# Toward a Multimodal Computer-Aided Diagnostic Tool for Alzheimer's Disease Conversion

Danilo Pena[1], Jessika Suescun[2], Mya Schiess[2], Timothy M. Ellmore[3], Luca Giancardo[1]* and the Alzheimer's Disease Neuroimaging Initiative

[1] Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, United States, [2] Department of Neurology, McGovern Medical School, University of Texas Health Science Center, Houston, TX, United States, [3] Department of Psychology, The City College of New York, New York, NY, United States

Alzheimer's disease (AD) is a progressive neurodegenerative disorder. It is one of the leading sources of morbidity and mortality in the aging population AD cardinal symptoms include memory and executive function impairment that profoundly alters a patient's ability to perform activities of daily living. People with mild cognitive impairment (MCI) exhibit many of the early clinical symptoms of patients with AD and have a high chance of converting to AD in their lifetime. Diagnostic criteria rely on clinical assessment and brain magnetic resonance imaging (MRI). Many groups are working to help automate this process to improve the clinical workflow. Current computational approaches are focused on predicting whether or not a subject with MCI will convert to AD in the future. To our knowledge, limited attention has been given to the development of automated computer-assisted diagnosis (CAD) systems able to provide an AD conversion diagnosis in MCI patient cohorts followed longitudinally. This is important as these CAD systems could be used by primary care providers to monitor patients with MCI. The method outlined in this paper addresses this gap and presents a computationally efficient pre-processing and prediction pipeline, and is designed for recognizing patterns associated with AD conversion. We propose a new approach that leverages longitudinal data that can be easily acquired in a clinical setting (e.g., T1-weighted magnetic resonance images, cognitive tests, and demographic information) to identify the AD conversion point in MCI subjects with AUC = 84.7. In contrast, cognitive tests and demographics alone achieved AUC = 80.6, a statistically significant difference ($n = 669$, $p < 0.05$). We designed a convolutional neural network that is computationally efficient and requires only linear registration between imaging time points. The model architecture combines Attention and Inception architectures while utilizing both cross-sectional and longitudinal imaging and clinical information. Additionally, the top brain regions and clinical features that drove the model's decision were investigated. These included the thalamus, caudate, planum temporale, and the Rey Auditory Verbal Learning Test. We believe our method could be easily translated into the healthcare setting as an objective AD diagnostic tool for patients with MCI.

Keywords: mild cognitive impairment, ADNI, longitudinal, deep learning, neuroimaging, clinical features, multimodal

# INTRODUCTION

Alzheimer's disease (AD) is a progressive cognitive decline that severely disrupts activities of daily living. It is estimated that the number of people affected by AD will triple to over 120 million people by 2050, costing the United States alone billions of dollars in healthcare expenses (Lane et al., 2018). Further, no medications are currently available that can either reverse or stop the cognitive decline in subjects with AD. There is a clear need to develop novel treatments for those with AD. To accomplish this, early detection and identification of AD will facilitate the development of biomarkers and support the discovery of novel molecules by providing the right population for clinical trials.

Early dementia detection is paramount to decrease the chance of further comorbidities and mortality (Ahmed et al., 2019). This is especially relevant in clinical environments outside large academic centers, such as community hospitals, where resources are limited. Subjects with mild cognitive impairment (MCI) have many of the neurological deficits found in AD subjects. Additionally, about 10–15% of subjects with MCI will progress to AD every year (Plassman et al., 2008). This estimate is variable, with higher rates in clinical centers and some treatment trials; and lower numbers in population-based studies. Hence, subjects with MCI represent the perfect prodromal population for the exploration of conversion biomarkers, which has been one focus of the neurocognitive field (Desikan et al., 2010; Jack et al., 2010; Landau et al., 2010; Young et al., 2014; Liu et al., 2017; Ottoy et al., 2019; Giorgio et al., 2020). Creating a computer-assisted diagnosis (CAD) tool would provide an objective instrument for early AD diagnosis in patients with MCI. The vast majority of community hospitals can perform basic neuropsychological assessments and T1 magnetic resonance imaging (MRI); as such, we propose a multi-modal approach that combines both data sources to objectively and efficiently confirm the AD diagnosis in patients with MCI (which are at high risk of conversion).

For years, researchers have been investigating neuroimaging-based biomarkers in conjunction with computational tools to find early signs of AD within MCI subjects. Studies have looked at the differences between all of the combinations of healthy controls (CN), AD subjects, MCI subjects who have converted to AD (cMCI), and MCI subjects who have stayed stable (sMCI) (Mateos-Pérez et al., 2018). To determine the crucial features of an MCI subject which eventually converts to AD, we decided to focus on a cMCI vs. sMCI comparison. Current works have combined many types of data and a host of machine learning techniques. Recent papers have used T1-weighted MRI images and linear support vector machines (Sun et al., 2017; Tong et al., 2017), positron emission tomography (PET) and random forests (Nozadi et al., 2018), clinical information/neuropsychological measurements with ensemble learning (Grassi et al., 2019), and T1-weighted and diffusion MRI with linear models (Xu et al., 2019) to predict MCI conversion. However, many of these techniques require dimensionality reduction techniques, feature selection, lengthy image pre-processing pipelines, and other tabular data

transformations that all require *a priori* hypotheses and increase the model and hyperparameter search space (Moradi et al., 2015; Ahmed et al., 2019).

Thus, scientists have turned to deep learning methods to abstract some of these steps that may incur bias throughout the pipeline. This class of models allows the incorporation of different types of data that form complex, non-linear relationships that could potentially provide more information about the conversion risk of an MCI subject. Some of the recent deep learning techniques for MCI classification use multimodal data types. These include T1-weighted MRI imaging with clinical variables (Spasov et al., 2019), cerebrospinal fluid imaging and longitudinal brain volumetric features (Lee et al., 2019a), T1-weighted and hippocampal imaging (Li et al., 2019), and a recurrent neural network (RNN) structure that uses cerebrospinal fluid, cognitive, and imaging biomarkers (Lee et al., 2019b). Using an array of data has been shown to have additive effects over using one data type alone for MCI classification. Researchers are also interested in developing a better understanding of the disease progression. Groups have predicted MCI clinical trajectories through a longitudinal feature framework (Bhagwat et al., 2018) and have used gray matter density maps at multiple time points as inputs to an RNN (Cui et al., 2019). This extension of data through time within one subject's trajectory has proven a complicated but necessary problem to be able to incorporate all potential clinically available data (Lawrence et al., 2017). This is a non-exhaustive list of neuroimaging deep learning models for AD/MCI detection and prediction, and we refer to recent comprehensive reviews (Rathore et al., 2017; Ansart et al., 2021) for a complete list. This body of work focuses on the prediction of future AD in MCI subjects or diagnosis of AD using cohorts of subjects included in studies after their AD diagnosis, and therefore likely to have the disease for many years. To our knowledge, limited to no attention has been given to the development of automated CAD systems able to diagnose the conversion from MCI to AD, in patient cohorts followed longitudinally. This is important as these CAD systems could be used by neurologists and non-specialized physicians to monitor their MCI patients.

In our work, we propose to fill in this gap with a model that combines multi-modal longitudinal data that can be easily acquired in the vast majority of clinical settings in the industrialized world (e.g., T1-weighted magnetic resonance images, cognitive tests, and demographic information). This model is based on a compact convolutional neural network architecture that combines Attention and Inception modules which is computationally efficient and requires only linear registration between imaging time points. We test the conversion diagnosis performance of our model in a cohort of subjects that received a confirmed AD diagnosis after having MCI in a previous visit (cMCI) and subjects that remained with a stable MCI diagnosis (sMCI). Our dataset has a relatively large sample size (440 sMCI vs. 229 cMCI) compared to related methodological studies, which has been a common criticism (Mateos-Pérez et al., 2018).

# MATERIALS AND METHODS

## Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1] in October 2019. The ADNI was launched in 2003 as a public–private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

Demographic information used in this study is displayed in **Table 1**. The majority of subjects were categorized as white (>93%) and non-Hispanic (>97%). Differences between sex counts were tested using Fisher's exact test, and differences in baseline age, time between sessions, and years of education were evaluated with Wilcoxon rank-sum tests. $p$-Values of less than 0.05 were considered statistically significant.

For each subject, T1-weighted structural magnetic resonance images (MRI) were taken at two different time points in addition to clinical and demographic variables (age, sex, and years of education) available from ADNI. The clinical variables included APOe4 genotypes, neuropsychological cognitive tests like Montreal Cognitive Assessment (MoCA), Mini-Mental State Exam (MMSE), and the Dementia Rating Scale (CDRSB), the AD Assessment Scale (ADAS13, ADAS11, and ADASQ4), memory evaluations from the Rey Auditory Verbal Learning Test (RAVLT), and the functional activities questionnaire (FAQ). Additionally, we used AD and CN subjects to pre-train the model, and these subjects' demographics are in **Supplementary Material**.

