



# Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data

Taeho Jo<sup>1,2,3\*</sup>, Kwangsik Nho<sup>1,2,3</sup> and Andrew J. Saykin<sup>1,2,3</sup>

<sup>1</sup> Department of Radiology and Imaging Sciences, Center for Neuroimaging, Indiana University School of Medicine, Indianapolis, IN, United States, <sup>2</sup> Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, United States, <sup>3</sup> Indiana University Network Science Institute, Bloomington, IN, United States

Deep learning, a state-of-the-art machine learning approach, has shown outstanding performance over traditional machine learning in identifying intricate structures in complex high-dimensional data, especially in the domain of computer vision. The application of deep learning to early detection and automated classification of Alzheimer's disease (AD) has recently gained considerable attention, as rapid progress in neuroimaging techniques has generated large-scale multimodal neuroimaging data. A systematic review of publications using deep learning approaches and neuroimaging data for diagnostic classification of AD was performed. A PubMed and Google Scholar search was used to identify deep learning papers on AD published between January 2013 and July 2018. These papers were reviewed, evaluated, and classified by algorithm and neuroimaging type, and the findings were summarized. Of 16 studies meeting full inclusion criteria, 4 used a combination of deep learning and traditional machine learning approaches, and 12 used only deep learning approaches. The combination of traditional machine learning for classification and stacked auto-encoder (SAE) for feature selection produced accuracies of up to 98.8% for AD classification and 83.7% for prediction of conversion from mild cognitive impairment (MCI), a prodromal stage of AD, to AD. Deep learning approaches, such as convolutional neural network (CNN) or recurrent neural network (RNN), that use neuroimaging data without pre-processing for feature selection have yielded accuracies of up to 96.0% for AD classification and 84.2% for MCI conversion prediction. The best classification performance was obtained when multimodal neuroimaging and fluid biomarkers were combined. Deep learning approaches continue to improve in performance and appear to hold promise for diagnostic classification of AD using multimodal neuroimaging data. AD research that uses deep learning is still evolving, improving performance by incorporating additional hybrid data types, such as—omics data, increasing transparency with explainable approaches that add knowledge of specific disease-related features and mechanisms.

**Keywords:** artificial intelligence, machine learning, deep learning, classification, Alzheimer's disease, neuroimaging, magnetic resonance imaging, positron emission tomography

## OPEN ACCESS

### Edited by:

James H. Cole,  
King's College London,  
United Kingdom

### Reviewed by:

Donghuan Lu,  
Simon Fraser University, Canada  
Zheng Wang,  
University of Miami, United States

### \*Correspondence:

Taeho Jo  
tjo@iu.edu

**Received:** 01 March 2019

**Accepted:** 02 August 2019

**Published:** 20 August 2019

### Citation:

Jo T, Nho K and Saykin AJ (2019)  
Deep Learning in Alzheimer's Disease:  
Diagnostic Classification and  
Prognostic Prediction Using  
Neuroimaging Data.  
*Front. Aging Neurosci.* 11:220.  
doi: 10.3389/fnagi.2019.00220

## INTRODUCTION

Alzheimer's disease (AD), the most common form of dementia, is a major challenge for healthcare in the twenty-first century. An estimated 5.5 million people aged 65 and older are living with AD, and AD is the sixth-leading cause of death in the United States. The global cost of managing AD, including medical, social welfare, and salary loss to the patients' families, was \$277 billion in 2018 in the United States, heavily impacting the overall economy and stressing the U.S. health care system (Alzheimer's Association, 2018). AD is an irreversible, progressive brain disorder marked by a decline in cognitive functioning with no validated disease modifying treatment (De strooper and Karran, 2016). Thus, a great deal of effort has been made to develop strategies for early detection, especially at pre-symptomatic stages in order to slow or prevent disease progression (Galvin, 2017; Schelke et al., 2018). In particular, advanced neuroimaging techniques, such as magnetic resonance imaging (MRI) and positron emission tomography (PET), have been developed and used to identify AD-related structural and molecular biomarkers (Veitch et al., 2019). Rapid progress in neuroimaging techniques has made it challenging to integrate large-scale, high dimensional multimodal neuroimaging data. Therefore, interest has grown rapidly in computer-aided machine learning approaches for integrative analysis. Well-known pattern analysis methods, such as linear discriminant analysis (LDA), linear program boosting method (LPBM), logistic regression (LR), support vector machine (SVM), and support vector machine-recursive feature elimination (SVM-RFE), have been used and hold promise for early detection of AD and the prediction of AD progression (Rathore et al., 2017).

In order to apply such machine learning algorithms, appropriate architectural design or pre-processing steps must be predefined (Lu and Weng, 2007). Classification studies using machine learning generally require four steps: feature extraction, feature selection, dimensionality reduction, and feature-based classification algorithm selection. These procedures require specialized knowledge and multiple stages of optimization, which may be time-consuming. Reproducibility of these approaches has been an issue (Samper-Gonzalez et al., 2018). For example, in the feature selection process, AD-related features are chosen from various neuroimaging modalities to derive more informative combinatorial measures, which may include mean subcortical volumes, gray matter densities, cortical thickness, brain glucose metabolism, and cerebral amyloid- $\beta$  accumulation in regions of interest (ROIs), such as the hippocampus (Riedel et al., 2018).

In order to overcome these difficulties, deep learning, an emerging area of machine learning research that uses raw neuroimaging data to generate features through "on-the-fly" learning, is attracting considerable attention in the field of large-scale, high-dimensional medical imaging analysis (Plis et al., 2014). Deep learning methods, such as convolutional neural networks (CNN), have been shown to outperform existing machine learning methods (Lecun et al., 2015).

We systematically reviewed publications where deep learning approaches and neuroimaging data were used for the early detection of AD and the prediction of AD progression. A

PubMed and Google Scholar search was used to identify deep learning papers on AD published between January 2013 and July 2018. The papers were reviewed and evaluated, classified by algorithms and neuroimaging types, and the findings were summarized. In addition, we discuss challenges and implications for the application of deep learning to AD research.

## DEEP LEARNING METHODS

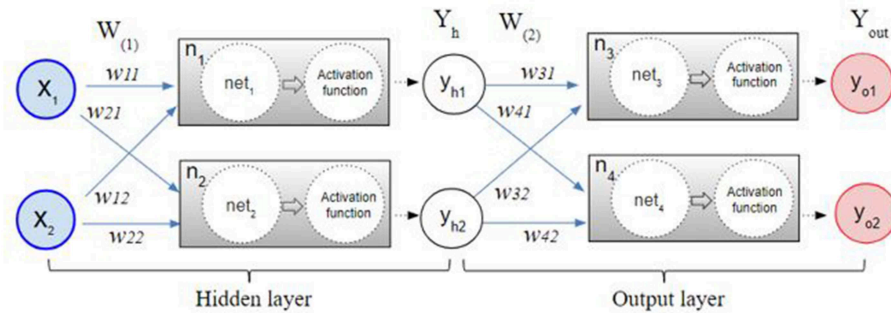
Deep learning is a subset of machine learning (Lecun et al., 2015), meaning that it learns features through a hierarchical learning process (Bengio, 2009). Deep learning methods for classification or prediction have been applied in various fields, including computer vision (Ciregan et al., 2012; Krizhevsky et al., 2012; Farabet et al., 2013) and natural language processing (Hinton et al., 2012; Mikolov et al., 2013), both of which demonstrate breakthroughs in performance (Boureau et al., 2010; Russakovsky et al., 2015). Because deep learning methods have been reviewed extensively in recent years (Bengio, 2013; Bengio et al., 2013; Schmidhuber, 2015), we focus here on basic concepts of Artificial Neural Networks (ANN) that underlie deep learning (Hinton and Salakhutdinov, 2006). We also discuss architectural layouts of deep learning that have been applied to the task of AD classification and prognostic prediction. ANN is a network of interconnected processing units called artificial neurons that were modeled (McCulloch and Pitts, 1943) and developed with the concept of Perceptron (Rosenblatt, 1957, 1958), Group Method of Data Handling (GMDH) (Ivakhnenko and Lapa, 1965; Ivakhnenko, 1968, 1971), and the Neocognitron (Fukushima, 1979, 1980). Efficient error functions and gradient computing methods were discussed in these seminal publications, spurred by the demonstrated limitation of the single layer perceptron, which can learn only linearly separable patterns (Minsky and Papert, 1969). Further, the back-propagation procedure, which uses gradient descent, was developed and applied to minimize the error function (Werbos, 1982, 2006; Rumelhart et al., 1986; Lecun et al., 1988).

