4	0.97	0.95	0.97	0.97	0.92	0.97	0.97	0.97	0.93	0.97
	5	S0	0.96	0.94	0.96	0.96	0.89	0.96	0.96	0.96	0.92	0.96
		S2	0.97	0.95	0.96	0.97	0.91	0.97	0.97	0.97	0.93	0.96
		S3	0.95	0.94	0.95	0.95	0.86	0.95	0.95	0.95	0.89	0.95
		S4	0.96	0.94	0.96	0.96	0.9	0.96	0.96	0.96	0.92	0.96
NC_MCI	1	S0	0.99	0.99	0.99	0.99	0.95	0.99	0.99	0.99	0.97	0.99
		S2	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.96	0.98
		S3	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.96	0.98
		S4	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
	2	S0	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.96	0.98
		S2	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.98	0.96	0.98
		S3	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.99	0.97	0.99
		S4	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
	3	S0	0.99	0.98	0.99	0.99	0.95	0.99	0.99	0.99	0.96	0.99
		S2	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
		S3	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96	0.98
		S4	0.99	0.99	0.99	0.99	0.95	0.99	0.99	0.99	0.97	0.99
	4	S0	0.99	0.99	0.99	0.99	0.96	0.99	0.99	0.99	0.97	0.99
		S2	0.99	0.99	0.99	0.99	0.95	0.99	0.99	0.99	0.97	0.99
		S3	0.99	0.98	0.99	0.99	0.95	0.99	0.99	0.99	0.97	0.99
		S4	0.97	0.95	0.97	0.97	0.9	0.97	0.97	0.97	0.92	0.97
	5	S0	0.97	0.96	0.97	0.97	0.9	0.97	0.97	0.97	0.93	0.97
		S2	0.97	0.96	0.97	0.97	0.91	0.97	0.97	0.97	0.93	0.97
		S3	0.98	0.95	0.98	0.98	0.95	0.98	0.98	0.98	0.95	0.98
		S4	0.97	0.95	0.97	0.97	0.91	0.97	0.97	0.97	0.93	0.97

Step1 : Calculate the number of AD or NC slices in a given class where i is slice number.

```

for  $i=1$  to  $N$  do
  if  $slice(i) \in AD$  then
    Counter AD +=1
  else if  $slice(i) \in NC$  then
    Counter NC +=1
end

```

Step2 : Calculate the probability of each class by dividing the number of slices in a class by the total number of slices.

$Probability(AD) = \text{Counter of AD} / \text{Total number of slices in the subject}$

$Probability(NC) = \text{Counter of NC} / \text{Total number of slices in the subject}$

Step3 : Compare the two probabilities for the AD and NC classes.

```

if  $Probability(AD) = Probability(NC)$  then
  Flag = Decision is to be made based on clinical
  measures.
if  $Probability(AD) > Probability(NC)$  then
  Flag = AD
else
  Flag = NC
end

```

Step4 : Vote for the majority and assign the label of “majority group” to the given subject. Assign Flag to the given subject.

liated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Company, Inc.; Meso ScaleDiagnostics, 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. (Danielle D. Desouza, John Anderson, and Cristina Saverino contributed equally to this work.)

APPENDIX

See Tables 2–8 and Figure 6.

