

# Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data

Harshit Parmar<sup>Ⓞ</sup>,<sup>a,\*</sup> Brian Nutter,<sup>a</sup> Rodney Long,<sup>b</sup> Sameer Antani<sup>Ⓞ</sup>,<sup>b</sup> and Sunanda Mitra<sup>a</sup>

<sup>a</sup>Texas Tech University, Department of Electrical and Computer Engineering, Lubbock, Texas, United States

<sup>b</sup>National Institutes of Health, Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, Maryland, United States

## Abstract

**Purpose:** Through the last three decades, functional magnetic resonance imaging (fMRI) has provided immense quantities of information about the dynamics of the brain, functional brain mapping, and resting-state brain networks. Despite providing such rich functional information, fMRI is still not a commonly used clinical technique due to inaccuracy involved in analysis of extremely noisy data. However, ongoing developments in deep learning techniques suggest potential improvements and better performance in many different domains. Our main purpose is to utilize the potentials of deep learning techniques for fMRI data for clinical use.

**Approach:** We present one such synergy of fMRI and deep learning, where we apply a simplified yet accurate method using a modified 3D convolutional neural networks (CNN) to resting-state fMRI data for feature extraction and classification of Alzheimer's disease (AD). The CNN is designed in such a way that it uses the fMRI data with much less preprocessing, preserving both spatial and temporal information.

**Results:** Once trained, the network is successfully able to classify between fMRI data from healthy controls and AD subjects, including subjects in the mild cognitive impairment (MCI) stage. We have also extracted spatiotemporal features useful for classification.

**Conclusion:** This CNN can detect and differentiate between the earlier and later stages of MCI and AD and hence, it may have potential clinical applications in both early detection and better diagnosis of Alzheimer's disease.

© 2020 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.7.5.056001](https://doi.org/10.1117/1.JMI.7.5.056001)]

**Keywords:** 3D convolutional neural networks; Alzheimer's disease; clinical functional magnetic resonance imaging; deep learning; spatiotemporal functional magnetic resonance imaging features.

Paper 20077R received Apr. 2, 2020; accepted for publication Oct. 7, 2020; published online Oct. 27, 2020.

## 1 Introduction

Magnetic resonance imaging (MRI) is one of the most advanced non-invasive neuroimaging techniques in common use. In 1990, Ogawa et al.<sup>1</sup> found that MRI can be used to obtain functional information about the brain. This modality is commonly known as functional MRI (fMRI). fMRI measures indirect effects of neuronal activities in a brain region due to changes in magnetic susceptibility between oxygenated and deoxygenated blood. This type of signal is called the blood oxygenation level dependent (BOLD) signal. Being paramagnetic, deoxyhemoglobin causes slightly more local inhomogeneity in the local magnetic field, in turn causing slightly faster dephasing of the spins (of hydrogen ions in the water molecules) compared to the diamagnetic oxyhemoglobin. This change in spin dephasing causes a small change in local T2\*

---

\*Address all correspondence to Harshit Parmar, [harshit.parmar@ttu.edu](mailto:harshit.parmar@ttu.edu)

relaxation time, but it is enough to measure by statistical analysis of the MRI signal. BOLD signal changes are observed in both task fMRI and rest fMRI.<sup>2,3</sup>

In spite of providing such rich functional information, fMRI is not yet a common clinical technique for diagnosis<sup>4,5</sup> due to two major problems. First, fMRI data are highly susceptible to noise.<sup>6</sup> The useful BOLD signal change amounts to only 2% to 5% of the absolute signal intensity. Useful information is extracted from the raw fMRI data using sophisticated statistical analysis. Even a small amount of noise or artifact can change the result of the analysis significantly. This becomes a major issue when dealing with clinical analysis. A small artifact can lead to false interpretation and affect the diagnosis. Second, the fMRI data structure is very complicated. A traditional x-ray or an anatomical MRI scan can readily be visualized and assessed by medical professionals. fMRI is difficult to visualize in its entirety without efficient processing of the statistically analyzed data.

Machine learning techniques such as deep learning algorithms have been recently developed to extract relevant information for classification of data in various fields. Because of excellent performance in image processing and computer vision applications, we anticipate great potential for using deep learning techniques for other applications. Here, we attempted to establish a synergy between fMRI and deep learning, and investigate this procedure as a step toward clinical application of fMRI. However, some challenges remain in applying deep learning algorithms directly to fMRI data. The majority of the deep learning algorithms, including the well-known convolutional neural networks (CNNs), were designed for images. fMRI data on the other hand are four dimensional, with three spatial and one temporal dimension. Both the network and the data must undergo some preprocessing to make them compatible with each other. Some researchers have tried to use CNNs on modified fMRI image data, but limitations to those solutions have been noted.<sup>7</sup> In addition, a tradeoff exists between the preprocessing complexity of the fMRI data and the complexity of the network. Moreover, in the case of fMRI, we do not have the privilege of a large quantity of training subjects. A typical fMRI study may only have few tens of subjects. With open-source data banks, such as Alzheimer's Disease Neuroimaging Initiative (ADNI)<sup>8</sup> and OpenfMRI,<sup>9</sup> one has access to additional fMRI data, but that is still not enough in many cases.

In this study, we designed and trained a volumetric 3D CNN to do multiclass classification. The CNN is able to extract spatiotemporal features from a given section of BOLD fMRI data, and then use those features to classify the data into either Alzheimer's disease (AD), mild cognitive impairment (MCI), or both early cognitive impairment (EMCI) and late cognitive impairment (LMCI), and healthy controls (CN). This is a modified extension of our previous work where we tested the validity of the network design using a binary classifier.<sup>10</sup> Two main reasons encourage working with AD. First, AD is the most common neurological disorder found in humans. After the age of 60, the chances of having AD doubles nearly every five years.<sup>11</sup> Second, a large amount of data is available from the ADNI database.

Currently, no cure exists for AD. After the brain tissue damage has occurred, it cannot be reversed with medication. However, if detected in an earlier stage, the rate of progression of AD can be reduced and controlled by the use of medication. The proposed classifier can identify MCI, which is associated with high risk of progression to AD. Detection of MCI can serve as a potential application toward an earlier detection tool for AD. The aim of this study is to take a step toward developing fMRI as a clinical technique for early detection of AD.