### Gradient Computation

The back-propagation procedure is used to calculate the error between the network output and the expected output. The back propagation calculates the gap repeatedly, changing weights and stopping the calculation when the gap is no longer updated (Rumelhart et al., 1986; Bishop, 1995; Ripley and Hjort, 1996; Schalkoff, 1997). **Figure 1** illustrates the process of the neural network made by multilayer perceptron. After the initial error value is calculated from the given random weight by the least squares method, the weights are updated until the differential value becomes 0. For example, the  $w_{31}$  in **Figure 1** is updated by the following formula:

$$w_{31}(t+1) = w_{31}t - \frac{\partial \text{Error} Y_{out}}{\partial w_{31}}$$

$$\text{Error} Y_{out} = \frac{1}{2} (y_{t1} - y_{o1})^2 + \frac{1}{2} (y_{t2} - y_{o2})^2$$



**FIGURE 1 |** The multilayer perceptron procedure. After the initial error value is calculated from the given random weight by the least squares method, the weights are updated by a back-propagation algorithm until the differential value becomes 0.

The  $Error_{Y_{out}}$  is the sum of error  $y_{o1}$  and error  $y_{o2}$ .  $y_{t1}$ ,  $y_{t2}$  are constants that are known through the given data. The partial derivative of  $Error_{Y_{out}}$  with respect to  $w_{31}$  can be calculated by the chain rule as follows.

$$\frac{\partial Error_{Y_{out}}}{\partial w_{31}} = \frac{\partial Error_{Y_{out}}}{\partial y_{o1}} \cdot \frac{\partial y_{o1}}{\partial net_3} \cdot \frac{\partial net_3}{\partial w_{31}}$$

Likewise,  $w_{11}$  in the hidden layer is updated by the chain rule as follows.

$$\frac{\partial Error_{Y_{out}}}{\partial w_{11}} = \frac{\partial Error_{Y_{out}}}{\partial y_{h1}} \cdot \frac{\partial y_{h1}}{\partial net_1 y} \cdot \frac{\partial net_1}{\partial w_{11}}$$

Detailed calculation of the weights in the backpropagation is described in **Supplement 1**.

### Modern Practical Deep Neural Networks

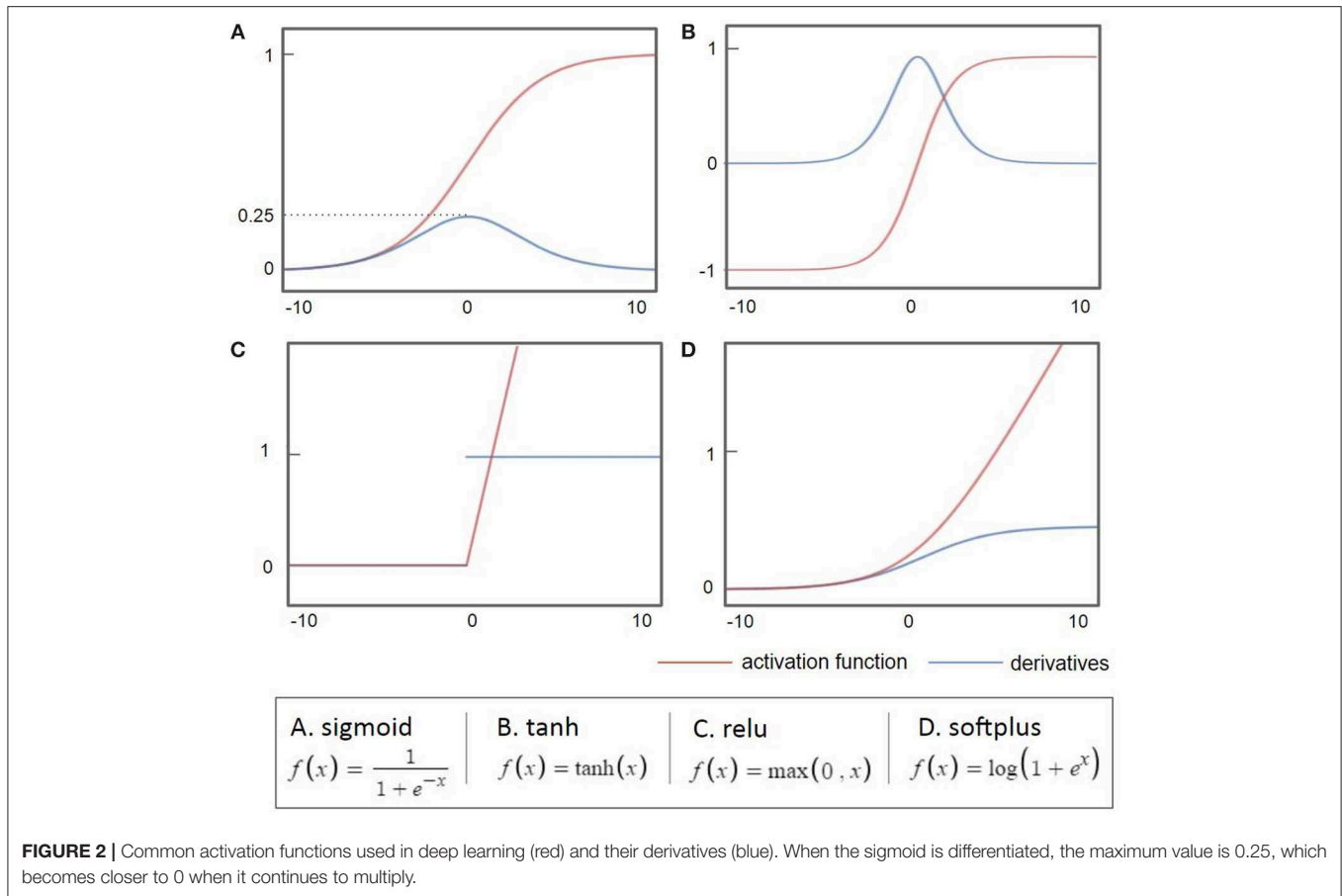
As the back-propagation uses a gradient descent method to calculate the weights of each layer going backwards from the output layer, a vanishing gradient problem occurs as the layer is stacked, where the differential value becomes 0 before finding the optimum value. As shown in **Figure 2A**, when the sigmoid is differentiated, the maximum value is 0.25, which becomes closer to 0 when it continues to multiply. This is called a vanishing gradient issue, a major obstacle of the deep neural network. Considerable research has addressed the challenge of the vanishing gradient (Goodfellow et al., 2016). One of the accomplishments of such an effort is to replace the sigmoid function, an activation function, with several other functions, such as the hyperbolic tangent function, ReLu, and Softplus (Nair and Hinton, 2010; Glorot et al., 2011). The hyperbolic tangent (tanh, **Figure 2B**) function expands the range of derivative values of the sigmoid. The ReLu function (**Figure 2C**), the most used activation function, replaces a value with 0 when the value is  $<0$  and uses the value if the value is  $>0$ . As the derivative becomes 1, when the value is larger than 0, it becomes possible to adjust the weights without disappearing up to the first layer through the stacked hidden layers. This simple method allows building multiple layers and accelerates the development of deep learning.

The Softplus function (**Figure 2D**) replaces the ReLu function with a gradual descent method when ReLu becomes zero.

While a gradient descent method is used to calculate the weights accurately, it usually requires a large amount of computation time because all of the data needs to be differentiated at each update. Thus, in addition to the activation function, advanced gradient descent methods have been developed to solve speed and accuracy issues. For example, Stochastic Gradient Descent (SGD) uses a subset that is randomly extracted from the entire data for faster and more frequent updates (Bottou, 2010), and it has been extended to Momentum SGD (Sutskever et al., 2013). Currently, one of the most popular gradient descent method is Adaptive Moment Estimation (Adam). Detailed calculation of the optimization methods is described in **Supplement 2**.