REFERENCES

- [1] B. Reisberg, S. H. Ferris, A. Kluger, E. Franssen, J. Wegiel, and M. J. De Leon, “Mild cognitive impairment (MCI): A historical perspective,” *Int. Psychogeriatrics*, vol. 20, no. 1, pp. 18–31, 2008.
- [2] R. R. Tampi, D. J. Tampi, S. Chandran, A. Ghori, and M. Durning, “Mild cognitive impairment: A comprehensive review,” *Healthy Aging Res.*, vol. 4, pp. 1–11, Jan. 2015.
- [3] R. C. Petersen, J. C. Stevens, M. Ganguli, E. G. Tangalos, J. Cummings, and S. DeKosky, “Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American academy of neurology,” *Neurology*, vol. 56, no. 9, pp. 1133–1142, 2001.
- [4] T. Luck, M. Lupp, S. Briel, and S. G. Riedel-Heller, “Incidence of mild cognitive impairment: A systematic review,” *Dementia Geriatric Cogn. Disorders*, vol. 29, no. 2, pp. 164–175, 2010.
- [5] L. Jean, M.-É. Bergeron, S. Thivierge, and M. Simard, “Cognitive intervention programs for individuals with mild cognitive impairment: Systematic review of the literature,” *Amer. J. Geriatric Psychiatry*, vol. 18, no. 4, pp. 281–296, 2010.
- [6] R. C. Petersen, “Mild cognitive impairment as a diagnostic entity,” *J. Internal Med.*, vol. 256, no. 3, pp. 183–194, 2004.
- [7] L. G. Apostolova and J. L. Cummings, “Neuropsychiatric manifestations in mild cognitive impairment: A systematic review of the literature,” *Dementia Geriatric Cogn. Disorders*, vol. 25, no. 2, pp. 115–126, 2008.
- [8] R. Monastero, F. Mangialasche, C. Camarda, S. Ercolani, and R. Camarda, “A systematic review of neuropsychiatric symptoms in mild cognitive impairment,” *J. Alzheimer’s Disease*, vol. 18, no. 1, pp. 11–30, 2009.
- [9] D. Gallagher, C. E. Fischer, and A. Iaboni, “Neuropsychiatric symptoms in mild cognitive impairment: An update on prevalence, mechanisms, and clinical significance,” *Can. J. Psychiatry*, vol. 62, no. 3, pp. 161–169, 2017.
- [10] D. Fitzpatrick-Lewis, R. Warren, M. U. Ali, D. Sherifali, and P. Raina, “Treatment for mild cognitive impairment: A systematic review and meta-analysis,” *CMAJ Open*, vol. 3, no. 4, pp. E419–E427, 2015.
- [11] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, “Mild cognitive impairment represents early-stage alzheimer disease,” *Arch. Neurol.*, vol. 58, no. 3, pp. 397–405, 2001.
- [12] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, “Mild cognitive impairment: Clinical characterization and outcome,” *Arch. Neurol.*, vol. 56, no. 3, pp. 303–308, 1999.
- [13] M. Grundman, R. C. Petersen, S. H. Ferris, R. G. Thomas, P. S. Aisen, D. A. Bennett, N. L. Foster, C. R. Jack, Jr., D. R. Galasko, and R. Doody, “Mild cognitive impairment can be distinguished from alzheimer disease and normal aging for clinical trials,” *Arch. Neurol.*, vol. 61, no. 1, pp. 59–66, 2004.
- [14] C. R. Jack, Jr., V. J. Lowe, M. L. Senjem, S. D. Weigand, B. J. Kemp, M. M. Shiung, D. S. Knopman, B. F. Boeve, W. E. Klunk, C. A. Mathis, and R. C. Petersen, “ ^{11}C PiB and structural MRI provide complementary information in imaging of Alzheimer’ disease and amnestic mild cognitive,” *Brain*, vol. 131, no. 3, pp. 665–680, 2008.
- [15] C. R. Jack, Jr., V. J. Lowe, S. D. Weigand, H. J. Wiste, M. L. Senjem, D. S. Knopman, M. M. Shiung, J. L. Gunter, B. F. Boeve, B. J. Kemp, and M. Weiner, “Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer’s disease: Implications for sequence of pathological events in Alzheimer’s disease,” *Brain*, vol. 132, no. 5, pp. 1355–1365, 2009.
- [16] B. Winblad et al., “Mild cognitive impairment—Beyond controversies, towards a consensus: Report of the international working group on mild cognitive impairment,” *J. Internal Med.*, vol. 256, no. 3, pp. 240–246, 2004.
- [17] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies, and C. H. Phelps, “The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the national institute on aging-Alzheimer’s association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s Dementia*, vol. 7, no. 3, pp. 270–279, 2011.
- [18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
- [20] I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning—a new frontier in artificial intelligence research,” *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, Nov. 2010.
- [21] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, “Why does unsupervised pre-training help deep learning?” *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, Feb. 2010.
- [22] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