## 2 Methodology

### 2.1 Dataset and Preprocessing

All the fMRI volumes used for this study were acquired from the ADNI online database. Resting-state fMRI volumes were acquired for a total of 120 subjects over four classes—AD, LMCI, EMCI, and CN. Each class contained 30 subjects. The data were grouped into classes with the following parameters: gender (F/M); age (mean  $\pm$  standard deviation).

- AD – 16F/14M, age:  $73.5 \pm 7.5$  years.
- CN – 14F/16M, age:  $75 \pm 6.5$  years.

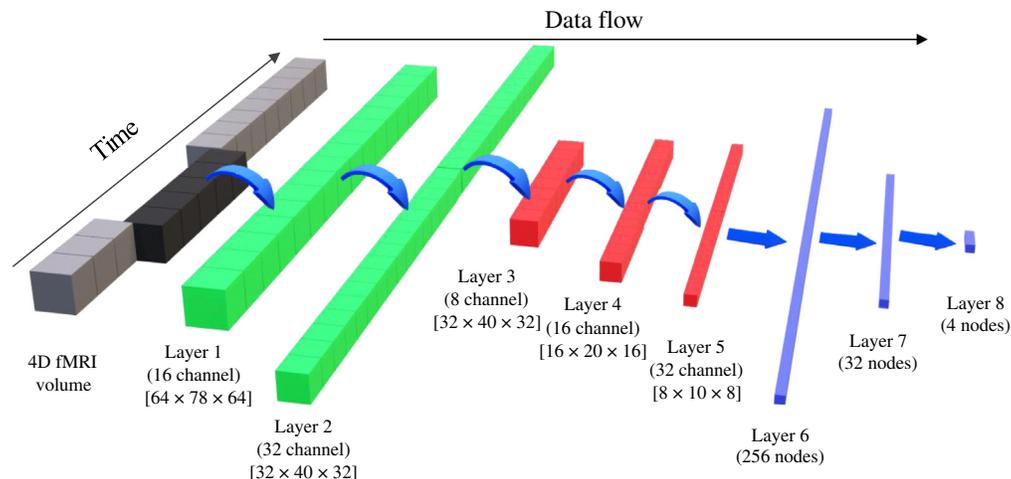
- EMCI – 17F/13M, age:  $72 \pm 7.5$  years.
- LMCI – 12F/18M, age:  $72.5 \pm 9$  years.

Each resting-state volume was acquired on a 3T Philips MRI Medical System. The flip angle was 80 deg with an echo time and repetition time of 30 ms and 3000 ms, respectively. About 140 full brain volumes were acquired, accounting for a time series of 140 time points for each voxel. Each volume was acquired with 48 axial slices of slice thickness of 3.313 mm. Each slice has a matrix size of  $64 \times 64$ , with in-plane resolution of  $3.125 \text{ mm} \times 3.125 \text{ mm}$ .

All the fMRI data went through the standard preprocessing pipeline before being input to the CNN. All the preprocessing was done in MATLAB 2019a and its SPM12 toolbox.<sup>12</sup> The preprocessing included motion correction, intensity normalization, coregistration, global mean and temporal signal drift removal, and intensity-based thresholding and masking. The first step of the preprocessing was motion correction. All the fMRI volumes were corrected for motion, with the reference taken as the temporal mean volume. For three of the subjects (two in AD and one in EMCI), the motion was more than one voxel, so they were replaced by other volumes from the ADNI database. The fMRI volume then went through intensity normalization. Next, all the fMRI volumes were coregistered to the Montreal Neurological Institute (MNI) space to ensure that all the subject's data were in same spatial size and dimensions. The MNI atlas used for coregistration had a size of  $64 \times 78 \times 64$  with a voxel size of  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ . After all the volumes were in the same MNI space, we applied a binary mask to remove the background outside the region of interest. The same binary mask, in the MNI space, was used for all volumes to provide uniformity for further analysis. Finally, temporal signal drift and global signal fluctuations were removed using a PCA-based approach.<sup>13</sup>

## 2.2 CNN Architecture

The architecture of the 3D CNN used is shown in Fig. 1. The left-most block shows the entire preprocessed fMRI volume. From that, a section of five consecutive brain volumes is given as input to the CNN. By sampling a smaller subset of brain volumes, we can increase the number of training samples available. For this study, we obtained 20 samples from each fMRI volume. The size of each input is  $64 \times 78 \times 64$  with five channel depth. The size of each volume corresponds to the reshaped standard MNI atlas, with a voxel size of  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ . The reason for a standard input volume size is that fMRI volumes from any scanner can be registered to MNI space and made compatible to the network. This scaling increases the generalization of this network, which is useful in working with clinical data across different sites and scanners. Moreover, the spatial size of  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$  MNI atlas is  $181 \times 217 \times 181$  while the



**Fig. 1** 3D CNN architecture with network parameters.

**Table 1** CNN parameters.

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
No. of filters	16	32	8	16	32
Filter size	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$5 \times 5 \times 5$	$3 \times 5 \times 3$	$3 \times 3 \times 3$
Filter stride	[1,1,1]	[1,1,1]	[1,1,1]	[1,1,1]	[1,1,1]
Padding	[0,1,0]	[0,0,0]	[0,0,0]	[0,0,0]	[0,0,0]
Pool size	$2 \times 2 \times 2$	—	$2 \times 2 \times 2$	$2 \times 2 \times 2$	$2 \times 2 \times 2$
Pool stride	[2,2,2]	—	[2,2,2]	[2,2,2]	[2,2,2]
Pool type	Max	—	Max	Max	Max
Batch norm	No	Yes	No	No	Yes

spatial size of the reshaped  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$  MNI atlas is  $64 \times 78 \times 46$  which is closer to a typical resting-state fMRI volume size.

For the network architecture, the first convolutional layers have a filter size of  $1 \times 1 \times 1$  with a depth of 5. The aim of this filter is to capture just the temporal aspect of the data, which in this case is across the channels. About 32 different  $1 \times 1 \times 1$  kernels were learned by the network, accounting for 32 different temporal features. This is equivalent of learning 32 unique regions of interest (ROIs) using the BOLD signal from all voxels. The second convolutional layer also have a kernel size of  $1 \times 1 \times 1$  with a depth of 32. Again, this layer learns higher level temporal features which are obtained by nonlinear combination of features obtained by first layer. The first and second layer together extracts temporal features from the BOLD fMRI volume. After that, the learned temporal features are passed through spatial convolutional layers at three different spatial scales: half scale ( $32 \times 40 \times 32$ ), quarter scale ( $16 \times 20 \times 16$ ), and one-eighth scale ( $8 \times 10 \times 8$ ). At the end of the third spatial convolutional layer, the network outputs learned spatiotemporal features. These spatiotemporal features are then used by the fully connected dense layers to perform classification.