### Architectures of Deep Learning

Overfitting has also played a major role in the history of deep learning (Schmidhuber, 2015), with efforts being made to solve it at the architectural level. The Restricted Boltzmann Machine (RBM) was one of the first models developed to overcome the overfitting problem (Hinton and Salakhutdinov, 2006). Stacking the RBMs resulted in building deeper structures known as the Deep Boltzmann Machine (DBM) (Salakhutdinov and Larochelle, 2010). The Deep Belief Network (DBN) is a supervised learning method used to connect unsupervised features by extracting data from each stacked layer (Hinton et al., 2006). DBN was found to have a superior performance to other models and is one of the reasons that deep learning has gained popularity (Bengio, 2009). While DBN solves the overfitting problem by reducing the weight initialization using RBM, CNN efficiently reduces the number of model parameters by inserting convolution and pooling layers that lead to a reduction in complexity. Because of its effectiveness, when given enough data, CNN is widely used in the field of visual recognition. **Figure 3** shows the structures of RBM, DBM, DBN, CNN, Auto-Encoders (AE), sparse AE, and stacked AE, respectively. Auto-Encoders (AE) are an unsupervised learning method that make the output value approximate to the input value by using the back-propagation and SGD (Hinton and Zemel, 1994). AE



engages the dimensional reduction, but it is difficult to train due to the vanishing gradient issue. Sparse AE has solved this issue by allowing for only a small number of the hidden units (Makhzani and Frey, 2015). Stacked AE stacks sparse AE like DBN.

DNN, RBM, DBM, DBN, AE, Sparse AE, and Stacked AE are deep learning methods that have been used for Alzheimer's disease diagnostic classification to date (see **Table 1** for the definition of acronyms). Each approach has been developed to classify AD patients from cognitively normal controls (CN) or mild cognitive impairment (MCI), which is the prodromal stage of AD. Each approach is used to predict the conversion of MCI to AD using multi-modal neuroimaging data. In this paper, when deep learning is used together with traditional machine learning methods, i.e., SVM as a classifier, it is referred to as a "hybrid method."

## MATERIALS AND METHODS

We conducted a systematic review on previous studies that used deep learning approaches for diagnostic classification of AD with multimodal neuroimaging data. The search strategy is outlined in detail using the PRISMA flow diagram (Moher et al., 2009) in **Figure 4**.

## Identification

From a total of 389 hits on Google scholar and PubMed search, 16 articles were included in the systematic review.

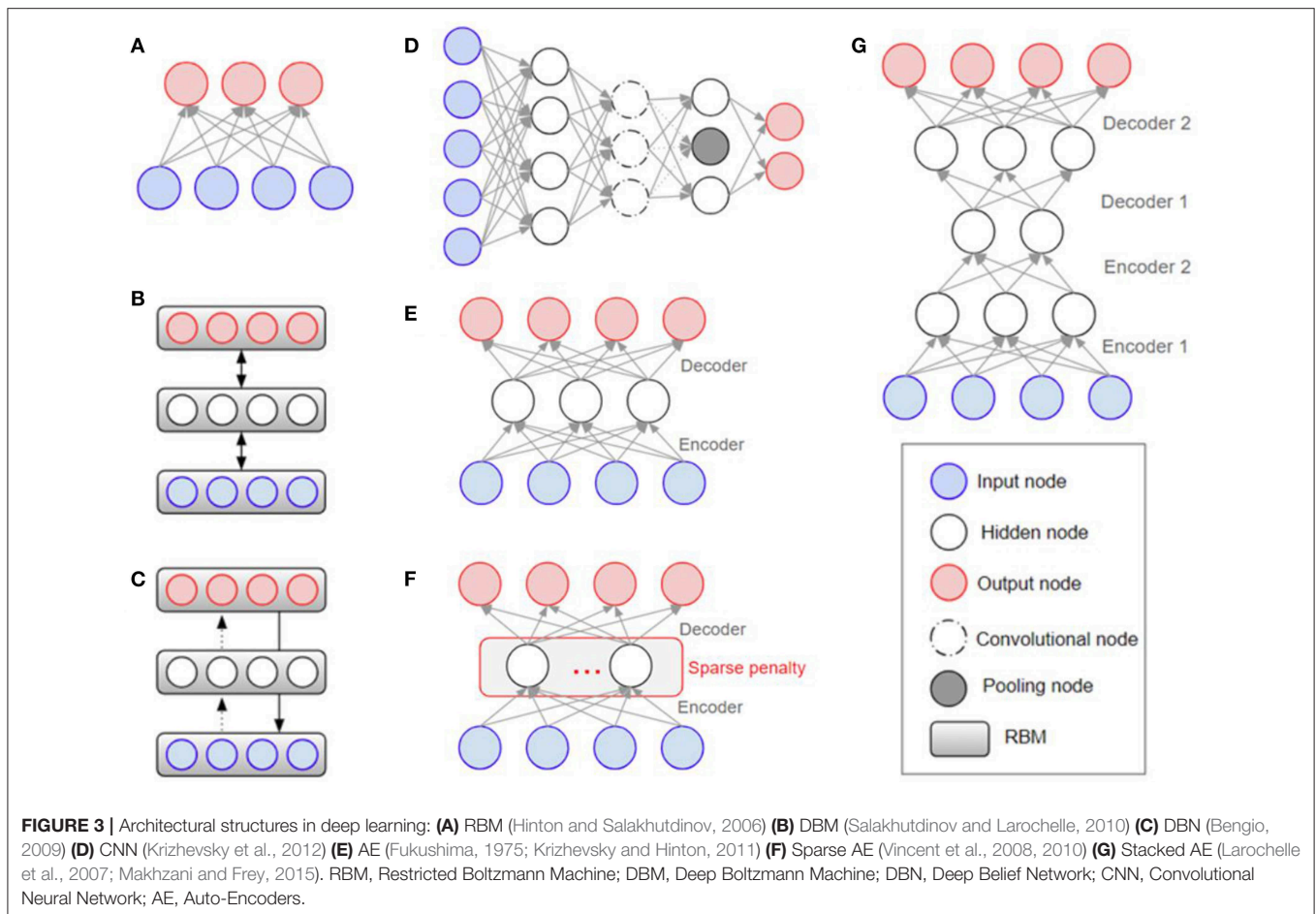
Google Scholar: We searched using the following key words and yielded 358 results ("Alzheimer disease" OR "Alzheimer's disease"), ("deep learning" OR "deep neural network" OR "CNN" OR "CNN" OR "Autoencoder" OR "DBN" OR "RBM"), ("Neuroimaging" OR "MRI" OR "multimodal").

PubMed: The keywords used in the Google Scholar search were reused for the search in PubMed, and yielded 31 search results ("Alzheimer disease" OR "Alzheimer's disease") AND ("deep learning" OR "deep neural network" OR "CNN" OR "recurrent neural network" OR "Auto-Encoder" OR "Auto Encoder" OR "RBM" OR "DBN" OR "Generative Adversarial Network" OR "Reinforcement Learning" OR "Long Short Term Memory" OR "Gated Recurrent Units") AND ("Neuroimaging" OR "MRI" OR "multimodal").

Among the 389 relevant records, 25 overlapping records were removed.

## Screening Based on Article Type

We first excluded 38 survey papers, 22 theses, 19 Preprint, 34 book chapters, 20 conference abstract, 13 none English papers, 5 citations, and 10 patents. We also excluded 11 papers of which



**TABLE 1 |** Definition of acronyms.

Acronym	Description	Acronym	Description
ANN	Artificial neural network	CNN	Convolutional neural network
DNN	Deep neural network	RNN	Recurrent neural network
RBM	Restricted Boltzmann machine	GAN	Generative adversarial networks
DBM	Deep Boltzmann machine	SGD	Stochastic gradient descent
DBN	Deep belief network	SVM	Support vector machine
AE	Auto-encoders	ROI	Regions of interest
SAE	Stacked auto-encoder	HMM	Hidden markov model

the full text was not accessible. The remaining 192 articles were downloaded for review.