- [23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [27] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 5987–5995.
- [28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.
- [29] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: <https://arxiv.org/abs/1602.07360>
- [30] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," 2015, *arXiv:1511.00561*. [Online]. Available: <https://arxiv.org/abs/1511.00561>
- [31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [32] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr./May 2014, pp. 1015–1018.
- [33] H.-I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2013, pp. 583–590.
- [34] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.
- [35] H.-I. Suk, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, 2015.
- [36] H.-I. Suk and D. Shen, "Deep learning in diagnosis of brain disorders," in *Recent Progress in Brain and Cognitive Engineering*. Dordrecht, The Netherlands: Springer, 2015, pp. 203–213.
- [37] V. K. Ithapu, V. Singh, O. C. Okonkwo, R. J. Chappell, N. M. Dowling, and S. C. Johnson, "Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment," *Alzheimer's Dementia*, vol. 11, no. 12, pp. 1489–1499, 2015.
- [38] S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage, Clin.*, vol. 21, 2019, Art. no. 101645.
- [39] U. Senanayake, A. Sowmya, L. Dawes, N. A. Kochan, W. Wen, and P. S. Sachdev, "Deep learning approach for classification of mild cognitive impairment subtypes," in *Proc. ICPHAM*, 2017, pp. 655–662.
- [40] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," 2015, *arXiv:1502.02506*. [Online]. Available: <https://arxiv.org/abs/1502.02506>
- [41] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, and M. J. Fulham, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.
- [42] E. Arvesen, "Automatic classification of Alzheimer' disease from structural MRI," M.S. thesis, Dept. Comput. Sci., Østfold Univ. College, Halden, Norway, 2015.
- [43] F. Liu and C. Shen, "Learning deep convolutional features for MRI based Alzheimer's disease classification," 2014, *arXiv:1404.3366*. [Online]. Available: <https://arxiv.org/abs/1404.3366>
- [44] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, M. Fulham, and D. Feng, "High-level feature based PET image retrieval with deep learning architecture," *J. Nucl. Med.*, vol. 55, p. 2028, May 2014.
- [45] T. Brosch and R. Tam, "Manifold learning of brain MRIs by deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2013, pp. 633–640.
- [46] L. Rampasek and A. Goldenberg, "TensorFlow: Biology's gateway to deep learning?" *Cell Syst.*, vol. 2, no. 1, pp. 12–14, 2016.
- [47] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 20–28.
- [48] E. P. Ijima and C. K. Mohan, "Hybrid deep neural network model for human action recognition," *Appl. Soft Comput.*, vol. 46, pp. 936–952, Sep. 2016.
- [49] S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, "Fusion of deep learning models of MRI scans, minimal state examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's Dementia, Diagnosis, Assessment Disease Monit.*, vol. 10, pp. 737–749, Jan. 2018.
- [50] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, and X. and Qu, "Convolutional neural networks-based MRI Image Analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers Neurosci.*, vol. 12, p. 777, Nov. 2018.
- [51] M. S. Mazrina, K. Chiotis, E. Colato, A. K. Nordberg, and E. Rodriguez-Vieitez, "Modelling the associations between [18F] AV1451,[18F] FDG PET and cognition in mild cognitive impairment and ad dementia," *Alzheimer's Dementia, J. Alzheimer's Assoc.*, vol. 14, no. 7, pp. P1573–P1574, 2018.
- [52] D. Wen, Z. Wei, Y. Zhou, G. Li, X. Zhang, and W. Han, "Deep learning methods to process fMRI data and their application in the diagnosis of cognitive impairment: A brief overview and our opinion," *Frontiers Neuroinform.*, vol. 12, p. 23, Apr. 2018.
- [53] A. Srinivasan, P. Battacharjee, A. P. I, and G. Sanyal, "Brain MR image analysis using discrete wavelet transform with fractal feature analysis," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 1660–1664.
- [54] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [55] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.
- [56] S. Sarraf and G. Tofghi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Proc. Future Technol. Conf. (FTC)*, Dec. 2016, pp. 816–820.
- [57] S. Sarraf et al., "DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI," *BioRxiv*, 2016, Art. no. 070441.
- [58] E. Hosseini-Asl, R. Keynto, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," 2016, *arXiv:1607.00455*. [Online]. Available: <https://arxiv.org/abs/1607.00455>
- [59] N. Amoroso, D. Diacono, A. Fanzini, M. La Rocca, A. Monaco, A. Lombardi, C. Guaragnella, R. Bellotti, and S. Tangaro, "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge," *J. Neurosci. Methods*, vol. 302, pp. 3–9, May 2018.
- [60] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.
- [61] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituev, T. P. Copeland, M. S. Aboian, C. M. Aparici, S. C. Behr, R. R. Flavell, S.-Y. Huang, K. A. Zalocusky, L. Nardo, Y. Seo, R. A. Hawkins, M. H. Pampaloni, D. Hadley, and B. L. Franc, "A deep learning model to predict a diagnosis of Alzheimer disease by using ¹⁸F-FDG pet of the brain," *Radiology*, vol. 290, no. 2, pp. 456–464, 2018.
- [62] S. Basaia, F. Agosta, L. Wagner, G. Magnani, and M. Filippi, "Automatic classification of patients with Alzheimer's disease (AD) and mild cognitive impairment (MCI) who will convert to AD using deep neural networks," p. 179, 2018.
- [63] E. Jabson, M. O. Ahmad, and M. N. S. Swamy, "Deep structural and clinical feature learning for semi-supervised multiclass prediction of Alzheimer's disease," in *Proc. IEEE 61st Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2018, pp. 791–794.
- [64] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 1952.
- [65] S. Spasov, L. Passamonti, A. Duggento, P. Lió, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019.
- [66] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage, Clin.*, vol. 21, Jan. 2019, Art. no. 101645.