Each convolution layer in the network is followed by a leaky-rectified linear unit activation, where the scaling factor for negative inputs is 0.1. This activation is followed by a pooling layer for all layers, except for the second temporal convolution layer. The pooling is done on a 3D window of size  $2 \times 2 \times 2$ , with no overlap between adjacent windows. Thus, each pooling layer reduces the spatial size by half. All the pooling layers are max-pooling layers. The last temporal and spatial convolutional layers, second and fifth, are followed by a batch normalization layer. The details of kernel size, number of filters, stride, and pooling are shown in Table 1.

After the spatial convolution and pooling layers, the CNN has three fully connected layers (layers 6, 7, and 8) with 256, 32, and 4 neurons in each layer, respectively. Each fully connected layer has a random dropout rate of 50% to prevent overfitting to the training data. After the last fully connected layer, the output is passed through the SoftMax layer, which computes the prediction probabilities for each class. The network has a total of eight learnable layers.

### 2.3 Training Parameters

The network was designed and trained using MATLAB 2019a on a Windows machine with 32 GB RAM and 4 GB NVIDIA GPU card. About 30 subjects for each class and 20 samples from each subject's fMRI data makes a total of 2400 data samples and 600 samples per class. For performing a five-fold cross validation, samples from each class were divided into five folds, 120 samples per fold per class. Then the network was trained using four folds from each class and tested on the remaining fold. This was done on all five folds. Once five-fold cross validation was performed, main network training was done. For that, samples from each class were distributed in three categories, about two-thirds for training (400 samples/class), one-sixth for validation (100 samples/class), and one-sixth for testing (100 samples/class). The network training was

done using minibatch stochastic gradient descent optimizer with momentum. The training was done for a total of 100 epochs with minibatch of 10 samples. The initial learning rate was set to 0.01. After every 10 epochs, the learning rate was reduced by a factor of 0.5. The momentum was set to 0.9. To prevent overfitting, L2 regularization was used with a regularization constant of 0.1. To prevent the network from learning correlations within minibatches, all the training samples were randomized after each epoch. The validation frequency was set to 100 iterations. The total training time for the network was slightly more than 1200 h.

### 3 Results

The network performance was validated using five-fold cross validation. The mean testing accuracy for five-fold cross validation is 98.96 %. Along with that, for each fold, we also computed class-wise performance parameters such as accuracy, precision, sensitivity, specificity, and F1 score for each of the four classes. Figure 2 shows the class-wise box plots for each performance parameters.

From the figure, we observe that the network is able to perform well on the testing data. Once validation was done, the network was trained on two-thirds of the entire data set. The trained network was able to classify a given section of 4D BOLD fMRI data into either AD, CN, EMCI, or LMCI. The training (1600 samples), validation (400 samples), and testing (400 samples) accuracies are 99.4%, 96.75%, and 93.00%, respectively. For validation and testing data, we also computed the confusion matrix as shown in Fig. 3. Apart from that, performance parameters such as precision, sensitivity, specificity, F1-score, and area under the curve were computed for each class. Figure 4 shows the graph of performance parameters while Fig. 5 shows the receiver operating characteristic (ROC) curve for all classes.

The classification accuracy obtained by our approach is comparable, in fact slightly better than previously published studies on AD classification using 4D fMRI data. Table 2 shows a

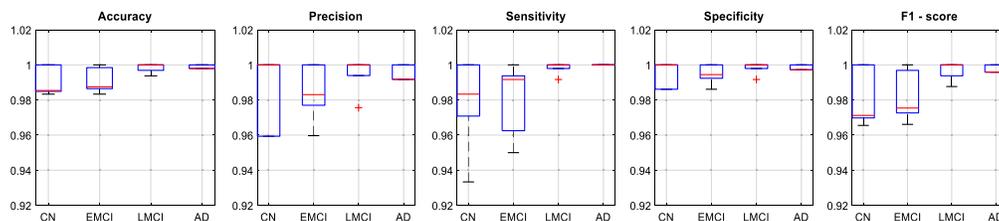


Fig. 2 Five-fold cross-validation performance parameters.