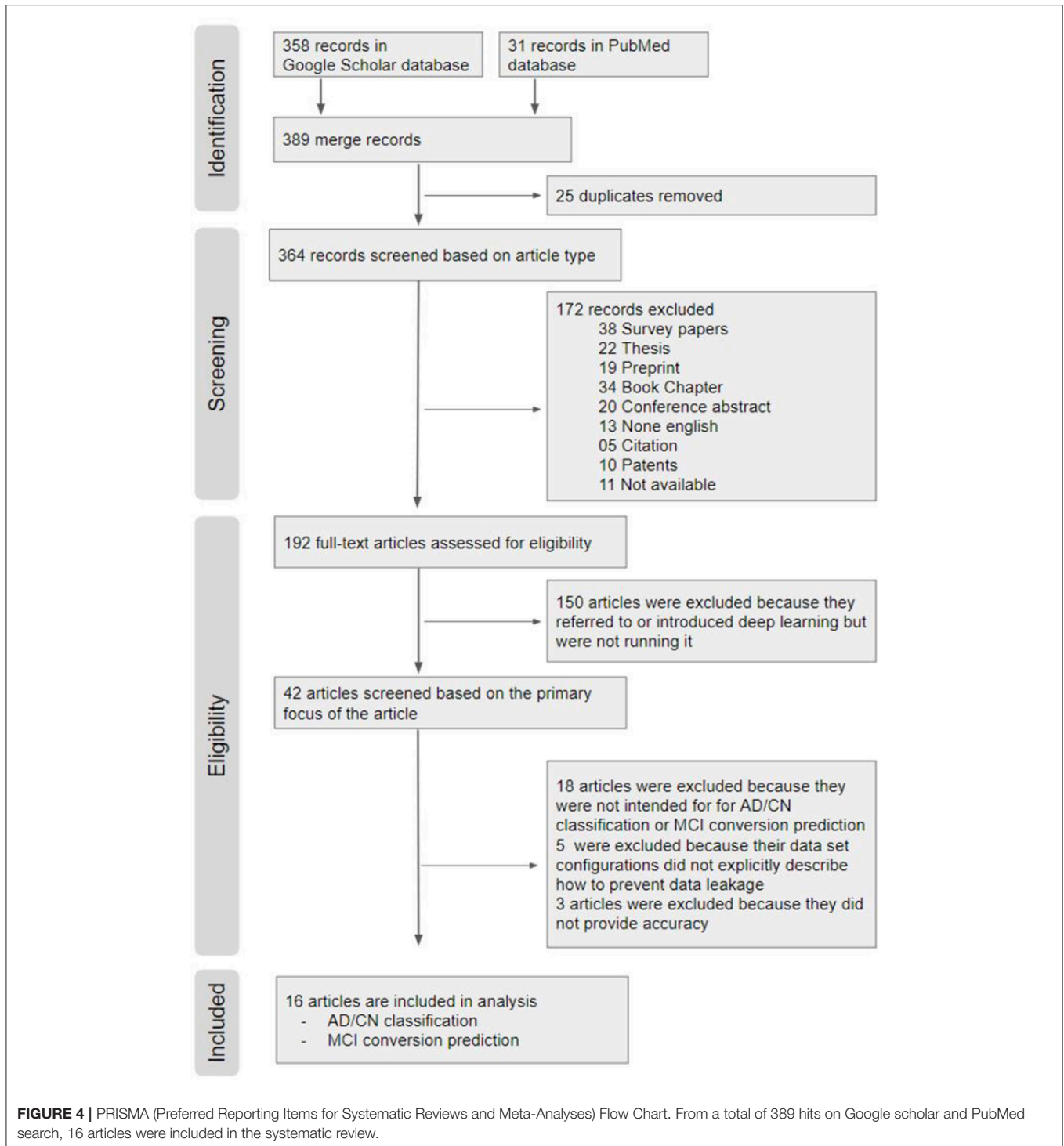
### Eligibility Screening

Out of the 192 publications retrieved, 150 articles were excluded because the authors only introduced or mentioned deep learning but did not use it. Out of the 42 remaining publications, (1) 18 articles were excluded because they did not perform deep learning approaches for AD classification

and/or prediction of MCI to AD conversion; (2) 5 articles were excluded because their neuroimaging data were not explicitly described; and (3) 3 articles were excluded because performance results were not provided. The remaining 16 papers were included in this review for AD classification and/or prediction of MCI to AD conversion. All of the final selected and compared papers used ADNI data in common.

### RESULTS

From the 16 papers included in this review, **Table 2** provides the top results of diagnostic classification and/or prediction of MCI to AD conversion. We compared only binary classification results. Accuracy is a measure used consistently in the 16 publications. However, it is only one metric of the performance characteristics of an algorithm. The group composition, sample sizes, and number of scans analyzed are also noted together because accuracy is sensitive to unbalanced distributions. **Table S1** shows the full results sorted according to the performance accuracy as well as the number of subjects, the deep learning approach, and the neuroimaging type used in each paper.



## Deep Learning for Feature Selection From Neuroimaging Data

Multimodal neuroimaging data have been used to identify structural and molecular/functional biomarkers for AD. It has been shown that volumes or cortical thicknesses in pre-selected AD-specific regions, such as the hippocampus and entorhinal

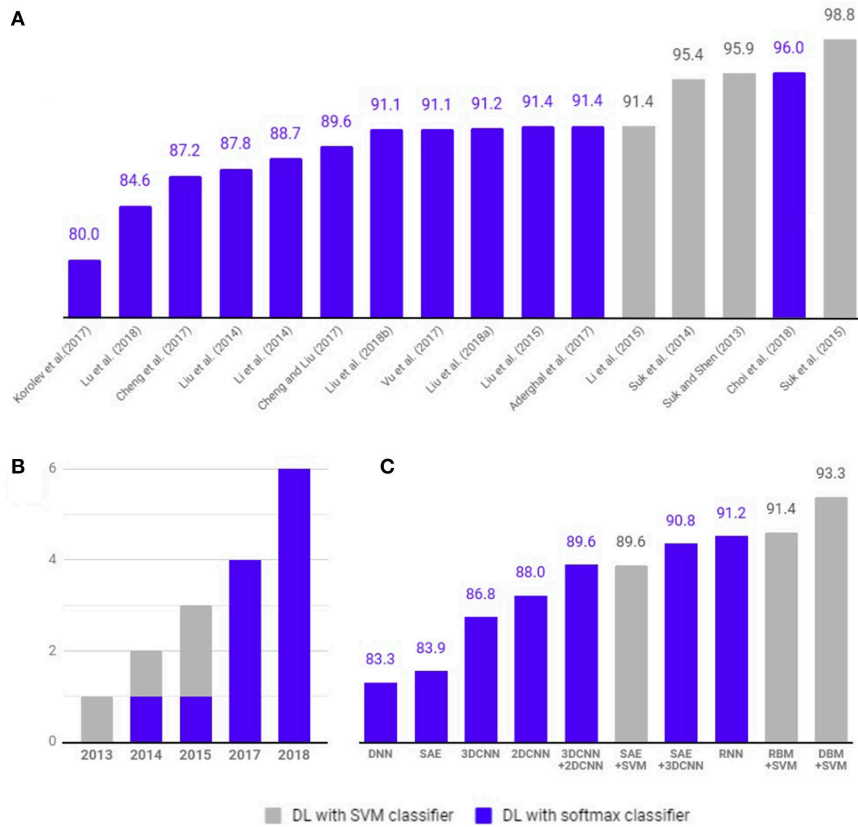
cortex, could be used as features to enhance the classification accuracy in machine learning. Deep learning approaches have been used to select features from neuroimaging data.