The time points used for cMCI subjects were chosen by selecting the session when the subjects were diagnosed with AD (session two) and the previous session where the subjects were still not converted (session one). Sessions for the subjects in the other cohorts (sMCI, AD, and CN) were chosen by selecting two consecutive sessions where both imaging and clinical evaluation were present. The current dataset did not allow a design to match the time between sessions for the whole cohort, this potential confounder is accounted for in our analysis.

Mild cognitive impairment conversion was clinically adjudicated by trained clinicians as described in the ADNI protocol. Any subject who converted back from AD to MCI was excluded from the study. Any subject included in the

---

[1]adni.loni.usc.edu

---

**TABLE 1** | Demographics and time between imaging sessions of MCI subjects used in this study (1 SD).

| | sMCI | cMCI | p-Value |
|---|---|---|---|
| Number of subjects | 440 | 229 | |
| Baseline age, years [mean (SD)] | 73.4 (7.7) | 74.2 (7.1) | 0.167 |
| Time between sessions, years [mean (SD)] | 2.9 (2.2) | 3.7 (1.9) | <0.0001 |
| Years of education [mean (SD)] | 15.8 (2.9) | 15.8 (2.7) | 0.831 |
| Sex [male, n (%)] | 260 (59.1) | 134 (58.5) | 0.934 |

sMCI cohort remained stable for all sessions present in the ADNI dataset. For the subjects who converted (cMCI), the MRI images selected were based on the closest imaging session to the conversion adjudication; as such, we assumed that the T1 brain image would be representative of the status of the subject at the time of conversion as it is unlikely to significantly change in this time period. The average elapsed time between the time of conversion and the second imaging session was $-0.7 \pm 1.4$ years.

The information on the conversion date can be found in the DXSYM_PDXCONV_ADNIALL.csv file from the ADNI database.
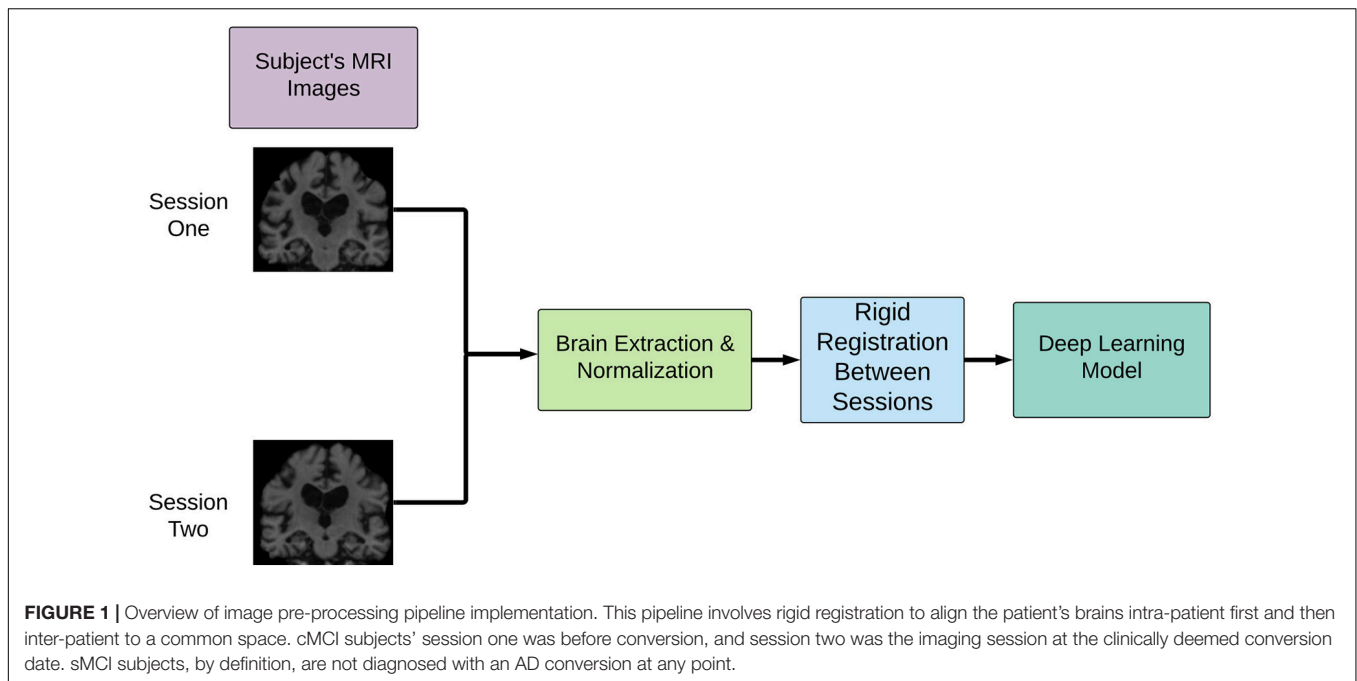
## Image Preprocessing Pipeline

As shown in **Figure 1**, the T1-weighted MRI images were pre-processed according to the steps outlined in our previous work (Pena et al., 2019). In summary, the two images at two time points were normalized and aligned to each other first and then registered to a common space using a linear registration algorithm. The normalization involved motion correction, non-uniform intensity normalization, and skull strip as implemented in the first pre-processing stages of the Freesurfer 6.0 pipeline. The final common interpatient space was derived from 2 mm MNI T1 template which was cropped of the background space to reduce the computational complexity of the network for a final resolution of $64 \times 80 \times 64$. This pipeline was shown to drastically decrease the pre-processing time compared with conventional image processing pipelines such as the wull FreeSurfer-based ones (Pena et al., 2019). These steps were extended to the full MCI cohort used in this study.

Nine clinical variables were used at two different imaging sessions (e.g., cross-sectional variables). In addition, the longitudinal signed differences for each of these variables. Note that while the APOe4 genotype is not expected to change between sessions, it has followed the same processing for consistency and simplifying the evaluation of the feature importance. Age, sex, and years of education were also concatenated in the final feature vector used in the model. These clinical variables were all normalized by their mean value.

## Deep Learning Pipeline

### Experimental Design

A 10-fold stratified cross-validation procedure was employed for model training and evaluation. Each fold was split into training, validation, and test sets with proportions of 80, 10, and 10%, respectively. Each fold maintained the distribution of sMCI/cMCI. Binary cross-entropy and the Adam were the loss function and optimizers used, respectively (Kingma and Ba, 2015). Each of the 10-folds had 75 epochs, and an early stopping condition of 10 epochs was implemented based on the model's validation loss. Cyclical learning rates were used to dynamically change the learning rate throughout the training process (Smith, 2017). This method has been shown to potentially allow the model to "jump" out of local minima to subsequently find a lower minimum to reduce the overall loss. The upper and lower bounds for the learning rates were 1e−5 to 1e−8. A batch size of 4 was used in the experiments. The area under the receiver operating curves (AUC) and balanced accuracy were the experimental

**FIGURE 1 |** Overview of image pre-processing pipeline implementation. This pipeline involves rigid registration to align the patient's brains intra-patient first and then inter-patient to a common space. cMCI subjects' session one was before conversion, and session two was the imaging session at the clinically deemed conversion date. sMCI subjects, by definition, are not diagnosed with an AD conversion at any point.

evaluation metrics. The DeLong's test for statistical significance was used to test differences between AUC curves (DeLong et al., 1988). AUC curves' 95% confidence intervals were calculated using a Monte Carlo resampling simulation with 1,000 iterations, and in each iteration, 80% of the total subjects' probabilities were randomly chosen.

Two of the experiments used a transfer learning approach where an additional set of 190 AD and 243 CN subjects were first used to pre-train the network aimed at a simpler task first. None of the 433 AD/CN subjects in this pretraining step were used for the cross-validation, this avoided any risk of data leakage. The pretraining step followed all image pre-processing, hyperparameters, and initializations as stated in the text above, except for the cross-validation procedure. Finally, the pre-trained model was then used as the starting point for the weights used in the MCI classification task.

A fully connected network using only clinical variables was tested to obtain baseline comparison with the multi-modal network. This model is effectively equivalent to a logistic regression trained using the same optimization technique and validation approach as the multi-modal network; as such, it will allow for a fair evaluation of the relative improvement of adding brain imaging to the clinical data. The input feature vectors were the clinical variables, and the outputs were the same as the multi-modal network.