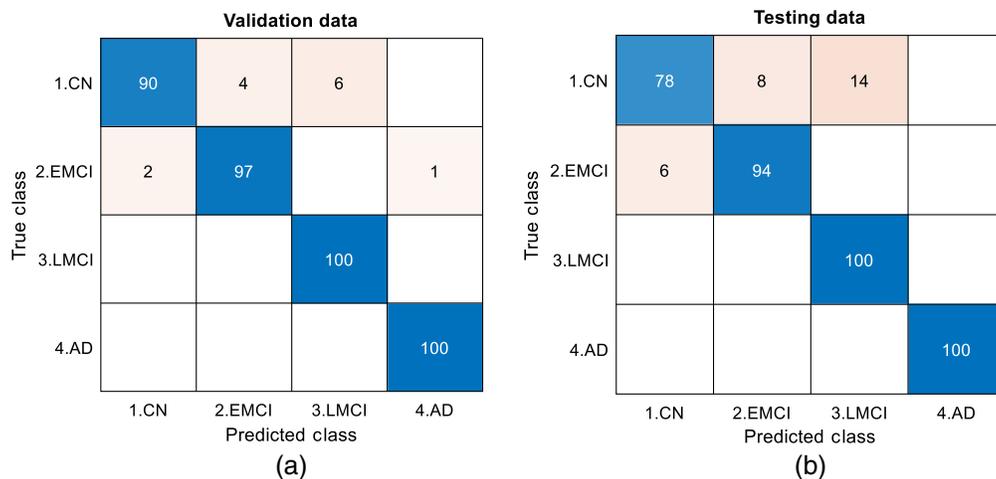
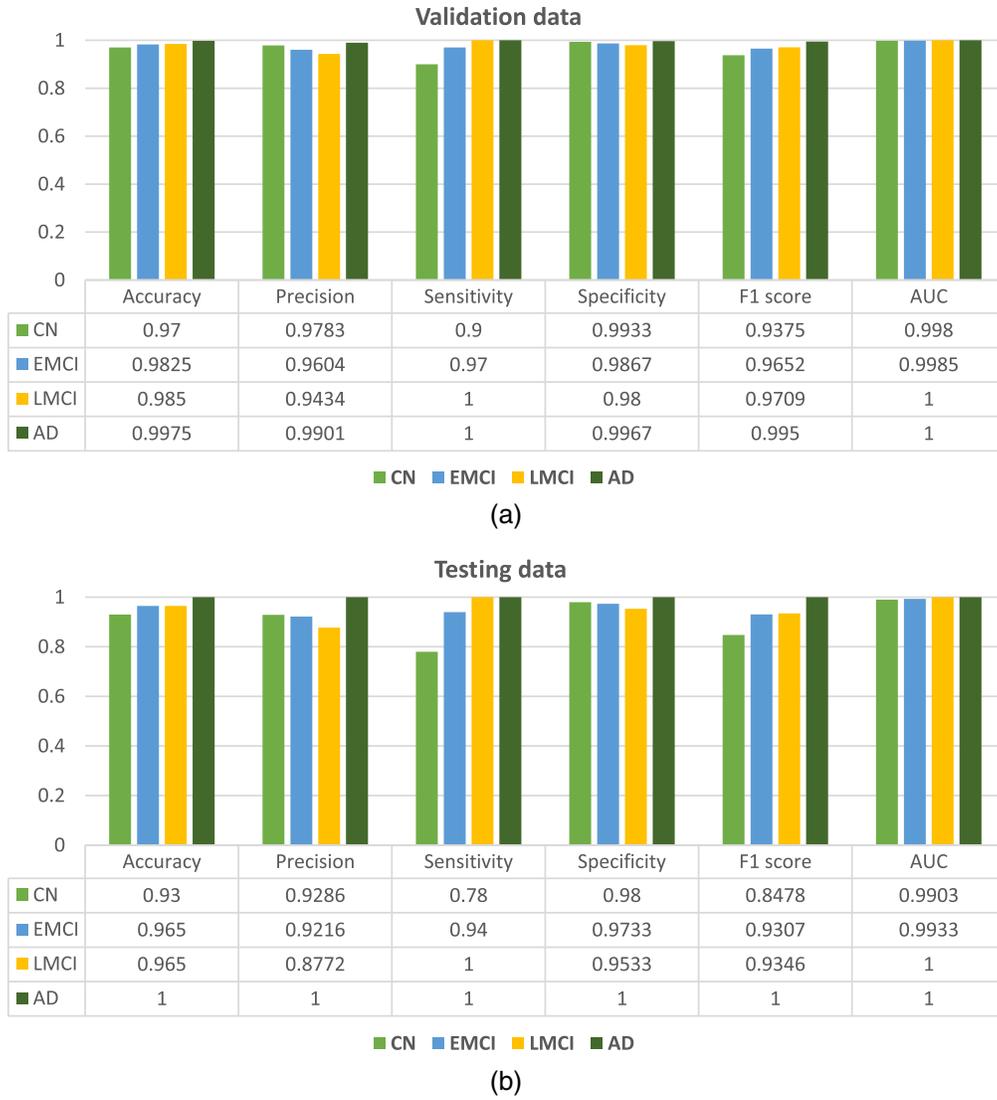
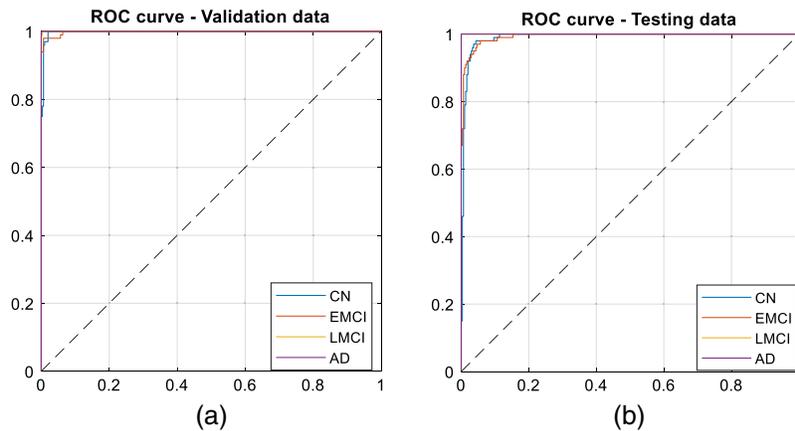


Fig. 3 Confusion matrix for (a) validation and (b) testing dataset.



**Fig. 4** Network performance for (a) validation and (b) testing dataset.



**Fig. 5** ROC curve for (a) validation and (b) testing dataset.

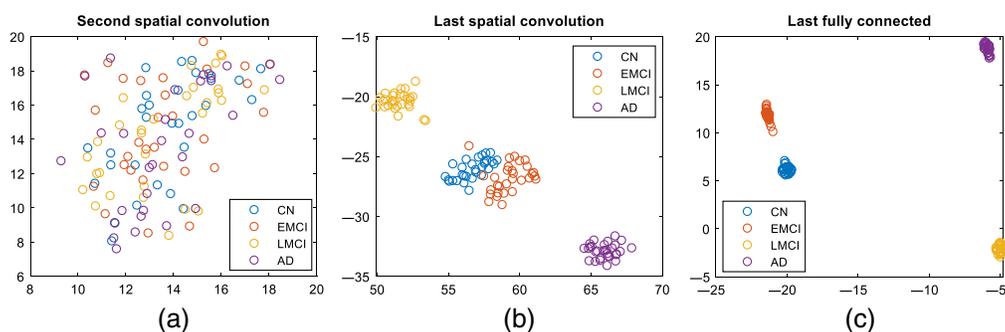
**Table 2** Comparison of few Alzheimer's disease classifiers based on modality of input data and type of classifier.