As shown in **Figure 5**, 4 studies have used hybrid methods that combine deep learning for feature selection from neuroimaging data and traditional machine learning, such as the SVM as a

**TABLE 2 |** Summary of 16 previous studies to systematically be reviewed.

References	Modality	Data processing/training	Classifier	AD:NC acc.	SEN	SPE	cMCI:ncMCI acc.	SEN	SPE	AD	cMCI	ncMCI	NC	Total
Suk and Shen (2013)	MRI, PET, CSF	SAE	SVM	<b>95.9</b>			<b>75.8</b>			51	43	56	52	202
Liu et al. (2014)	MRI, PET	SAE + NN	Softmax	<b>87.76</b>	88.57	87.22	76.92 (MCI:NC)	74.29	78.13	65	67	102	77	311
Suk et al. (2014)	MRI, PET	DBM	SVM	<b>95.35</b>	94.65	95.22	<b>75.92</b> 86.75 (MCI:NC)	48.04 95.37	95.23 65.87	93	76	128	101	398
Li et al. (2014)	MRI, PET	3D CNN	Logistic regression	<b>92.87</b>			76.21 (MCI:NC)			198	167	236	229	830
Li et al. (2015)	MRI, PET, CSF	RBM + drop out	SVM	<b>91.4</b>			<b>57.4</b> 76.21 (MCI:NC)			51	43	56	52	202
Suk et al. (2015)	MRI, PET, CSF	SAE + sparse learning	SVM	<b>98.8</b>			<b>83.3</b> 90.7 (MCI:NC)			51	43	56	52	202
Liu et al. (2015)	MRI, PET	SAE with zero-masking	Softmax	<b>91.4</b>	92.32	90.42	82.1 (MCI:NC)	60.0	92.32	77	67	102	85	331
Cheng et al. (2017)	MRI	3D CNN	Softmax	<b>87.15</b>	86.36	85.93				199			229	428
Cheng and Liu (2017)	MRI, PET	3D CNN + 2D CNN	Softmax	<b>89.64</b>	87.10	92.00				93			100	193
Aderghal et al. (2017)	MRI	2D CNN	Softmax	<b>91.41</b>	93.75	89.06	65.62 (MCI:NC)	66.25	65.0	188	399 (MCI)		228	815
Korolev et al. (2017)	MRI	3D CNN	Softmax	<b>80</b>	87 (AUC)		61 (IMCI:NC) 56 (IMCI:NC)	65 (AUC) 58 (AUC)		50	43 (IMCI)	77 (eMCI)	61	111
Vu et al. (2017)	MRI, PET	SAE + 3D CNN	Softmax	<b>91.14</b>						145			172	317
Liu et al. (2018a)	PET	RNN	Softmax	<b>91.2</b>	91.4	91.0	78.9 (MCI:NC)	78.01	80.0	93	146 (MCI)		100	339
Liu et al. (2018b)	MRI	Landmark detection + 3D CNN	Softmax	<b>91.09</b>	88.05	93.50	<b>76.9</b>	42.11	82.43	159	38	239	200	636
Lu et al. (2018)	MRI, PET	DNN + NN	Softmax	<b>84.6</b>	80.2	91.8	<b>82.93</b>	79.69	83.84	238	217	409	360	1224
Choi and Jin (2018)	PET	3D CNN	Softmax	<b>96</b>	93.5	97.8	<b>84.2</b>	81.0	87.0	139	79	92	182	492

SEN = TP/(TP + FN), SPE = TN/(TN + FP). TP, true positive; TN, true negative; FP, false positive; FN, false negative. All data on this table were from ADNI.



**FIGURE 5 |** Comparison of diagnostic classification accuracy of pure deep learning and hybrid approach. Four studies (gray) have used hybrid methods that combine deep learning for feature selection from neuroimaging data and traditional machine learning, such as the SVM as a classifier. Twelve studies (blue) have used deep learning method with softmax classifier for diagnostic classification and/or prediction of MCI to AD conversion. **(A)** Accuracy comparison between articles. **(B)** Number of studies published per year. **(C)** Average classification accuracy of each methods.

classifier. Suk and Shen (2013) used a stacked auto-encoder (SAE) to construct an augmented feature vector by concatenating the original features with outputs of the top hidden layer of the representative SAEs. Then, they used a multi-kernel SVM for classification to show 95.9% accuracy for AD/CN classification and 75.8% prediction accuracy of MCI to AD conversion. These methods successfully tuned the input data for the SVM classifier. However, SAE as a classifier (Suk et al., 2015) yielded 89.9% accuracy for AD/CN classification and 60.2% accuracy for prediction of MCI to AD conversion. Later Suk et al. (2015) extended the work to develop a two-step learning scheme: greedy layer-wise pre-training and fine-tuning in deep learning. The same authors further extended their work to use the DBM to find latent hierarchical feature representations by combining heterogeneous modalities during the feature representation learning (Suk et al., 2014). They obtained 95.35% accuracy for AD/CN classification and 74.58% prediction accuracy of MCI to AD conversion. In addition, the authors initialized SAE parameters with target-unrelated samples and tuned the optimal parameters with target-related samples to have 98.8% accuracy for AD/CN classification and 83.7% accuracy for prediction of MCI to AD conversion (Suk et al., 2015). Li et al. (2015) used

the RBM with a dropout technique to reduce overfitting in deep learning and SVM as a classifier, which produced 91.4% accuracy for AD/CN classification and 57.4% prediction accuracy of MCI to AD conversion.

### Deep Learning for Diagnostic Classification and Prognostic Prediction

To select optimal features from multimodal neuroimaging data for diagnostic classification, we usually need several pre-processing steps, such as neuroimaging registration and feature extraction, which greatly affect the classification performance. However, deep learning approaches have been applied to AD diagnostic classification using original neuroimaging data without any feature selection procedures.

As shown in **Figure 5**, 12 studies have used only deep learning for diagnostic classification and/or prediction of MCI to AD conversion. Liu et al. (2014) used stacked sparse auto-encoders (SAEs) and a softmax regression layer and showed 87.8% accuracy for AD/CN classification. Liu et al. (2015) used SAE and a softmax logistic regressor as well as a zero-mask strategy for data fusion to extract complementary information from multimodal neuroimaging data (Ngiam et al., 2011), where