The experiments outlined were completed using Python 3.7, Keras version 2.2, and TensorFlow 1.14. The graphical processing units used were GeForce RTX 2080 Ti with 11 GB RAM. The training times varied between 30 and 90 s per epoch, depending on the architecture and experimental setup. The computational performance at inference time, which is more relevant to evaluate the ease of deployment of the model in a clinical environment, is discussed in section "Results."
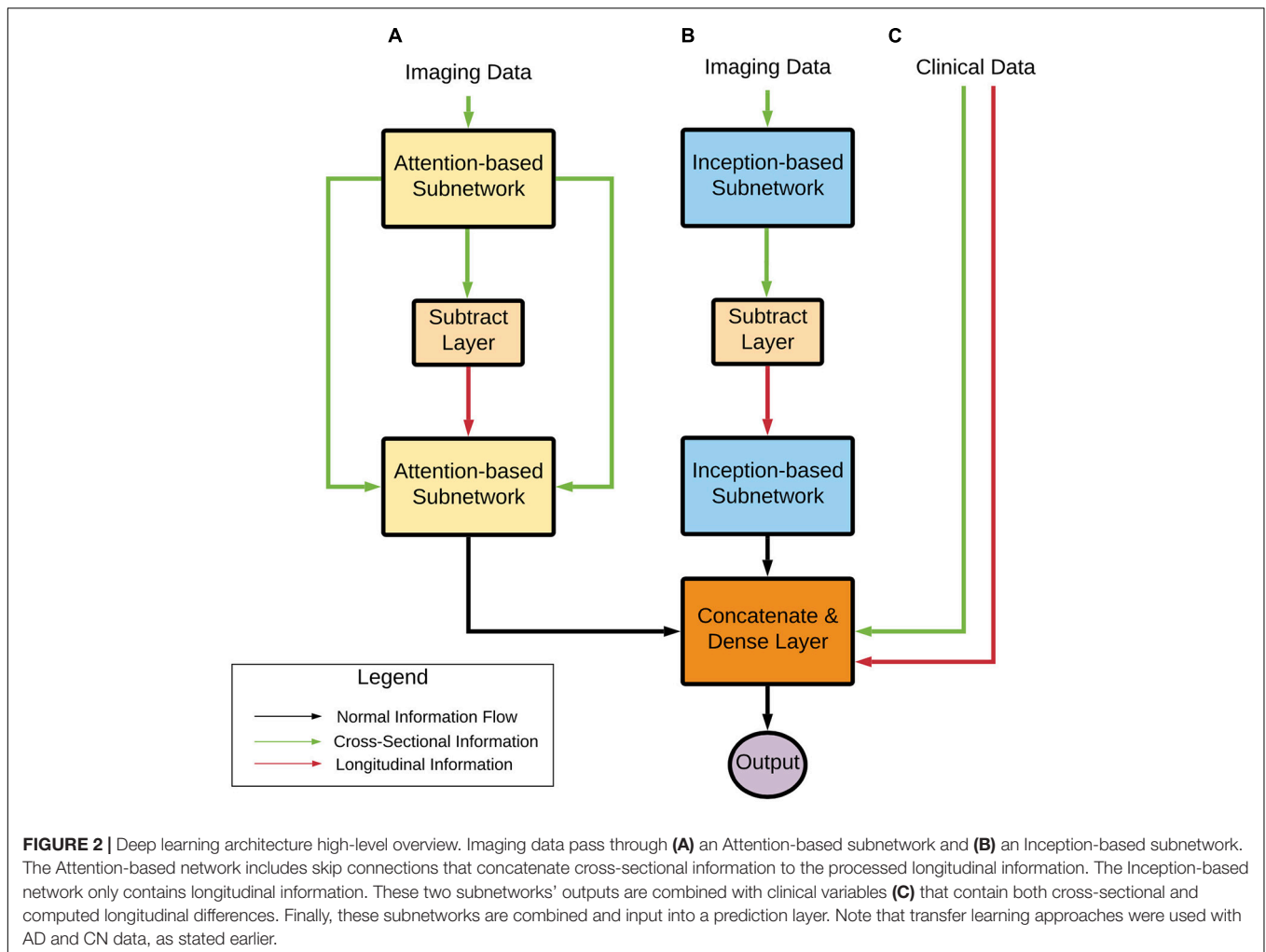
## Deep Learning Architecture

The network architecture employed was inspired by a model that learned from spatial symmetry between brain hemispheres in the stroke detection task (Barman et al., 2019; Sheth et al., 2019). Our previous work extended this model in the AD-progression and time domain (Pena et al., 2019). This study has implemented a new network that combines cross-sectional and longitudinal imaging data with clinical features, which can be trained end-to-end on the MCI conversion classification task. Further, we focused our efforts on a less parameterized network to improve computational efficiency, as we are primarily concerned with the clinical application of this class of methods. This was possible through residual attention-based modules (Wang et al., 2017), allowing the network to focus on specific areas of the image with an Inception-based network, which leads to learning convolutional filters at different scales.

From a high level (**Figure 2**), the model can learn a complex representation of two images at different time points through the two subnetworks (Attention and Inception) in addition to the temporal differences of the two brains through the subtraction layer. This subtraction layer is sensitive to changes and is referred to as the "longitudinal" portion of the network. In the attention module, cross-sectional information is added through skip connections. The output from these two subnetworks is combined with clinical variables in a final dense layer for prediction.

This model has several benefits:

- It has the potential to identify structural changes in T1-weighted MRI scans over time, which is vital for determining MCI conversion while utilizing commonly available clinical information.

**FIGURE 2 |** Deep learning architecture high-level overview. Imaging data pass through **(A)** an Attention-based subnetwork and **(B)** an Inception-based subnetwork. The Attention-based network includes skip connections that concatenate cross-sectional information to the processed longitudinal information. The Inception-based network only contains longitudinal information. These two subnetworks' outputs are combined with clinical variables **(C)** that contain both cross-sectional and computed longitudinal differences. Finally, these subnetworks are combined and input into a prediction layer. Note that transfer learning approaches were used with AD and CN data, as stated earlier.

- It uses attention-based networks and deliberately leverages a less parameterized network that inherently regularizes the weights to only focus on important information.
- It incorporates an inception-based network that allows the model to use multi-resolution to represent the images at different scales in a non-sparse fashion.

## Residual Attention Modules

Wang et al. (2017) extended the previously studied attention mechanism and applied it to their approach for image-level classification. Their overall network was composed of blocks named the residual attention module. These modules combined normal convolutional blocks (e.g., convolution, back normalization, and max pooling) with a U-Net inspired structure (Ronneberger et al., 2015) through a multiplication operator. The U-Net subunit allows the model to learn important information representing the input image through an encoder-decoder-like structure. The convolutional blocks allow the model to pay "attention" to these critical parts of the image through multiplication. This output then goes through another series of convolutional layers for further learning. Wang et al. (2017) stacked these residual attention modules to create a deep

structure with complex attention mechanisms at different scales of the images. However, to create a less parameterized network, we limited the proposed network to just one residual attention block. For additional details about these modules, we refer the readers to the original publications.