Study	Modality	Type	Accuracy (%)
Sarraf et al. <sup>7</sup>	fMRI-2D	Binary	96.80
Kazemi et al. <sup>14</sup>	fMRI-2D	Multiclass	97.64
Billones et al. <sup>15</sup>	fMRI-2D	Binary	98.33
		Multiclass	91.85
Jain et al. <sup>16</sup>	MRI-2D	Binary	99.14
		Multiclass	95.73
Basaia et al. <sup>17</sup>	MRI-2D	Binary	99.20
Hosseini-Asl et al. <sup>18</sup>	MRI-3D	Binary	97.60
		Multiclass	89.10
Duc et al. <sup>19</sup>	fMRI	Binary	85.27
Forouzaneshad et al. <sup>20</sup>	MRI + Positron Emission Tomography + Diffusion Imaging	Binary	94.70
Li et al. <sup>21</sup>	fMRI-4D	Binary	97.30
		Multiclass	89.47
Parmar et al. <sup>10</sup>	fMRI-4D	Binary	94.58
Current study	fMRI-4D	Multiclass	93

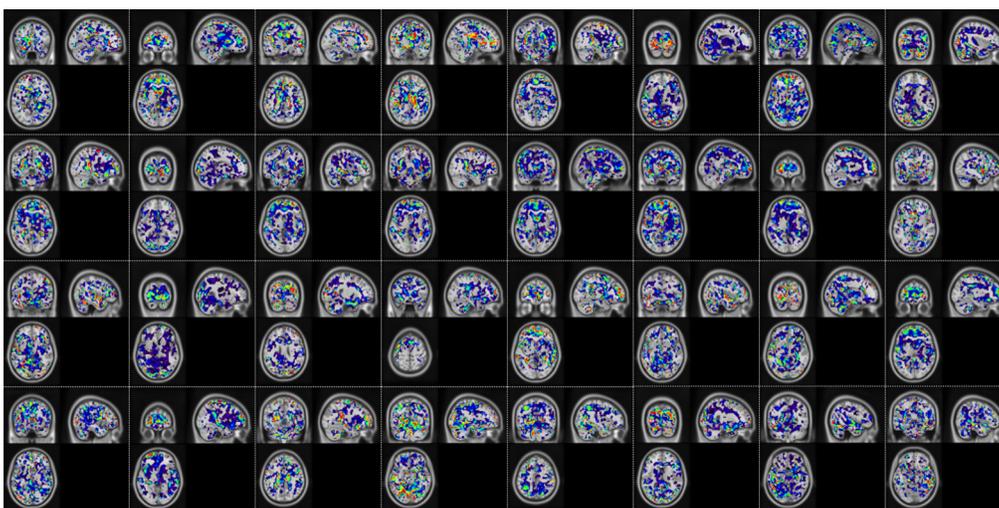
comparison of various binary and multiclass classifiers for AD and MCI using MRI and fMRI data. In the table, we first show studies that uses 2D images from the fMRI data and use image-based networks for classifications. Because these approaches use a single image to make predictions, it does not consider temporal or volumetric neighborhood from the fMRI data. Some studies use 3D structural MRI but here also, classification is made mainly based on anatomical differences. Radiologists can identify progression of AD using anatomical differences. But by the time anatomical differences are visible, the disease has already progressed a lot further. Studies have shown that functional changes are observed in brain networks and activation even during earlier stages of dementia.<sup>22,23</sup> Thus, for early detection application, classifiers should use both functional and structural information rather than just structural. Lastly, there are some recent studies that uses both spatial and temporal information for making the classification. The proposed approach has a higher classification accuracy for a multiclass type classifier which uses the entire 4D fMRI data without significant modifications (Fig. 6).

We also used t-distributed stochastic neighbor embedding (t-SNE) to visualize the classification of different classes. Figure 6 shows the t-SNE visualization for the outputs of the second and the last spatial convolutional layers and the output of the last fully connected layers. Each point in the plot represents a subject, and each class is shown in a different color. We can observe from the figure that through the last convolutional layer, the difference between the classes was not very dominant, and the difference became more and more dominant with the addition of fully connected layers. This observation indicates that the convolutional layers contributes more to feature extraction, while the fully connected layers contribute more toward classification.

Along with the classification performance parameters, we also show the extracted temporal features. The temporal features were extracted as the network activation maps at the output of both the temporal convolutional layers. Here we show the network activation maps for one of the subjects in Figs. 7 and 8. Network activation from just a single random subject is shown because, as seen from t-SNE results, in earlier layers there does not exist a significant difference between



**Fig. 6** t-SNE representation of all subjects at output of different layers in the network. We can see a progression within the network where successive layers introduce more and more difference between different classes (each dot is a subject; different classes are shown by different colors).