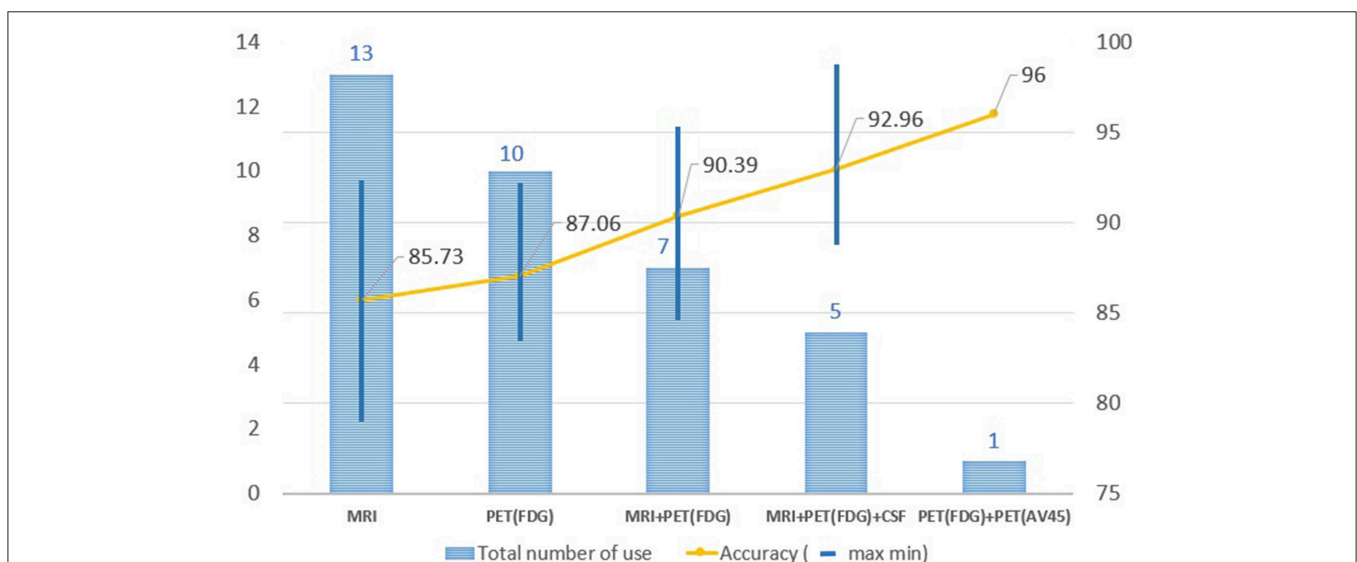


one of the modalities is randomly hidden by replacing the input values with zero to converge different types of image data for SAE. Here, the deep learning algorithm improved accuracy for AD/CN classification by 91.4%. Recently, Lu et al. (2018) used SAE for pre-training and DNN in the last step, which achieved an AD/CN classification accuracy of 84.6% and an MCI conversion prediction accuracy of 82.93%. CNN, which has shown remarkable performance in the field of image recognition, has also been used for the diagnostic classification of AD with multimodal neuroimaging data. Cheng et al. (2017) used image patches to transform the local images into high-level features from the original MRI images for the 3D-CNN and yielded 87.2% accuracy for AD/CN classification. They improved the accuracy to 89.6% by running two 3D-CNNs on neuroimage patches extracted from MRI and PET separately and by combining their results to run 2D CNN (Cheng and Liu, 2017). Korolev et al. (2017) applied two different 3D CNN approaches [plain (VoxCNN) and residual neural networks (ResNet)] and reported 80% accuracy for AD/CN classification, which was the first study that the manual feature extraction step was unnecessary. Aderghal et al. (2017) captured 2D slices from the hippocampal region in the axial, sagittal, and coronal directions and applied 2D CNN to show 85.9% accuracy for AD/CN classification. Liu et al. (2018b) selected discriminative patches from MR images based on AD-related anatomical landmarks identified by a data-driven learning approach and ran 3D CNN on them. This approach used three independent data sets (ADNI-1 as training, ADNI-2 and MIRIAD as testing) to yield relatively high accuracies of 91.09 and 92.75% for AD/CN classification from ADNI-2 and MIRIAD, respectively, and an MCI conversion prediction accuracy of 76.9% from ADNI-2. Li et al. (2014) trained 3D CNN models on subjects with both MRI and PET scans to encode the non-linear relationship between MRI and PET images and then

used the trained network to estimate the PET patterns for subjects with only MRI data. This study obtained an AD/CN classification accuracy of 92.87% and an MCI conversion prediction accuracy of 72.44%. Vu et al. (2017) applied SAE and 3D CNN to subjects with MRI and FDG PET scans to yield an AD/CN classification accuracy of 91.1%. Liu et al. (2018a) decomposed 3D PET images into a sequence of 2D slices and used a combination of 2D CNN and RNNs to learn the intra-slice and inter-slice features for classification, respectively. The approach yielded AD/CN classification accuracy of 91.2%. If the data is imbalanced, the chance of misdiagnosis increases and sensitivity decreases. For example, in Suk et al. (2014) there were 76 cMCI and 128 nMCI subjects and the obtained sensitivity of 48.04% was low. Similarly, Liu et al. (2018b) included 38 cMCI and 239 nMCI subjects and had a low sensitivity of 42.11%. Recently Choi and Jin (2018) reported the first use of 3D CNN models to multimodal PET images [FDG PET and [18F]florbetapir PET] and obtained 96.0% accuracy for AD/CN classification and 84.2% accuracy for the prediction of MCI to AD conversion.

## Performance Comparison by Types of Neuroimaging Techniques

In order to improve the performance for AD/CN classification and for the prediction of MCI to AD conversion, multimodal neuroimaging data such as MRI and PET have commonly been used in deep learning: MRI for brain structural atrophy, amyloid PET for brain amyloid- $\beta$  accumulation, and FDG-PET for brain glucose metabolism. MRI scans were used in 13 studies, FDG-PET scans in 10, both MRI and FDG-PET scans in 12, and both amyloid PET and FDG-PET scans in 1. The performance in AD/CN classification and/or prediction of MCI to AD conversion yielded better results in PET data compared to MRI. Two or more multimodal neuroimaging data types produced



**FIGURE 6 |** Changes in accuracy by types of image resource. MRI scans were used in 13 studies, FDG-PET scans in 10, both MRI and FDG-PET scans in 12, and both amyloid PET and FDG-PET scans in 1. The performance in AD/CN classification yielded better results in PET data compared to MRI. Two or more multimodal neuroimaging data types produced higher accuracies than a single neuroimaging technique.

higher accuracies than a single neuroimaging technique. **Figure 6** shows the results of the performance comparison by types of neuroimaging techniques.

## Performance Comparison by Deep Learning Algorithms

Deep learning approaches require massive amounts of data to achieve the desired levels of performance accuracy. In currently limited neuroimaging data, the hybrid methods that combine traditional machine learning methods for diagnostic classification with deep learning approaches for feature extraction yielded better performance and can be a good alternative to handle the limited data. Here, an auto-encoder (AE) was used to decode the original image values, making them similar to the original image, which it then included as input, thereby effectively utilizing the limited neuroimaging data. Although hybrid approaches have yielded relatively good results, they do not take full advantage of deep learning, which automatically extracts features from large amounts of neuroimaging data. The most commonly used deep learning method in computer vision studies is the CNN, which specializes in extracting characteristics from images. Recently, 3D CNN models using multimodal PET images [FDG-PET and [18F]florbetapir PET] showed better performance for AD/CN classification and for the prediction of MCI to AD conversion.

## DISCUSSION

Effective and accurate diagnosis of Alzheimer's disease (AD) is important for initiation of effective treatment. Particularly, early diagnosis of AD plays a significant role in therapeutic development and ultimately for effective patient care. In this study, we performed a systematic review of deep learning approaches based on neuroimaging data for diagnostic classification of AD. We analyzed 16 articles published between 2013 and 2018 and classified them according to deep learning algorithms and neuroimaging types. Among 16 papers, 4 studies used a hybrid method to combine deep learning and traditional machine learning approaches as a classifier, and 12 studies used only deep learning approaches. In a limited available neuroimaging data set, hybrid methods have produced accuracies of up to 98.8% for AD classification and 83.7% for prediction of conversion from MCI to AD. Deep learning approaches have yielded accuracies of up to 96.0% for AD classification and 84.2% for MCI conversion prediction. While it is a source of concern when experiments obtain a high accuracy using small amounts of data, especially if the method is vulnerable to overfitting, the highest accuracy of 98.8% was due to the SAE procedure, whereas the 96% accuracy was due to the amyloid PET scan, which included pathophysiological information regarding AD. The highest accuracy for the AD classification was 87% when 3DCNN was applied from the MRI without the feature extraction step (Cheng et al., 2017). Therefore, two or more multimodal neuroimaging data types have been shown to produce higher accuracies than a single neuroimaging type.

In traditional machine learning, well-defined features influence performance results. However, the greater the complexity of the data, the more difficult it is to select optimal features. Deep learning identifies optimal features automatically from the data (i.e., the classifier trained by deep learning finds features that have an impact on diagnostic classification without human intervention). Because of its ease-of-use and better performance, deep learning has been used increasingly for medical image analysis. The number of studies of AD using CNN, which show better performance in image recognition among deep learning algorithms, has increased drastically since 2015. This is consistent with a previous survey showing that the use of deep learning for lesion classification, detection, and segmentation has also increased rapidly since 2015 (Litjens et al., 2017).

Recent trends in the use of deep learning are aimed at faster analysis with better accuracy than human practitioners. Google's well-known study for the diagnostic classification of diabetic retinopathy (Gulshan et al., 2016) showed classification performance that goes well beyond that of a skilled professional. The diagnostic classification by deep learning needs to show consistent performance under various conditions, and the predicted classifier should be interpretable. In order for diagnostic classification and prognostic prediction using deep learning to reach readiness for real world clinical applicability, several issues need to be addressed, as discuss below.

## Transparency

Traditional machine learning approaches may require expert involvement in preprocessing steps for feature extraction and selection from images. However, since deep learning does not require human intervention but instead extracts features directly from the input images, the data preprocessing procedure is not routinely necessary, allowing flexibility in the extraction of properties based on various data-driven inputs. Therefore, deep learning can create a good, qualified model at each time of the run. The flexibility has shown deep learning to achieve a better performance than other traditional machine learning that relies on preprocessing (Bengio, 2013). However, this aspect of deep learning necessarily brings uncertainty over which features would be extracted at every epoch, and unless there is a special design for the feature, it is very difficult to show which specific features were extracted within the networks (Goodfellow et al., 2016). Due to the complexity of the deep learning algorithm, which has multiple hidden layers, it is also difficult to determine how those selected features lead to a conclusion and to the relative importance of specific features or subclasses of features. This is a major limitation for mechanistic studies where understanding the informativeness of specific features is desirable for model building. These uncertainties and complexities tend to make the process of achieving high accuracy opaque and also make it more difficult to correct any biases that arise from a given data set. This lack of clarity also limits the applicability of obtained results to other use cases.