## Inception Modules

The inception modules used in this paper were inspired by the work done by Szegedy et al. (2015) and were extended to the 3-dimensional space (3D). The inception modules used were a combination of multi-resolution 3D convolutional layers. These layers were composed of three parallel operations: $1 \times 1 \times 1$, $3 \times 3 \times 3$, and $5 \times 5 \times 5$ convolutions with two filters. Previous work has shown that this module can produce meaningful results in neuroimaging applications (Barman et al., 2019, 2020; Pena et al., 2019). This style of operation allows the model to view an image or input at different scales to learn different types of spatial information. The outputs were then concatenated and served as input to the next network layer. For additional details about these modules, we refer the readers to the original publications.
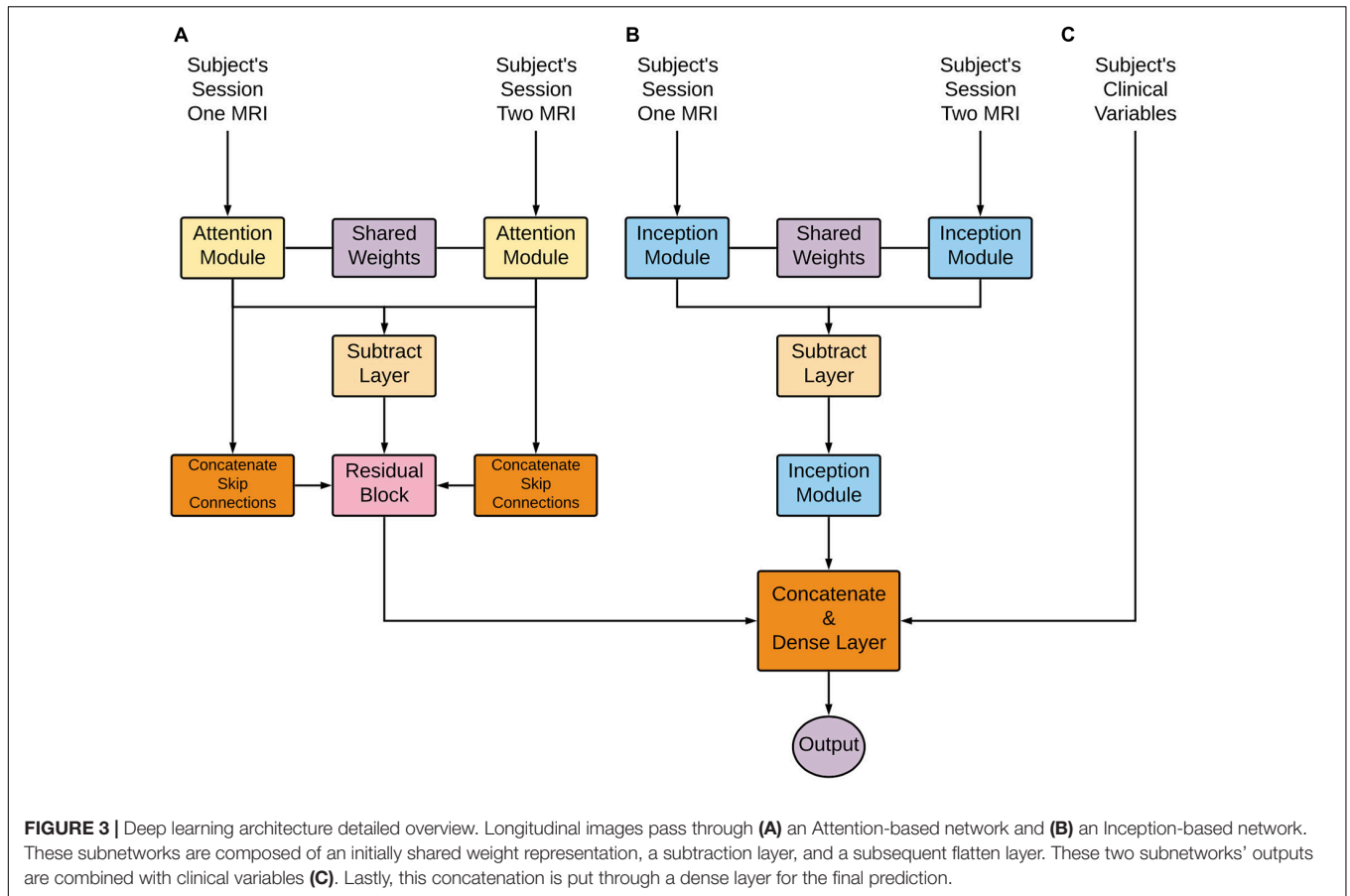
**FIGURE 3 |** Deep learning architecture detailed overview. Longitudinal images pass through **(A)** an Attention-based network and **(B)** an Inception-based network. These subnetworks are composed of an initially shared weight representation, a subtraction layer, and a subsequent flatten layer. These two subnetworks' outputs are combined with clinical variables **(C)**. Lastly, this concatenation is put through a dense layer for the final prediction.
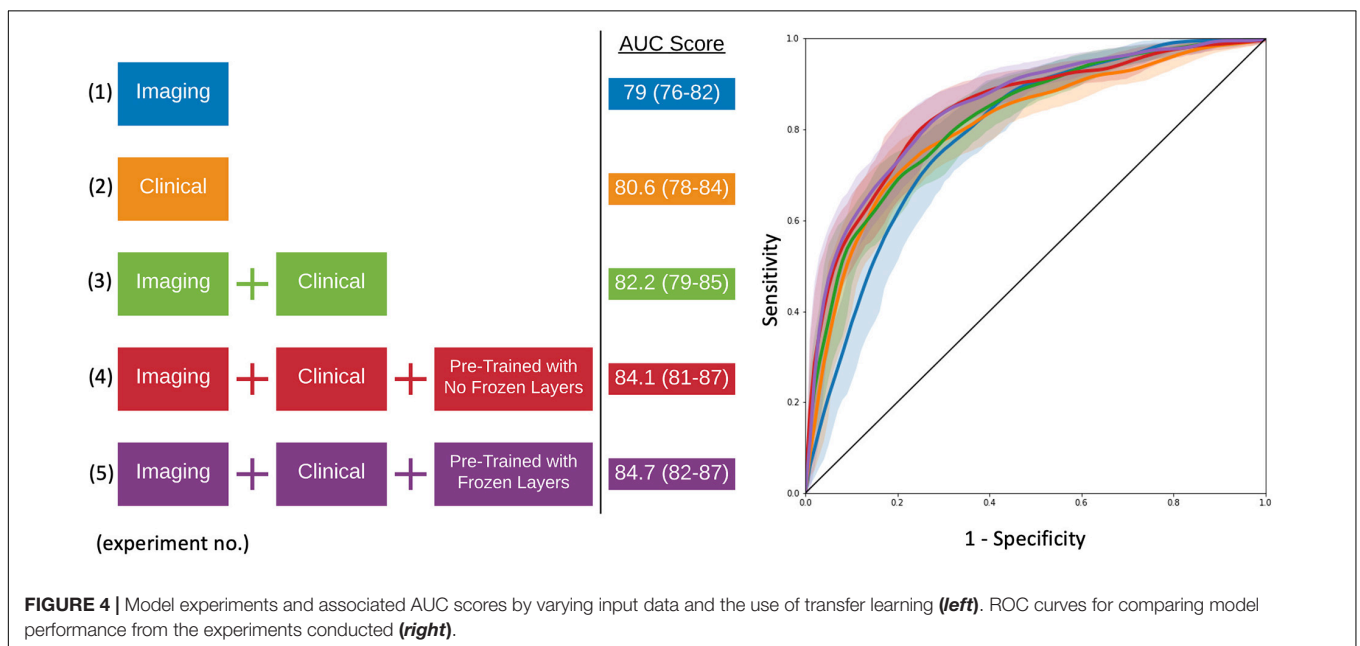


**FIGURE 4 |** Model experiments and associated AUC scores by varying input data and the use of transfer learning (*left*). ROC curves for comparing model performance from the experiments conducted (*right*).

## Layers

From the overall network perspective (**Figure 3** below), the first module layers learned a representation shared between the first and second imaging time points. This representation proceeded to a subtraction layer that took the difference between the two sessions, and this difference was the input to another module.

These longitudinal outputs then went through another module to further learn from the differences between the two sessions. Note that the attention subnetwork incorporated longitudinal and cross-sectional information through the addition of skip connections, as seen in the figure below. Next, these outputs were flattened and concatenated with each other to form an imaging and clinical feature vector. Finally, the prediction layer was used for prediction utilizing the SoftMax activation function.

The code repository for this publication can be found at https://gitlab.com/lgianca/deepsymnet-att.

## Confounding Variable Adjustment

A logistic regression model was fit with the deep neural network's probability output, baseline age, time between imaging sessions, and sex to adjust for any potential confounders inherent in the data chosen for the model. The logistic regression model coefficients, the 95% confidence intervals, and corresponding *p*-values were reported.

## Feature Importance

To develop an intuition about which voxels from the T1-weighted MRI images and features from the clinical variables, we employed the epsilon layer-wise relevance propagation (e-LRP) method (Bach et al., 2015). The e-LRP method starts from the prediction layer and works its way backward through the network. Layer-by-layer, the relevance of each of the previous layer's nodes is computed until the operation reaches the input data layer. Each feature in the input data is assigned a final relevance score that describes how important that feature was for the final prediction. The codebase used in our experiments follows the implementation of DeepExplain (Ancona et al., 2018).

As stated above, relevance scores are projected onto the input data, which, in this study, are the two T1-weighted images and the subject's clinical feature vector. We used a global and regional method to compute the magnitude of relative importance for the voxels in the MRI images. For the global method, each subject's MRI relevance map was added for both sessions, and absolute values were used to remove the risk of canceling out relevance scores. Then, a heatmap allowed for the visualization of this global method.

For the regional method, the cortical and subcortical regions were segmented for each subject *via* the Harvard–Oxford atlas (Caviness et al., 1996). Then, for each subject and session, the summation of all the voxels' magnitude in each region was calculated. This value was then divided by the volume of that particular region, resulting in a normalized relevance magnitude for a particular brain region. This final value allowed the different regions to be compared to one another on a similar scale. A similar approach was used to find the relative importance of the clinical features. The unsigned value for a clinical feature was added for each subject and then ranked in order of importance based on the magnitude of the total value. The overall method is described in greater detail in our previous work (Pena et al., 2019).

# RESULTS

This study aims to (1) evaluate the use of different deep learning architectures, input data modalities, and transfer learning for MCI conversion classification using a computationally efficient architecture and to (2) investigate the important imaging and clinical features that drove the model's decision based on the e-LRP method.

## Model Evaluation

As seen in **Figure 4** and **Table 2**, the model that used imaging and clinical input data was pre-trained using AD, and CN subjects with frozen weights had the highest AUC score (Experiment 5). This model was considered the "best" performing model in this paper. The pre-trained model where all of the weights could be fine-tuned had the best-balanced accuracy. **Table 2** also shows that the improvement between solely using clinical variables (Experiment 1) and the best model that combined clinical and T1 imaging was statistically significant. Further, our best model was the only one significantly greater than the model that used clinical variables only. The average time taken to pre-process an image and for the model to make a prediction was $129.7 \pm 19.8$ and $0.12 \pm 0.05$ s, respectively.