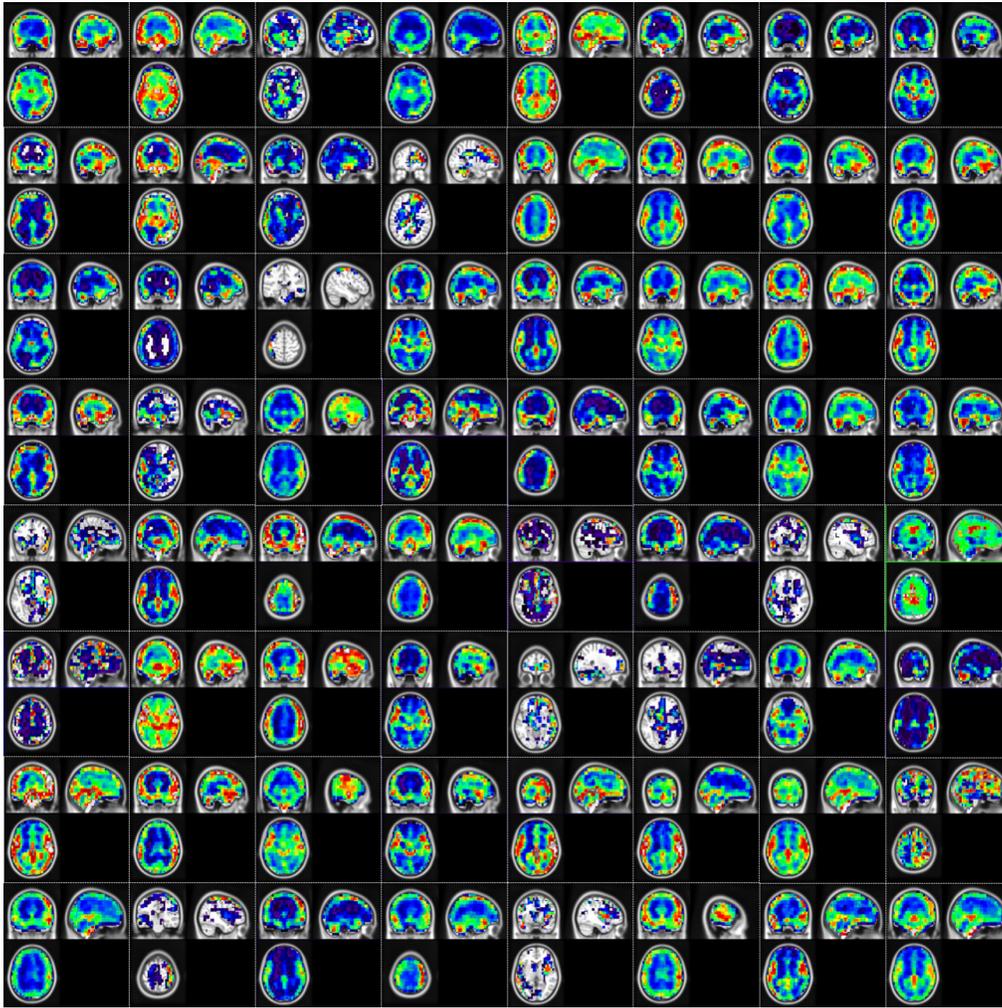


**Fig. 7** Network activation map from the output of first temporal convolution layer (32 channels) mapped onto MNI brain atlas. The color coding indicates the absolute response after the convolution and non-linearity operation, where red being higher and blue being lower.

different classes and all the subjects would have a similar response. From Fig. 7 we can observe that the first convolutional layer extracts regions having similar temporal response. The regions or ROIs are small, scattered, and without any definite structure. All 32 channels learn different temporally similar regions. Figure 8 shows the output of the second convolutional layer, all 64 channels. Here, the second convolution combines different ROIs from the first layer to obtain more homogenous and definite regions. One more thing to observe is that in the output of most of the channels, the main voxels of interests are from the gray matter region. This is a good sign as it suggests that the network classifies based on temporal BOLD fluctuations, as we expect BOLD changes mainly in the gray matter region.

#### 4 Discussion

The uniqueness of this approach lies in the way the network is designed. 3D CNNs with spatio-temporal feature extractors have been used recently for video processing and classification.<sup>24,25</sup> However, when applied to functional neuroimaging data, the features of interest are a little different than features extracted from a natural video. Recently, 3D convolutions have been used for classification of neurological disorders, however, there are some factors that make this approach different from those. The most important aspect is the architecture of the CNN. It consists of three main parts: the temporal feature extractor, the spatial feature extractor, and the classifier.



**Fig. 8** Network activation map from the output of second temporal convolution layer (64 channels) mapped onto MNI brain atlas. The color coding indicates the absolute response after the convolution and non-linearity operation, where red being higher and blue being lower.

The first two convolutional layers extract features based only on temporal changes without any contribution from spatial structure. This is achieved by having a kernel with size  $1 \times 1 \times 1 \times N$ , where  $N$  is the number of channels. Kernels with filter size  $1 \times 1 \times 1$  are generally used for depth reduction or combining outputs of different kernels.<sup>26</sup> Unlike that, here the  $1 \times 1 \times 1$  filters are actually used for extracting temporal features from the input data. Such use of  $1 \times 1 \times 1$  kernels for temporal feature extraction is not common in 3D CNN. For the first layer, the output is based solely on the temporal changes of any given voxel. Thus, voxels having similar time series would have similar response to the convolutions. Thus, the output at different channels can be treated as the extracted ROIs, which is nothing but low-level temporal features. The second temporal convolutional layer combines these low-level temporal features to come up with higher level temporal features, which can be understood as complex combination of ROIs. Again, because the kernel size here is  $1 \times 1 \times 1 \times N$ , it has no contribution from spatial structure. This unique kernel size in the first two layers prevents the network from learning any anatomical correlations within a subject. Because of that, different samples from a same fMRI time series may end up having different temporal response based on the voxel-wise temporal fluctuations. Moreover, because of preprocessing steps, such as global signal and temporal drift removal, the input to the network is BOLD signal change volumes without any anatomical intensity information.

Next, the learned temporal features are given as inputs to the following three spatial convolutional layers. Here, the network extracts spatial features at different spatial scales from the

already extracted temporal features. At the end of the 3D convolutional layers, we end up with spatiotemporal features which are then further used to make classification. Traditionally for a typical resting-state feature extractor, we first parcellate the brain into functionally similar ROIs using either atlas or such techniques as spatial independent component analysis (ICA). We then compute a connectivity or a covariance matrix using pairwise similarity between representative time series of each regions. The covariance matrices for different time windows act as features for a classification algorithms. Features extracted in such a fashion require accurate functional parcellation. Unlike that process, in this algorithm, the entire process of functional parcellation and feature extraction is learned by the network. The t-SNE plots show that the convolutional layers act more like a feature extractor, while the later fully connected layers contribute more toward the classification task. The advantage of using deep learning is that it can extract complex useful features on its own that might be otherwise difficult to extract using traditional feature extraction techniques.

Our approach also makes use of the comparatively least altered BOLD signal, with a minimum amount of preprocessing. An earlier study uses individual axial images as inputs to a relatively simpler network, AlexNet5, to classify AD from CN.<sup>7</sup> Though such an approach achieved high classification accuracy, 96.86%, it does not take into consideration the true dynamic information. Because individual images are the input to the network, the decision is made mainly based on anatomical features rather than dynamic temporal features. Some studies do take into account the temporal information but not in a direct manner. Li et al.<sup>27</sup> use a two-channel 3D CNN for autism spectrum disorder classification and feature extraction. To account for temporal information, they compute the mean and standard deviation for all voxels and use those 3D volumes as input to the network. Riaz et al.<sup>28</sup> did use an end to end deep network for fMRI data classification named DeepFMRI, but the complexity of the network makes it difficult in practice for clinical application. The entire network has three main parts. The first part is the feature extractor network, which uses 90 CNN to extract temporal features. The next part is the similarity extractor network, with 4005 fully connected networks, which computes pairwise similarity between the 90 features extracted earlier. Finally, the classification network takes inputs from the outputs of the 4005 earlier networks and makes predictions. Apart from that, a recent study by Li et al.<sup>21</sup> uses a combination of 3D convolutions and long-short term memory (LSTMs) for AD classification. They use 3D CNN to extract features and then use the extracted features with an LSTM for making the classification. Somewhat similar to that, the proposed network has a feature extractor and a classifier part. However, the proposed network is relatively simpler as compared to the LSTM. Additionally, the architecture in this work uses time point volumes from the actual preprocessed fMRI data, without any special processing, as input to the network. This reduces both computational time and complexity.