The issue of transparency is linked to the clarity of the results from machine learning and is not a problem limited to deep learning (Kononenko, 2001). Despite the simple principle, the complexity of the algorithm makes it difficult to describe mathematically. When one perceptron advances to a neural network by adding more hidden layers, it becomes even more difficult to explain why a particular prediction was made. AD classification based on 3D multimodal medical images with deep learning involves non-linear convolutional layers and pooling that have different dimensionality from the source data, making it very difficult to interpret the relative importance of discriminating features in original data space. This is a fundamental challenge in view of the importance of anatomy in the interpretation of medical images, such as MRI or PET scans. The more advanced algorithm generates plausible results, but the mathematical background is difficult to explain, although the output for diagnostic classification should be clear and understandable.

## Reproducibility

Deep learning performance is sensitive to the random numbers generated at the start of training, and hyper-parameters, such as learning rates, batch sizes, weight decay, momentum, and dropout probabilities, may be tuned by practitioners (Hutson, 2018). To produce the same experimental result, it is important to set the same random seeds on multiple levels. It is also important to maintain the same code bases (Vaswani et al., 2018), even though the hyper-parameters and random seeds were not, in most cases, provided in our study. The uncertainty of the configuration and the randomness involved in the training procedure may make it difficult to reproduce the study and achieve the same results.

When the available neuroimaging data is limited, careful consideration at the architectural level is needed to avoid the issues of overfitting and reproducibility. Data leakage in machine learning (Smialowski et al., 2009) occurs when the data set framework is designed incorrectly, resulting in a model that uses inessential additional information for classification. In the case of diagnostic classification for the progressive and irreversible Alzheimer's disease, all subsequent MRI images should be labeled as belonging to a patient with Alzheimer's disease. Once the brain structure of the patient is shared by both the training and testing sets, the morphological features of the patient's brain greatly influence the classification decision, rather than the biomarkers of dementia. In the present study, articles were excluded from the review if the data set configurations did not explicitly describe how to prevent data leakage (**Figure 4**).

Future studies ultimately need to replicate key findings from deep learning on entirely independent data sets. This is now widely recognized in genetics (König, 2011; Bush and Moore, 2012) and other fields but has been slow to penetrate deep learning studies employing neuroimaging data. Hopefully the emerging open ecology of medical research data, especially in the AD and related disorders field (Toga et al., 2016; Reas, 2018), will provide a basis to remediate this problem.

## OUTLOOK AND FUTURE DIRECTION

Deep Learning algorithms and applications continue to evolve, producing the best performance in closed-ended cases, such as image recognition (Marcus, 2018). It works particularly well when inference is valid, i.e., the training and test environments are similar. This is especially true in the study of AD when using neuroimages (Litjens et al., 2017). One weakness of deep learning is that it is difficult to modify potential bias in the network when the complexity is too great to guarantee transparency and reproducibility. The issue may be solved through the accumulation of large-scale neuroimaging data and by studying the relationships between deep learning and features. Disclosing the parameters used to obtain the results and mean values from sufficient experimentations can mitigate the issue of reproducibility.

Not all problems can be solved with deep learning. Deep learning that extracts attributes directly from the input data without preprocessing for feature selection has difficulty integrating different formats of data as an input, such as neuroimaging and genetic data. Because the adjustment of weights for the input data is performed automatically within a closed network, adding additional input data into the closed network causes confusion and ambiguity. A hybrid approach, however, puts the additional information into machine learning parts and the neuroimages into deep learning parts before combining the two results.

Progress will be made in deep learning by overcoming these issues while presenting problem-specific solutions. As more and more data are acquired, research using deep learning will become more impactful. The expansion of 2D CNN into 3D CNN is important, especially in the study of AD, which deals with multimodal neuroimages. In addition, Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) may be applicable for generating synthetic medical images for data augmentation. Furthermore, reinforcement learning (Sutton and Barto, 2018), a form of learning that adapts to changes in data as it makes its own decision based on the environment, may also demonstrate applicability in the field of medicine.

AD research using deep learning is still evolving to achieve better performance and transparency. As multimodal neuroimaging data and computer resources grow rapidly, research on the diagnostic classification of AD using deep learning is shifting toward a model that uses only deep learning algorithms rather than hybrid methods, although methods need to be developed to integrate completely different formats of data in a deep learning network.

## DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

TJ and AS: conceptualization and study design. TJ: data collection and analysis and drafting manuscript. TJ, KN, and AS: revision of the manuscript for important scientific content and final approval.

## FUNDING

This review was supported, in part, by grants from the National Institutes of Health (NIH) and include the following sources: P30 AG10133, R01 AG19771, R01 AG057739, R01 CA129769, R01 LM012535, and R03 AG054936. Many studies reviewed here analyzed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) that was funded by the National Institutes of Health (U01 AG024904) and Department

## REFERENCES

- Aderghal, K., Benois-Pineau, J., Afdel, K., and Catheline, G. (2017). "FuseMe: classification of sMRI images by fusion of deep CNNs in 2D+ $\epsilon$  projections," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing* (New York, NY).
- Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's Dementia* 14, 367–429. doi: 10.1016/j.jalz.2018.02.001
- Bengio, Y. (2009). Learning deep architectures for AI. *Found. Trends Mach. Learn.* 2, 1–127. doi: 10.1561/2200000006
- Bengio, Y. (2013). "Deep learning of representations: looking forward," in *International Conference on Statistical Language and Speech Processing* (Tarragona: Springer), 1–37.
- Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1798–1828. doi: 10.1109/TPAMI.2013.50
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press.
- Bottou, L. (2010). "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010* (Paris: Springer), 177–186.
- Boureau, Y.-L., Ponce, J., and Lecun, Y. (2010). "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa), 111–118.
- Bush, W. S., and Moore, J. H. (2012). Genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822. doi: 10.1371/journal.pcbi.1002822
- Cheng, D., and Liu, M. (2017). "CNNs based multi-modality classification for AD diagnosis," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (Shanghai), 1–5.
- Cheng, D., Liu, M., Fu, J., and Wang, Y. (2017). "Classification of MR brain images by combination of multi-CNNs for AD diagnosis," in *Ninth International Conference on Digital Image Processing (ICDIP 2017)* (Hong Kong: SPIE), 5.
- Choi, H., and Jin, K. H. (2018). Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav. Brain Res.* 344, 103–109. doi: 10.1016/j.bbr.2018.02.017
- Ciregan, D., Meier, U., and Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (Providence, RI), 3642–3649.
- De strooper, B., and Karran, E. (2016). The cellular phase of Alzheimer's disease. *Cell* 164, 603–615. doi: 10.1016/j.cell.2015.12.056
- Farabet, C., Couprie, C., Najman, L., and Lecun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1915–1929. doi: 10.1109/TPAMI.2012.231
- Fukushima, K. (1975). Cognitron: a self-organizing multilayered neural network. *Biol. Cybernet.* 20, 121–136. doi: 10.1007/BF00342633

of Defense (W81XWH-12-2-0012) and a consortium of private partners.

## ACKNOWLEDGMENTS

We are grateful to all of the study participants and their families that participated in the neuroimaging research on Alzheimer's disease reviewed here. We are also indebted to the clinical and computational researchers who reported their results, facilitating the analyses and discussion in this systematic review. We thank Paula J. Bice, Ph.D., for editorial assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2019.00220/full#supplementary-material>

- Fukushima, K. (1979). Neural network model for a mechanism of pattern recognition unaffected by shift in position-Neocognitron. *IEICE Tech. Rep. A* 62, 658–665.
- Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202. doi: 10.1007/BF00344251
- Galvin, J. E. (2017). Prevention of Alzheimer's disease: lessons learned and applied. *J. Am. Geriatr. Soc.* 65, 2128–2133. doi: 10.1111/jgs.14997
- Glorot, X., Bordes, A., and Bengio, Y. (2011). "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (Fort Lauderdale, FL), 315–323.
- Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Montreal, QC), 2672–2680.
- Gulshan, V., Peng, L., Coram, M., Stumpe MC, Wu D, Narayanaswamy A, et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. doi: 10.1001/jama.2016.17216
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.* 29, 82–97. doi: 10.1109/MSP.2012.2205597
- Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527
- Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647
- Hinton, G. E., and Zemel, R. S. (1994). "Autoencoders, minimum description length and Helmholtz free energy," in *Advances in Neural Information Processing Systems 6*, eds J. D. Cowan, G. Tesauro, and J. Alsppector (Denver, CO), 3–10.
- Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science* 359, 725–726. doi: 10.1126/science.359.6377.725
- Ivakhnenko, A. G. (1968). The group method of data of handling; a rival of the method of stochastic approximation. *Sov. Autom. Control* 13, 43–55.
- Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE Trans. Syst. Man Cybern.* SMC-1, 364–378. doi: 10.1109/TSMC.1971.4308320
- Ivakhnenko, A. G. E., and Lapa, V. G. (1965). *Cybernetic Predicting Devices*. New York, NY: CCM Information Corporation.
- König, I. R. (2011). Validation in genetic association studies. *Brief. Bioinformatics* 12, 253–258. doi: 10.1093/bib/bbq074

- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artif. Intell. Med.* 23, 89–109. doi: 10.1016/S0933-3657(01)00077-X
- Korolev, S., Safullin, A., Belyaev, M., and Dodonova, Y. (2017). “Residual and plain convolutional neural networks for 3D brain MRI classification,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (Melbourne, VIC), 835–838.
- Krizhevsky, A., and Hinton, G. E. (2011). “Using very deep autoencoders for content-based image retrieval,” in *Proceedings of the 19th European Symposium on Artificial Neural Networks: ESANN 2011* (Bruges), 2.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Stateline, NV), 1097–1105.
- Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y. (2007). “An empirical evaluation of deep architectures on problems with many factors of variation,” in *Proceedings of the 24th International Conference on Machine Learning* (Corvallis, OR: ACM), 473–480.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521:436. doi: 10.1038/nature14539
- Lecun, Y., Touresky, D., Hinton, G., and Sejnowski, T. (1988). “A theoretical framework for back-propagation,” in *Proceedings of the 1988 Connectionist Models Summer School: CMU* (Pittsburgh, PA: Morgan Kaufmann), 21–28.
- Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., and Li, J. (2015). A robust deep model for improved classification of AD/MCI patients. *IEEE J. Biomed. Health Inform.* 19, 1610–1616. doi: 10.1109/JBHI.2015.2429556
- Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., et al. (2014). “Deep learning based imaging data completion for improved brain disease diagnosis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Vol 17* (Boston, MA), 305–312.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005
- Liu, M., Cheng, D., Yan, W., and Alzheimer's Disease Neuroimaging Initiative. (2018a). Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front. Neuroinform.* 12:35. doi: 10.3389/fninf.2018.00035
- Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018b). Landmark-based deep multi-instance learning for brain disease diagnosis. *Med. Image Anal.* 43, 157–168. doi: 10.1016/j.media.2017.10.005
- Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., et al. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's Disease. *IEEE Trans. Biomed. Eng.* 62, 1132–1140. doi: 10.1109/TBME.2014.2372011
- Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., and Feng, D. (2014). “Early diagnosis of Alzheimer's disease with deep learning,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (Beijing), 1015–1018.
- Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci. Rep.* 8:5697. doi: 10.1038/s41598-018-22871-z
- Lu, D., and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* 28, 823–870. doi: 10.1080/01431160600746456
- Makhzani, A., and Frey, B. (2015). “k-sparse autoencoders,” in *Advances in Neural Information Processing Systems 28* (Montreal, QC), 2791–2799.
- Marcus, G. (2018). Deep learning: a critical appraisal. *arXiv preprint*. arXiv:1801.00631.
- McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. doi: 10.1007/BF02478259
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). “Distributed representations of words and phrases and their compositionality,” in *Advances in Neural Information Processing Systems 26*, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Stateline, NV), 3111–3119.
- Minsky, M., and Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* 151, 264–269. doi: 10.7326/0003-4819-151-4-200908180-00135
- Nair, V., and Hinton, G. E. (2010). “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa), 807–814.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y. (2011). “Multimodal deep learning,” in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (Bellevue), 689–696.
- Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., et al. (2014). Deep learning for neuroimaging: a validation study. *Front. Neurosci.* 8:229. doi: 10.3389/fnins.2014.00229
- Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage* 155, 530–548. doi: 10.1016/j.neuroimage.2017.03.057
- Reas, E. (2018). ADNI: understanding Alzheimer's disease through collaboration and data sharing. *PLoS Blogs*. Retrieved from: <https://blogs.plos.org/neuro/2018/10/24/adni-understanding-alzheimers-disease-through-collaboration-and-data-sharing/> (accessed October 24, 2018).
- Riedel, B. C., Daianu, M., Ver Steeg, G., Mezher, A., Salminen, L. E., Galstyan, A., et al. (2018). Uncovering biologically coherent peripheral signatures of health and risk for Alzheimer's disease in the aging brain. *Front. Aging Neurosci.* 10:390. doi: 10.3389/fnagi.2018.00390
- Ripley, B. D., and Hjort, N. (1996). *Pattern Recognition and Neural Networks*. New York, NY: Cambridge University Press.
- Rosenblatt, F. (1957). *The Perceptron, A Perceiving and Recognizing Automaton Project Para*. Buffalo, NY: Cornell Aeronautical Laboratory.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65:386. doi: 10.1037/h0042519
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323:533. doi: 10.1038/323533a0
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comp. Vision* 115, 211–252. doi: 10.1007/s11263-015-0816-y
- Salakhutdinov, R., and Larochelle, H. (2010). “Efficient learning of deep Boltzmann machines,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Sardinia), 693–700.
- Samper-Gonzalez, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., et al. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. *NeuroImage* 183, 504–521. doi: 10.1016/j.neuroimage.2018.08.042
- Schalkoff, R. J. (1997). *Artificial Neural Networks*. New York, NY: McGraw-Hill.
- Schelke, M. W., Attia, P., Palenchar, D. J., Kaplan, B., Mureb, M., Ganzer, C. A., et al. (2018). Mechanisms of risk reduction in the clinical practice of Alzheimer's disease prevention. *Front. Aging Neurosci.* 10:96. doi: 10.3389/fnagi.2018.00096
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003
- Smialowski, P., Frishman, D., and Kramer, S. (2009). Pitfalls of supervised feature selection. *Bioinformatics* 26, 440–443. doi: 10.1093/bioinformatics/btp621
- Suk, H.-I., Lee, S.-W., Shen, D., and Alzheimer's Disease Neuroimaging Initiative. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* 101, 569–582. doi: 10.1016/j.neuroimage.2014.06.077
- Suk, H.-I., Lee, S.-W., Shen, D., and the Alzheimer's Disease Neuroimaging, I. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct. Funct.* 220, 841–859. doi: 10.1007/s00429-013-0687-3
- Suk, H.-I., and Shen, D. (2013). “Deep learning-based feature representation for AD/MCI classification,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Vol 16* (Nagoya), 583–590.
- Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). “On the importance of initialization and momentum in deep learning,” in *International Conference on Machine Learning* (Atlanta), 1139–1147.
- Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

- Toga, A. W., Bhatt, P., and Ashish, N. (2016). Global data sharing in Alzheimer's disease research. *Alzheimer Dis. Assoc. Disord.* 30:160. doi: 10.1097/WAD.0000000000000121
- Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A. N., Gouws, S., et al. (2018). "Tensor2tensor for neural machine translation," in *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas* (Boston, MA), 193–199.
- Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement* 15, 106–152. doi: 10.1016/j.jalz.2018.08.005
- Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P. A. (2008). "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning* (Indianapolis, IN: ACM), 1096–1103.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P. A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 11, 3371–3408.
- Vu, T. D., Yang, H.-J., Nguyen, V. Q., Oh, A. R., and Kim, M.-S. (2017). "Multimodal learning using convolution neural network and Sparse Autoencoder," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)* (Jeju), 309–312.
- Werbos, P. J. (1982). "Applications of advances in nonlinear sensitivity analysis," in *System Modeling and Optimization*, eds R. F. Drenick and F. Kozin (New York, NY: Springer), 762–770.
- Werbos, P. J. (2006). "Backwards differentiation in AD and neural nets: past links and new opportunities," in *Automatic Differentiation: Applications, Theory, and Implementations*, eds H. M. Bücker, G. Corliss, P. Hovland, U. Naumann, and B. Norris (New York, NY: Springer), 15–34.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Jo, Nho and Saykin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.