We evaluate the individual importance of the Inception and Attention subnetworks with ablation studies. We use as a base model and training strategy what has been described in Experiment 5. In order to account for the artificial advantage that the architectures might have solely on the basis of having more parameters, we increased the number of convolutional filters in each of the independent subnetworks to make them comparable with the full network. In **Table 3**, we show that the Inception-based subnetwork overperforms the Attention-based subnetworks. However, their combination (with the addition of the clinical data) outperformed the two architectures individually, even if the number of parameters was comparable.

In order to evaluate the computational efficiency of the model, we evaluated the time required to generate a prediction at inference time (i.e., after model training) on an off-the-shelf laptop without using any GPUs. We repeated this 100 times and achieved an average execution time of 1.56 s (0.10 std).

**TABLE 2** | Model experiments' metric comparison for balanced accuracy, AUC score, and testing for significant differences between AUC curves.

| Experiment | Balanced accuracy | AUC score | |
|---|---|---|---|
| (1) Clinical variables only | 75.5 | 80.6 | |
| (2) T1-weighted MRI only | 73.2 | 79 | p=0.51 |
| (3) Clinical variables + T1-weighted MRI | 75.4 | 82.2 | p=0.41 |
| (4) Clinical variables + T1-weighted MRI (pre-trained network: no frozen layers) | 78.2 | 84.1 | p=0.07 |
| (5) Clinical variables + T1-weighted MRI (pre-trained network: frozen layers) | 77.8 | 84.7 | p<0.05 |

*p-Values were computed from the DeLong test for correlated ROC curves to reject the null hypothesis that there is no statistical difference between the AUCs.*

**TABLE 3 |** Ablation studies indicate that the combination of the two Attention and Inception-based subnetworks overperform the two individual subnetworks.

| Experiment | Number of parameters | Balanced accuracy | AUC score |
|---|---|---|---|
| Attention-based subnetwork (imaging + pre-training with frozen layers) | 374,398 | 65.7 | 68.6 |
| Inception-based subnetwork (imaging + pre-training with frozen layers) | 333,352 | 73.6 | 77.6 |
| Best performing network (imaging + clinical + pre-training with frozen layers) | 355,572 | 77.8 | 84.7 |

*Note that the number of convolutional filters in the subnetworks were increased to have comparable parameters with the entire network (i.e., within 10%).*

This does not take into account the file conversion, initial brain extraction and linear registration steps required, which can take from tens of seconds to a few minutes, depending on the software used. This compares favorably to the "*de facto*" Freesurfer-based longitudinal pipeline that can take an average of 17 h per subject (Pena et al., 2019) or methods relying on non-linear registration and extraction of the warp field, taking each image into template space. For example Spasov et al. (2019) report approximately 19,200 h of CPU time on a high-performance parallel computing cluster to non-linearly register the images, which is ~19 h per subject.

## The Network as a Clinical Decision Support Tool

With the final model, we investigated the strength of the signal (deep network output probability) between MCI subjects who eventually converted to AD and those who stayed stable, as seen in **Figure 5**. The starting point for the cMCI subjects is higher than the sMCI subjects since there was some indication of AD conversion-like progression using MRI imaging time points before the actual conversion. However, this signal strengthens when an imaging time point around AD conversion is included, shown by the tendency toward higher probabilities on the right side of the figure. The sMCI group has a smaller slope with respect to time as there is no indication of AD conversion. This makes for a clear, qualitative difference between the two groups. The network derives a much stronger signal at the conversion point, indicating its ability to recognize patterns distinctly associated with AD conversion.
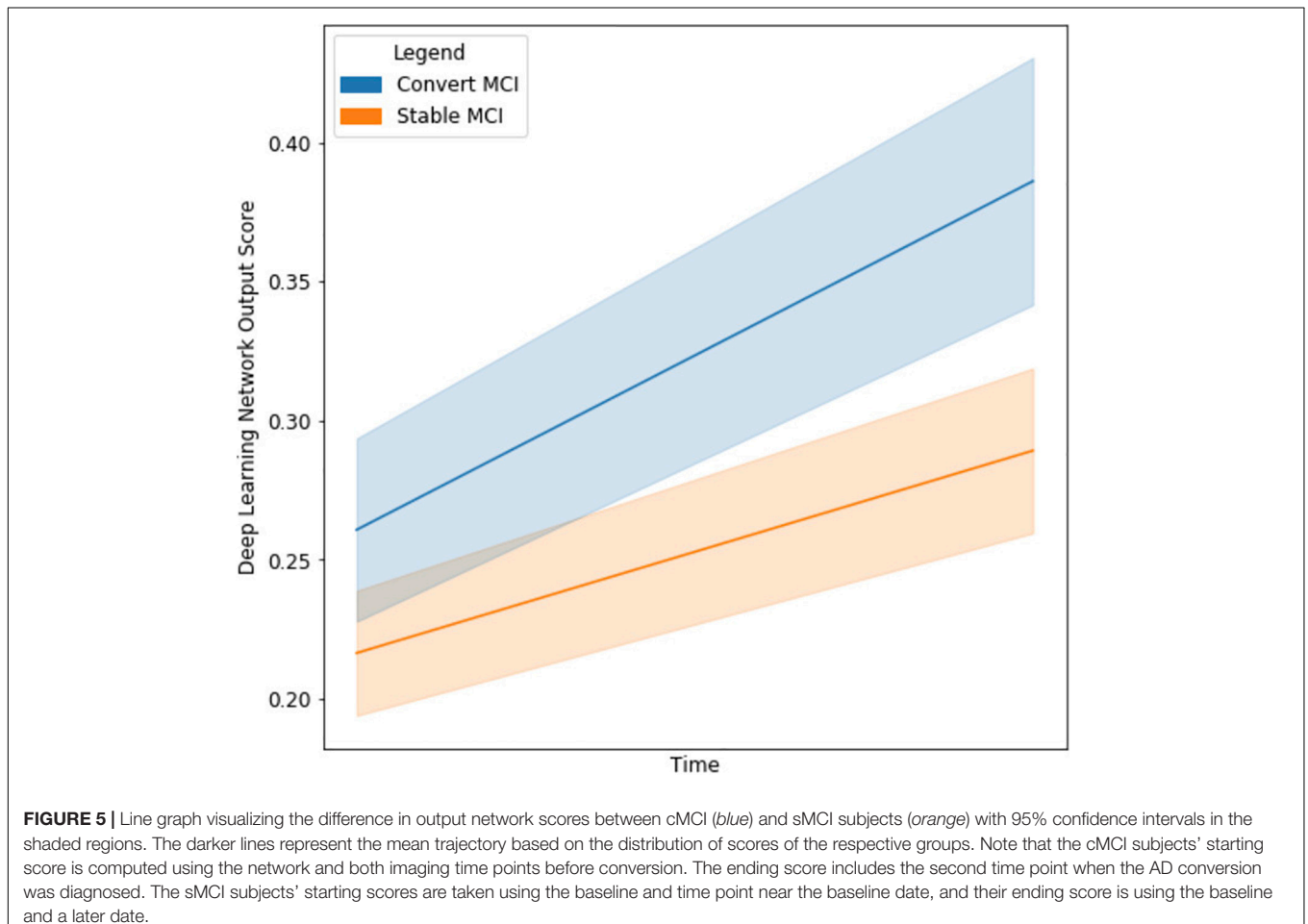


**FIGURE 5 |** Line graph visualizing the difference in output network scores between cMCI (*blue*) and sMCI subjects (*orange*) with 95% confidence intervals in the shaded regions. The darker lines represent the mean trajectory based on the distribution of scores of the respective groups. Note that the cMCI subjects' starting score is computed using the network and both imaging time points before conversion. The ending score includes the second time point when the AD conversion was diagnosed. The sMCI subjects' starting scores are taken using the baseline and time point near the baseline date, and their ending score is using the baseline and a later date.

**TABLE 4 |** Summary of the logistic regression coefficients, confidence intervals, and *p*-values for model output probability and confounding variables.

| Baseline age | Time between sessions | Years of education | Sex | Deep learning output probability |
|---|---|---|---|---|
| 0.0140 (0.001–0.027)* | 0.0916 (0.010–0.174) | 0.0361 (−0.024 to 0.096) | 0.0521 (−0.358 to 0.462) | −4.0714 (−4.700 to −3.443)*** |

*p < 0.05.*
***p < 0.0001.*

## Confounding Variable Adjustment

Further, the deep learning output probabilities were assessed for statistical significance with a logistic regression model and potential confounding variables. As seen in **Table 4**, the output probability remained statistically significant ($p < 0.0001$). Interestingly, though there were group differences between the imaging variable, as seen in **Table 1**, these differences were not significant when combined with the output probabilities.

## Feature Importance

Next, model feature importance was evaluated for both the imaging and clinical inputs. The imaging feature importance was completed on both a voxel-wise level and a brain regional level (subcortical vs. cortical regions), as seen in **Figure 6**. These saliency maps are smoother than our previous work (Pena et al., 2019), and we attributed this improvement to the use of attention-based modules and a less parameterized network. These model characteristics perform significant regularization, highlighting only the most informative regions for the given task.