Moreover, the results obtained in this and some of the previous studies show the possibility of using deep learning—fMRI synergy for effective diagnosis and classification of AD. A deep learning-based approaches can help improve diagnosis not only for AD but also for other neurological disorders. We can take advantage of such techniques and develop potential application of fMRI, mainly resting-state fMRI, in routine clinical practice. Unlike task fMRI scans, which have some clinical application as a presurgical procedure for brain tumors,<sup>29,30</sup> resting-state fMRI scans are easier to obtain and provide whole-brain functional information.<sup>31</sup> During a task fMRI, subject cooperation and participation are required for accurate analysis, while in a resting-state fMRI scan, the subject is not expected to participate in any task. This also reduces motion artifacts.<sup>32</sup> Functional scans also provide dynamic information about the brain that is not obtained by static anatomical scans. Taking advantage of deep learning, one can develop fMRI-based biomarkers for AD and other neurological disorders.<sup>33–35</sup>

In spite having slightly lower accuracy than some of the image-based classifiers (Table 2), the proposed deep learning functional classifier should be developed for clinical applications because of two main reasons. First, as discussed earlier, changes in the functional neuroimaging data appear before the changes in the anatomical structures which can allow for early detection. Second, the deep learning-based approach does not require complex and independent feature extraction stages as used by many of the other image-based classifiers. The network itself can extract useful features using the training data, making model simple to use. Another advantage of this study is that it can differentiate between MCI and AD, which are not easily differentiable

using just anatomical scans. Most of the time, by the time the symptoms of dementia are visible in the structural scan, it is already too late to reverse the damage. If we could detect dementia in the MCI stage by using fMRI classifiers, there is a good possibility to slow down the rate of progression to AD. As a secondary advantage, such a deep learning-based classifier has a potential to be used as an early detection tool for AD.

Just as with any other clinical machine learning algorithm, the main limitation of this algorithm is the accuracy and correctness of the training data. The true prediction labels for the training data were assigned by domain experts. In absence of gold standard biomarkers, it is sometimes difficult to distinguish between early stages of AD and later stages of MCI. Moreover, hard diagnostic decision boundaries between classes do not exist. As a result, one should be careful when applying such techniques for clinical purposes. A prediction error by the domain expert taken as ground truth would cause the network to learn incorrect features. Thus, correctness of training data becomes a major limitation. A second major limitation is the inability to generalize for unbalanced classes. If functional data with a rare and unique disorder is presented to the network, the network might not give accurate prediction. Though care is taken to prevent overfitting, most machine learning systems work well only for the kind of data that trained it. With extreme outliers, the network might tend to give inaccurate predictions and one must be aware of this tendency. Finally, the resting-state fMRI training data does not only contain differences related to AD and MCI. Various other factors, such as age, gender, physical and mental state, the emotion and thought process during the scan, physiology including respiration, and heart rate could introduce bias in the fMRI dataset. We used an age and gender-matched population from the same scanner and the same number of subjects per class to reduce some of the known bias.

## 5 Future Work

fMRI has long way to go before it can become a routine clinical practice. This study and similar studies indicate that fMRI has the potential to become a clinical diagnostic and early detection tool. This work is a step in that direction. As future work, we wish to train the network on an even larger dataset. With access to open source databases, researchers have access to a large dataset from different scan sites. If trained with a large number of data points from a variety of different locations, a CNN has higher chances of generalizing well to unanticipated data. Second, we will also study different kernel combinations in the same network and compare their performances. One main kernel combination to focus on would be the first temporal kernel. Currently, the temporal convolution is performed over five time points, which is equivalent to a sliding window of 5. Both larger and smaller kernels would have their own advantages and limitations. With a larger window, one might be able to capture better temporal similarity. The ROIs extracted from a larger window length might be comparable to resting-state functional networks such as the default mode network. The main limitation with a larger window is increased space and computation complexity. On the other hand, smaller kernels are computationally less expensive, but they do not capture enough temporal similarity and can be easily biased with noise and artifacts, which is quite common in fMRI studies. With well-studied network parameters and adequate amounts of data, we wish to investigate the possibility of developing fMRI-based biomarkers for AD and MCI.

## 6 Conclusion

The results obtained from deep learning-based classification show the possibility of using simple volumetric 3D CNN with 4D fMRI BOLD signal in detection and classification of AD. The unique architecture allows us to extract both spatial and temporal features from the entire 4D volume. With more training data, neuroimaging biomarkers can eventually be developed that could lead to early detection of AD. This development would also open new avenues for other clinical applications of fMRI.

## Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

## Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and Lister Hill National Center for Biomedical Communications (Contract No. HHSN276201800171P), and internal funding by Texas Tech University.

## Code, Data, and Materials Availability

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators 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](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

## References

1. S. Ogawa et al. "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Magn. Reson. Med.* **14**(1), 68–78 (1990).
2. P. A. Bandettini et al., "Spin-echo and gradient-echo EPI of human brain activation using BOLD contrast: a comparative study at 1.5 T," *NMR Biomed.* **7**(1–2), 12–20 (1994).
3. B. F. Biswal et al., "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magn. Reson. Med.* **34**(4), 537–541 (1995).
4. B. C. Dickerson, F. Agosta, and M. Filippi, "fMRI in neurodegenerative diseases: from scientific insights to clinical applications," *NeuroMethods* **119**, 699–739 (2016).
5. R. A. Poldrack, "The role of fMRI in cognitive neuroscience: where do we stand?" *Curr. Opin. Neurobiol.* **18**(2), 223–227 (2008).
6. R. M. Birn et al., "fMRI in the presence of task-correlated breathing variations," *Neuroimage* **47**(3), 1092–1104 (2009).
7. S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Future Technologies Conf.*, San Francisco, California, IEEE (2016).
8. S. G. Mueller et al., "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)," *Alzheimer's Dementia* **1**(1), 55–66 (2005).
9. R. A. Poldrack et al., "Toward open sharing of task-based fMRI data: the OpenfMRI project," *Front. Neuroinf.* **7**, 12 (2013).
10. H. S. Parmar et al., "Deep learning of volumetric 3D CNN for fMRI in Alzheimer's disease classification," *Proc. SPIE* **11317**, 113170C (2020).
11. World Health Organization, *Neurological Disorders: Public Health Challenges*, World Health Organization (2006).
12. J. Ashburner et al., *SPM12 Manual*, Wellcome Trust Centre for Neuroimaging, London (2014).
13. H. S. Parmar et al., "Automated signal drift and global fluctuation removal from 4D fMRI data based on principal component analysis as a major preprocessing step for fMRI data analysis," *Proc. SPIE* **10953**, 109531E (2019).
14. Y. Kazemi and S. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in *IEEE Conf. Comput. Intell. Bioinf. and Comput. Biol.*, IEEE, pp. 1–8 (2018).
15. C. D. Billones et al., "DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment," in *IEEE Region 10 Conf.*, IEEE, pp. 3724–3727 (2016).