- [67] K. R. Kruthika, A. P. Rajeswari, and H. D. Maheshappa, "CBIR system using capsule networks and 3D CNN for Alzheimer's disease diagnosis," *Inform. Med. Unlocked*, vol. 14, pp. 59–68,



SAMAN SARRAF (SM'17) received the B.Sc. degree in electrical and biomedical engineering and the M.Sc. degree in biomedical engineering from Science and Research University - Tehran, Iran, and the master's degree in electrical and computer engineering from McMaster University, Canada, in 2013.

He is a Senior Machine Learning Researcher working on industry-level projects in Silicon Valley - California. He has been with the IEEE for more than ten years. Formerly, he was with a Cognitive Aging Lab Manager with the Rotman Research Institute at Baycrest, which is affiliated with the University of Toronto. In 2016, he received the Best Paper Award from the IEEE FTC Conference, San Francisco, for his research into deep learning applications to recognize Alzheimer's disease using functional MRI data.



DANIELLE D. DESOUZA received the M.Sc. degree from the Department of Rehabilitation Science, University of Toronto, in 2009, and the Ph.D. degree from the Institute of Medical Science, University of Toronto, in 2015, both in neuroscience.

She is currently an Instructor with the Department of Neurology and Neurological Sciences, Stanford University. Her research focuses on multimodal neuroimaging to understand brain structure and function in patients with chronic pain.

Her research interest includes noninvasive neuromodulation approaches for the treatment of chronic pain and depression. She received several awards and honors, including the Canadian Institutes of Health Research (CIHR) Postdoctoral Fellowship, the CIHR Cell Signals Strategic Training Fellowship, and the Ontario Graduate Scholarship. She was named as a 2018 Massachusetts Institute of Technology Rising Star in Biomedical.



JOHN A. E. ANDERSON received the Ph.D. degree from the University of Toronto, in 2016. His Ph.D. work focused on how circadian arousal affects cognitive performance and brain network integrity in the elderly. He was a Postdoctoral Fellow with York University, examining language effects on cognitive reserve in multimodal MRI. He is currently with the Centre for Addiction and Mental Health (CAMH), where he is examining multishell MRI acquisitions from dementia populations. He has more than 22 publications. His research interests include aging, functional and structural MRI, and interventions leading to cognitive reserve in elderly populations.



CRISTINA SAVERINO received the Ph.D. degree from the Department of Psychology, University of Toronto, in 2015. Her major research focused on aging studies and brain connectivity. She is currently a Postdoctoral Fellow with Toronto Western Hospital, University Health Network.

...