Further, the top five regions from the cortical and subcortical regions were plotted in **Figures 6B,C**. After volume normalization, the thalamus, caudate, pallidum, and lateral ventricle subcortical regions contained the highest overall contribution to the model's decision. For the cortical regions, the planum temporale and parietal operculum cortex had the highest contributions. Both the cortical and subcortical regions had similar contribution magnitudes, as seen in the *x*-axis of **Figure 7**.

The clinical variables' contributions are shown in descending order in **Figure 8**. The RAVLT score from the second session was the most important clinical feature, with ADAS11 from the second session and MoCA scores from the first session following.
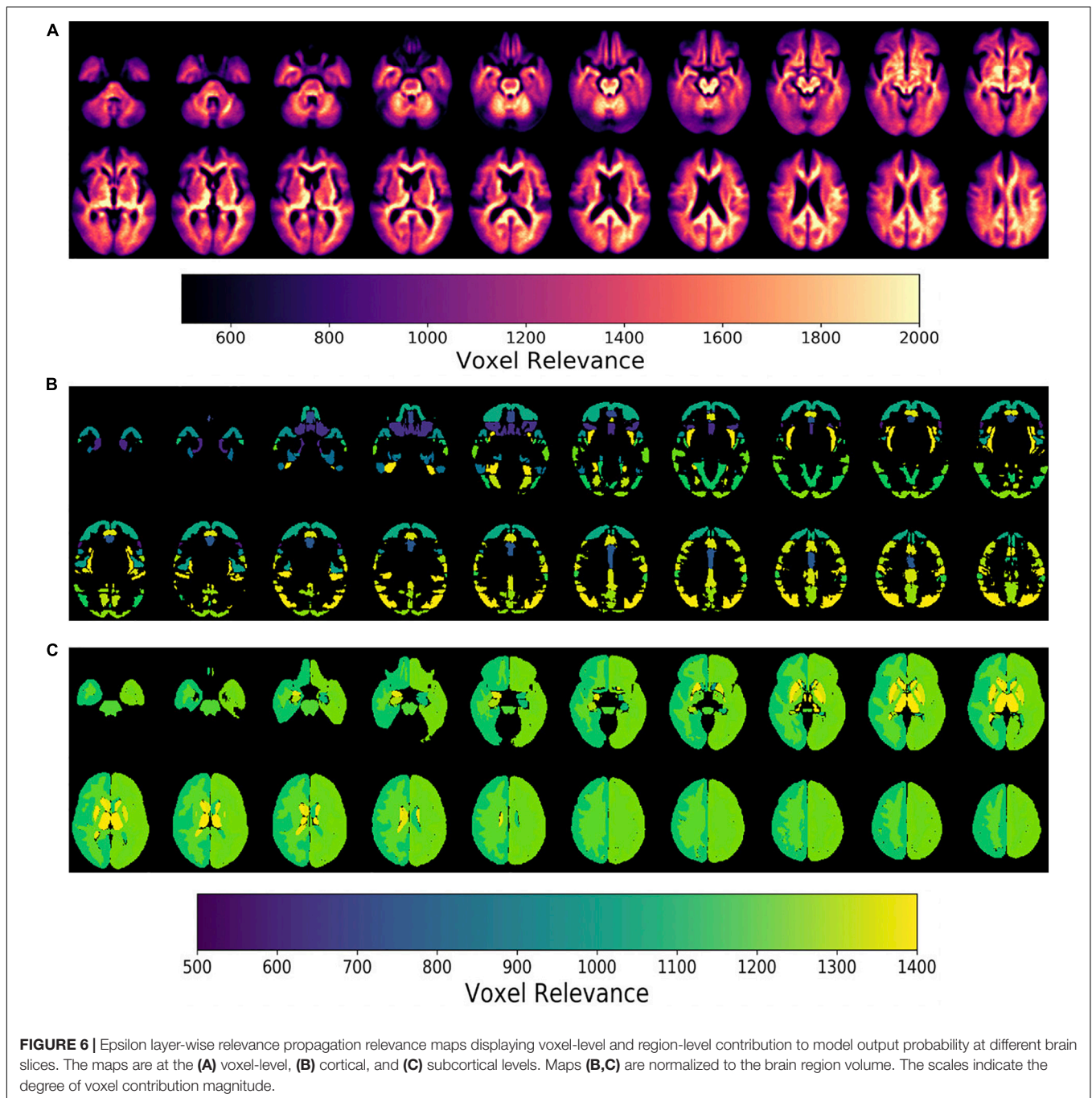
## DISCUSSION

This study employed a deep learning model to enable a CAD system able to provide an AD conversion diagnosis in an MCI cohort followed longitudinally. The model combined both Attention and Inception modules and was designed to be less parameterized to form a sparse yet rich representation of the input imaging and clinical features. The experiments performed demonstrated that the combination of imaging and clinical features produced a better model than using either type of data alone. Also, a model pre-trained on AD and CN subjects

that served as a baseline for MCI classification was a better starting point for subsequent model fine-tuning than random weight initialization. Further, the brain regions that drove the model's decision were visualized and quantified through the e-LRP method. The clinical features included in the model were also ranked and analyzed for relevance.

One of the main contributions of this network architecture is the combination of longitudinal and cross-sectional information. The subtraction operation was used between the two imaging and clinical time points and their respective features; thus, the network could learn from the differences over time. Further, this information was preserved throughout the training process by keeping the raw signal from the individual time points (e.g., cross-sectional data). The imaging-focused part of the network was divided into the Attention and Inception-based mechanisms. The attention module extended the residual attention used in computer vision, allowing the model to introduce sparsity into the network parameters. This allows the model to focus on certain parts of the brain input data related to MCI conversion to AD. The inception modules used 3D convolutional filters to find information at different spatial scales and granularity. We empirically show that using a combination of these modules, both the balanced accuracy and AUC were higher than using these modules individually in a network.

Further, we showed that the network improved AUC performance by incorporating more information in the time domain (cross-sectional and longitudinal) and in data modality (T1-weighted MRI and clinical features). This has been shown to be the case in related MCI and AD research (Goryawala et al., 2015; Spasov et al., 2019). The best model was also pre-trained on a cohort of AD and CN subjects. This model's only trainable layer was the dense layer right before the prediction layer. Exclusively fine-tuning of the penultimate layer allowed the model to focus on changing a smaller number of weights compared to the entire model. This transfer learning setup also assumed that the brain representation from the AD and CN subjects was a good representation for an MCI application, making intuitive sense since this is modeling a progression pattern in subjects at high risk of developing AD in their lifetime. Finally, after controlling for several potential confounding factors, the network output probabilities remained statistically significant. Once the model is trained, the whole model can run in ~1.5 s plus the time required to perform basic pre-processing involving file conversion, skull stripping, and linear registration (typically tens of seconds to a few minutes) on an off-the-shelf laptop without GPU. This would enable a neurologist to use this system as a computer-aided diagnostic tool during the office visit once the required imaging and/or clinical variables are acquired.
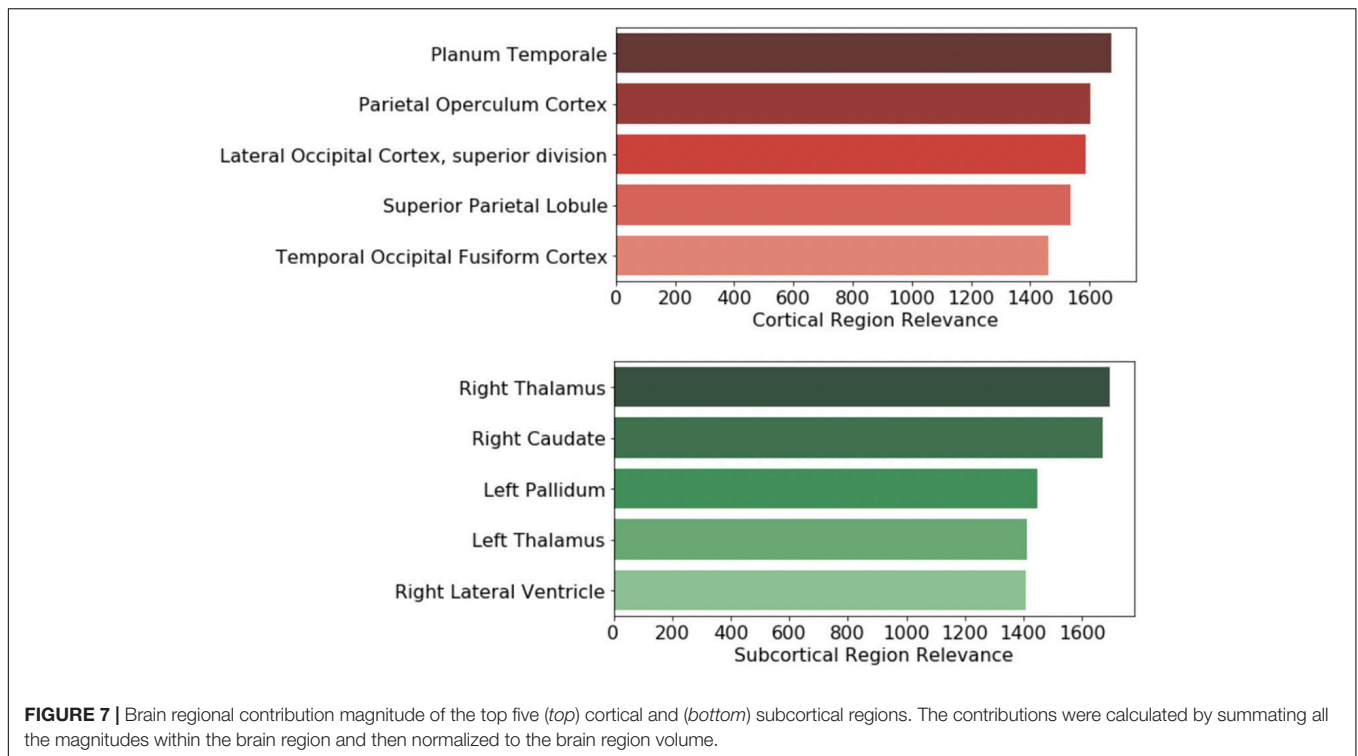
Once the experiments were completed, the crucial features for driving the model's decision were investigated. To narrow down the imaging analysis for interpretation, we focused on the subcortical and cortical regions. The thalamus, caudate, pallidum, and lateral ventricle had the highest overall activation magnitude for the subcortical regions. Cholinergic synapses have a high density in several parts of the brain, including the thalamus, and have played a central role in research in aging and