16. R. Jain et al., "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognit. Syst. Res.* **57**, 147–159 (2019).
17. S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clin.* **21**, 101645 (2019).
18. E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *IEEE Int. Conf. Image Process.*, IEEE, pp. 126–130 (2016).
19. N. T. Duc et al., "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics* **18**(1), 71–86 (2020).
20. P. Forouzannezhad et al., "A Gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging," *J. Neurosci. Methods* **333**, 108544 (2020).
21. W. Li, X. Lin, and X. Chen, "Detecting Alzheimer's disease Based on 4D fMRI: an exploration under deep learning framework," *Neurocomputing* **388**, 280–287 (2020).
22. M. A. A. Binnewijzend et al., "Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment," *Neurobiol. Aging* **33**(9), 2018–2028 (2012).
23. S. A. R. B. Rombouts et al., "Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study," *Hum. Brain Mapp.* **26**(4), 231–239 (2005).
24. Y. Cai et al., "Exploiting spatial-temporal relationships for 3D pose estimation via graph convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2272–2281 (2019).
25. J. Liu et al., "Skeleton-based action recognition using spatio-temporal LSTM network with trust gates," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(12), 3007–3021 (2017).
26. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).
27. X. Li et al., "2-channel convolutional 3D deep neural network (2CC3D) for fMRI analysis: ASD classification and feature learning," in *IEEE 15th Int. Symp. Biomed. Imaging*, IEEE, pp. 1252–1255 (2018).
28. A. Riaz et al., "Deep fMRI: an end-to-end deep network for classification of fMRI data," in *IEEE 15th Int. Symp. Biomed. Imaging*, IEEE, pp. 1419–1422 (2018).
29. A. Håberg et al., "Preoperative blood oxygen level-dependent functional magnetic resonance imaging in patients with primary brain tumors: clinical application and outcome," *Neurosurgery* **54**(4), 902–915 (2004).
30. P. M. Matthews, G. D. Honey, and E. T. Bullmore, "Neuroimaging: applications of fMRI in translational medicine and clinical practice," *Nat. Rev. Neurosci.* **7**(9), 732 (2006).
31. M. D. Fox and M. Greicius, "Clinical applications of resting state functional connectivity," *Frontiers Syst. Neurosci.* **4**, 19 (2010).
32. Z. Jiang et al., "Impaired fMRI activation in patients with primary brain tumors," *Neuroimage* **52**(2), 538–548 (2010).
33. F. de Vos et al., "A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease," *Neuroimage* **167**, 62–72 (2018).
34. P. Vemuri, D. T. Jones, and C. R. Jack, "Resting state functional MRI in Alzheimer's disease," *Alzheimer's Res. Ther.* **4**(1), 2 (2012).
35. K. Wang et al., "Discriminative analysis of early Alzheimer's disease based on two intrinsically anti-correlated networks with resting-state fMRI," *Lect. Notes Comput. Sci.* **4191**, 340–347 (2006).

**Harshit Parmar** received his BE degree in electronics engineering from Maharaja Sayajirao University of Baroda, India in 2016, and an MS degree in electrical engineering from Texas Tech University (TTU), Lubbock, Texas in 2018. He is currently pursuing his doctor of philosophy (PhD) in electrical engineering at TTU. He is working at the Biomedical Integrated Devices and Systems laboratory at TTU. His research interests include signal processing, pattern recognition, image processing, computer vision, and machine learning.

**Brian Nutter** received his BSEE and PhD degrees from Texas Tech University, Lubbock, Texas, in 1987 and 1990, respectively. He is an associate professor and associate chairman in the Department of Electrical and Computer Engineering, Texas Tech University. His interests include telecommunications, networks, biomedical signal and image processing, rapid prototyping, and real-time embedded systems.

**Rodney Long** leads a development group in creating applications for image-based biomedical information collection and dissemination. Currently, he is working in collaboration with the National Cancer Institute to develop a suite of tools for uterine cervix cancer databases. His research interests are in telecommunications, systems biology, image processing, and scientific/biomedical databases. He has an MA in applied mathematics from the University of Maryland.

**Sameer Antani** is currently (Acting) branch chief for the Communications Engineering Branch and the Computer Science Branch in the Lister Hill National Center for Biomedical Communications at the National Library of Medicine. He is a versatile lead researcher advancing the role of computational sciences and automated decision making in biomedical research, education, and clinical care. His research interests include topics in medical imaging and informatics, machine learning, data science, artificial intelligence, and global health. He applies his expertise in machine learning, biomedical image informatics, automatic medical image interpretation, data science, information retrieval, computer vision, and related topics in computer science and engineering technology.

**Sunanda Mitra** received her BSc degree in 1955 and MSc degree in 1957, both in physics, from Calcutta University, Calcutta, India, and her DSc (Doctors der Naturwissenschaften) degree in physics from Marburg University, Marburg, Germany, in 1966. She has served as the director of Computer Vision and Image Analysis Laboratory, Department of Electrical and Computer Engineering, Texas Tech University (TTU), Lubbock, Texas, since 1988. Prior to joining TTU in 1984, she worked as a research scientist at TTU's School of Medicine (1969 to 1983) and as a Visiting Faculty (1983 to 1984) at the Mount Sinai School of Medicine, New York. She also held a Faculty position (1960 to 1964) at Lady Brabourne College, Calcutta, India. She served on the Board of Scientific Counselors of the National Library of Medicine at the National Institutes of Health (USA) from 1997 to 2000. She holds the P.W. Horn Professorship at TTU and has published over 160 scientific articles including archival journal papers, invited papers, and book chapters. Her specialization includes medical image segmentation and analysis, data compression, 3-D modeling from stereo vision and pattern recognition. She has chaired the Technical Committee of Computational Medicine of the IEEE Computer Society.