**FIGURE 6 |** Epsilon layer-wise relevance propagation relevance maps displaying voxel-level and region-level contribution to model output probability at different brain slices. The maps are at the **(A)** voxel-level, **(B)** cortical, and **(C)** subcortical levels. Maps **(B,C)** are normalized to the brain region volume. The scales indicate the degree of voxel contribution magnitude.

cognitive decline. Cholinesterase inhibitors are considered first-line treatments for mild and moderate AD (Hampel et al., 2018). Likewise, an *in vivo* imaging study found reduced serotonin transporter availability in MCI subjects in the thalamus compared to controls (Smith et al., 2017). Qing et al. (2017) found that impairment of spatial navigation skills, a clinical feature of AD found in MCI subjects, was significantly correlated to neuroimaging variable changes in the pallidum and thalamus.

Similarly, Fischer et al. (2017) investigated mobility changes in subjects with MCI. They found that decreased gray matter volume in the caudate nucleus was associated with a lower speed in functional mobility tasks. Crocco et al. (2018) applied a cognitive stress test to AD and MCI subjects and showed that negative clinical results were related to dilation of the lateral ventricle, among other regions. Yi et al. (2016) found that gray matter volumes in subcortical regions, including, but not limited to, the thalamus, caudate, and pallidum, were significantly reduced in MCI subjects when compared to controls. Additionally, many of these subcortical volumes were correlated with cognitive function.

**FIGURE 7 |** Brain regional contribution magnitude of the top five (*top*) cortical and (*bottom*) subcortical regions. The contributions were calculated by summating all the magnitudes within the brain region and then normalized to the brain region volume.

For the cortical regions, the planum temporale, operculum cortex, and occipital cortex were some of the top regions with associated findings in AD and MCI literature. Researchers using several independent AD datasets found that anatomical changes in the planum temporale and thalamus were among the top features for their predictive model (Giraldo et al., 2018; Li et al., 2020). Others have found that changes in cortical minicolumn organization and premortem cognitive scores were significantly related in the planum temporale, potentially reflecting a phenomenon in brain atrophy in AD subjects (Chance et al., 2011). Alternatively, this might indicate the importance of auditory processing in MCI to AD progression as planum temporale is involved in auditory processing. A clinical study that used functional connectivity imaging and associated metrics found decreased intrinsic connectivity in the operculum cortex among MCI and AD subjects (Xie et al., 2012). Finally, a PET study demonstrated a significant and high overlap in hypoperfusion and hypometabolism in AD subjects in the occipital cortex (Riederer et al., 2018).

From the clinical features, ADAS, MoCA, and MMSE scores are among the top five variables with the most relevance for the model's decision. This is unsurprising as MoCA and MMSE are the most widely used screening tools in clinical practice. ADAS is frequently used as a progression measurement in both clinical settings and clinical trials.

Interestingly, the RAVLT, a recent memory test, was the variable with the most relevance for the model's decision for the first session, and it was ranked as one of the top five variables for the second session. Memory for recent events is distinctively impaired in AD and is served by the hippocampus, entorhinal cortex, and related structures in the medial temporal lobe.

This could indicate that the RAVLT provides more complementary information that is harder to directly learn from the imaging alone. Multiple clinical and neuroimaging studies have shown the importance of this variable in AD and MCI research; one of the earliest was performed by Estévez-González et al. (2003). More recently, Eliassen et al. (2017) used PET imaging and clinical scores to show that RAVLT were significant predictors in changes in cortical thickness between MCI and CN participants. A neuroimaging study conducted by Moradi et al. (2017) found that the MRI-based volumetric features were suitable variables for predicted parts of the RAVLT tool using an elastic net-based linear regression model. Russo et al. (2017) found that parts of the RAVLT assessment can have differences in discrimination accuracy and response bias between MCI and AD subjects, indicating there could be diagnostic specificity if using different test portions.

Our study has some limitations. First, the absolute classification performance of our method was lower than some found in the literature (Liu et al., 2017; Tong et al., 2017; Spasov et al., 2019) that report AUC scores around 90%. However, these models focus on the prediction of future AD rather than an actual diagnosis of the AD conversion, and they typically involve a very long pre-processing pipeline that would be hard to use in clinical settings. The use of longitudinal data to output an AD diagnosis can also be considered a limitation as it requires data from two-timepoints. The dataset used did include a majority of Caucasian non-Hispanic population, as such, the generalizability of the algorithm needs to be further

**FIGURE 8 |** Clinical variable-level contribution magnitude. These values were calculated by summation of the contribution across all subjects. Neurological clinical variables denoted with "first_ses" and "second_ses" correspond to the subjects' first and second clinical sessions, respectively. The clinical variables without a suffix represent the longitudinal change in that particular variable over time. Demographic variables included were age, sex, and years of education.

confirmed on an entirely external dataset including a more diverse population. Finally, while the model can handle missing imaging or clinical data (as a whole), it currently cannot leverage clinical data with missing variables (unless imputation is used).

Future work could extend the ADNI dataset to incorporate multiple sources. This would increase the model's generalizability to bias and errors that are inherent to different datasets. Also, adding more time points by extending this model using recursive neural networks or Gaussian processes algorithms could give a more nuanced trajectory signal that may unearth a strong signal for MCI progression and conversion to AD.

## CONCLUSION

In this paper, we introduce a novel method that utilizes T1-weighted MRI and clinical data at two-time points to diagnose AD in patients with MCI. At a high level, the model is a deep learning framework that combines residual Attention and

Inception modules while taking advantage of cross-sectional and longitudinal data. The epsilon layer-wise propagation method allowed the interpretation of essential brain regions and clinical features that drove the model's output. Some of the top subcortical and cortical regions included the thalamus, caudate, planum temporale, and operculum cortex. Further, RAVLT was the clinical feature that had the highest contribution to the final prediction. This method could easily be translated to the healthcare environment because it integrates variables commonly used in a clinical setting and has a fast image processing and prediction pipeline. This instrument could potentially be used as an objective and efficient diagnostic tool for patients at high risk of AD conversion.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: adni.loni.usc.edu.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DP, LG, MS, and JS contributed to conception and design of the study. DP performed the experiments and analysis, and wrote the first draft of the manuscript. DP, TE, and LG interpreted the results. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.744190/full#supplementary-material

## REFERENCES

Ahmed, M. R., Zhang, Y., Feng, Z., Lo, B., Inan, O. T., and Liao, H. (2019). Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. *IEEE Rev. Biomed. Eng.* 12, 19–33. doi: 10.1109/RBME.2018.2886237

Ancona, M., Ceolini, E., Oztireli, C., and Gross, M. (2018). Towards a better understanding of gradient-based attribution methods for deep neural networks. *ICLR* 2018, 1–16.

Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., et al. (2021). Predicting the progression of mild cognitive impairment using machine learning: a systematic, quantitative and critical review. *Med. Image Anal.* 67:101848. doi: 10.1016/j.media.2020.101848

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* 10:e0130140. doi: 10.1371/journal.pone.0130140

Barman, A., Inam, M. E., Lee, S., Savitz, S., Sheth, S., and Giancardo, L. (2019). "Determining ischemic stroke from CT-Angiography imaging using symmetry-sensitive convolutional networks," in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (Piscataway, NJ: IEEE), doi: 10.1109/ISBI.2019.8759475

Barman, A., Lopez-Rivera, V., Lee, S., Vahidy, F. S., Fan, J. Z., Savitz, S. I., et al. (2020). "Combining symmetric and standard deep convolutional representations for detecting brain hemorrhage," in *Proceedings of the Medical Imaging 2020: Computer-Aided Diagnosis*, (Bellingham, DC), doi: 10.1117/12.2549384

Bhagwat, N., Viviano, J. D., Voineskos, A. N., Chakravarty, M. M., and Initiative, A. D. N. (2018). Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput. Biol.* 14:e1006376. doi: 10.1371/journal.pcbi.1006376

Caviness, V. S., Meyer, J., Makris, N., and Kennedy, D. N. (1996). MRI-Based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *J. Cogn. Neurosci.* 8, 566–587. doi: 10.1162/jocn.1996.8.6.566

Chance, S. A., Clover, L., Cousijn, H., Currah, L., Pettingill, R., and Esiri, M. M. (2011). Microanatomical correlates of cognitive ability and decline: normal ageing, MCI, and Alzheimer's disease. *Cereb. Cortex* 21, 1870–1878. doi: 10.1093/cercor/bhq264

Crocco, E. A., Loewenstein, D. A., Curiel, R. E., Alperin, N., Sara, J., Harvey, P. D., et al. (2018). A novel cognitive assessment paradigm to detect Pre-mild cognitive impairment (PreMCI) and the relationship to biological markers of Alzheimer's disease. *J. Psychiatr. Res.* 96, 33–38. doi: 10.1016/j.jpsychires.2017.08.015.A

Cui, R., Liu, M., Alzheimer's Disease Neuroimaging Initiative (2019). RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* 73, 1–10. doi: 10.1016/J.COMPMEDIMAG.2019.01.005

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametetric approach. *Biometrics* 44, 837–845. doi: 10.2307/2531595

Desikan, R. S., Cabral, H. J., Settecase, F., Hess, C. P., Dillon, W. P., Glastonbury, C. M., et al. (2010). Automated MRI measures predict progression to Alzheimer's disease. *Neurobiol. Aging* 31, 1364–1374. doi: 10.1016/j.neurobiolaging.2010.04.023

Eliassen, C. F., Reinvang, I., Selnes, P., Fladby, T., and Hessen, E. (2017). Convergent results from neuropsychology and from neuroimaging in patients with mild cognitive impairment. *Dement. Geriatr. Cogn. Disord.* 43, 144–154. doi: 10.1159/000455832

Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., and García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. *Int. J. Geriatr. Psychiatry* 18, 1021–1028. doi: 10.1002/gps.1010

Fischer, B. L., Bacher, R., Bendlin, B. B., Birdsill, A. C., Ly, M., Hoscheidt, S. M., et al. (2017). An examination of brain abnormalities and mobility in individuals with mild cognitive impairment and Alzheimer's disease. *Front. Aging Neurosci.* 9:86. doi: 10.3389/fnagi.2017.00086

Giorgio, J., Landau, S. M., Jagust, W. J., Tino, P., Kourtzi, Z., and Alzheimer's Disease Neuroimaging Initiative (2020). Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. *NeuroImage Clin.* 26:102199. doi: 10.1016/j.nicl.2020.102199

Giraldo, D. L., García-Arteaga, J. D., Cárdenas-Robledo, S., and Romero, E. (2018). Characterization of brain anatomical patterns by comparing region intensity distributions: applications to the description of Alzheimer's disease. *Brain Behav.* 8:e00942. doi: 10.1002/brb3.942

Goryawala, M., Zhou, Q., Barker, W., Loewenstein, D. A., Duara, R., and Adjouadi, M. (2015). Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of alzheimer's disease and mild cognitive impairment. *Comput. Intell. Neurosci.* 2015:865265. doi: 10.1155/2015/865265

Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., et al. (2019). A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* 10:756. doi: 10.3389/fneur.2019.00756

Hampel, H., Mesulam, M. M., Cuello, A. C., Farlow, M. R., Giacobini, E., Grossberg, G. T., et al. (2018). The cholinergic system in the pathophysiology and treatment of Alzheimer's disease. *Brain* 141, 1917–1933. doi: 10.1093/brain/awy132

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 9:119. doi: 10.1016/S1474-4422(09)70299-6

Kingma, D. P., and Ba, J. (2015). Adam: a method for stochastic optimization. *arXiv [preprint]*. doi: 10.1063/1.4902458

Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., et al. (2010). Comparing predictors of conversion and decline in mild cognitive impairment(Podcast)(e–Pub ahead of print). *Neurology* 75, 230–238. doi: 10.1212/WNL.0b013e3181e8e8b8

Lane, C. A., Hardy, J., and Schott, J. M. (2018). Alzheimer's disease. *Eur. J. Neurol.* 25, 59–70. doi: 10.1111/ene.13439

Lawrence, E., Vegvari, C., Ower, A., Hadjichrysanthou, C., De Wolf, F., and Anderson, R. M. (2017). A systematic review of longitudinal studies which measure Alzheimer's disease biomarkers. *J. Alzheimers Dis.* 59, 1359–1379. doi: 10.3233/JAD-170261

Lee, G., Kang, B., Nho, K., Sohn, K. A., and Kim, D. (2019a). MildInt: deep learning-based multimodal longitudinal data integration framework. *Front. Genet.* 10:617. doi: 10.3389/fgene.2019.00617

Lee, G., Nho, K., Kang, B., Sohn, K. A., Kim, D., and the Alzheimer's Disease Neuroimaging Initiative (2019b). Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Nat. Sci. Rep.* 9:1952. doi: 10.1038/s41598-018-37769-z

Li, F., Liu, M., and the Alzheimer's Disease Neuroimaging Initiative (2019). A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease. *J. Neurosci. Methods* 323, 108–118. doi: 10.1016/J.JNEUMETH.2019.05.006

Li, X., Xia, J., Ma, C., Chen, K., Xu, K., Zhang, J., et al. (2020). Accelerating structural degeneration in temporal regions and their effects on cognition in aging of MCI patients. *Cereb. Cortex* 30, 326–338. doi: 10.1093/cercor/bhz090

Liu, K., Chen, K., Yao, L., Guo, X., and the Alzheimer's Disease Neuroimaging Initiative (2017). Prediction of mild cognitive impairment conversion using a combination of independent component analysis and the cox model. *Front. Hum. Neurosci.* 11:33. doi: 10.3389/fnhum.2017.00033

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., and Evans, A. C. (2018). Structural neuroimaging as clinical predictor: a review of machine learning applications. *NeuroImage Clin.* 20, 506–522. doi: 10.1016/j.nicl.2018.08.019

Moradi, E., Hallikainen, I., Hänninen, T., Tohka, J., and the Alzheimer's Disease Neuroimaging Initiative (2017). Rey's auditory verbal learning test scores can be predicted from whole brain MRI in alzheimer's disease. *NeuroImage Clin.* 13, 415–427. doi: 10.1016/j.nicl.2016.12.011

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and the Alzheimer's Disease Neuroimaging Initiative (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002

Nozadi, S. H., Kadoury, S., and the Alzheimer's Disease Neuroimaging Initiative (2018). Classification of Alzheimer's and MCI patients from semantically parcelled PET images: a comparison between AV45 and FDG-PET. *Int. J. Biomed. Imaging* 2018:1247430. doi: 10.1155/2018/1247430

Ottoy, J., Niemantsverdriet, E., Verhaeghe, J., De Roeck, E., Struyfs, H., Somers, C., et al. (2019). Association of short-term cognitive decline and MCI-to-AD dementia conversion with CSF, MRI, amyloid- and 18F-FDG-PET imaging. *NeuroImage Clin.* 22:101771. doi: 10.1016/j.nicl.2019.101771

Pena, D., Barman, A., Suescun, J., Jiang, X., Schiess, M. C., Giancardo, L., et al. (2019). Quantifying neurodegenerative progression with DeepSymNet, an end-to-end data-driven approach. *Front. Neurosci.* 13:1053. doi: 10.3389/fnins.2019.01053

Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., et al. (2008). Prevalence of cognitive impairment without dementia in the United States. *Ann. Intern. Med.* 148, 427–434.

Qing, Z., Li, W., Nedelska, Z., Wu, W., Wang, F., Liu, R., et al. (2017). Spatial navigation impairment is associated with alterations in subcortical intrinsic activity in mild cognitive impairment: a resting-state fMRI study. *Behav. Neurol.* 2017:e6364314. doi: 10.1155/2017/6364314

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

Riederer, I., Bohn, K. P., Preibisch, C., Wiedemann, E., Zimmer, C., Alexopoulos, P., et al. (2018). Alzheimer disease and mild cognitive impairment: integrated pulsed arterial spin-labeling MRI and 18F-FDG PET. *Radiology* 288, 198–206. doi: 10.1148/radiol.2018170575

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer), doi: 10.1007/978-3-319-24574-4_28

Russo, M. J., Cohen, G., Campos, J., Martin, M. E., Clarens, M. F., Sabe, L., et al. (2017). Usefulness of discriminability and response bias indices for the evaluation of recognition memory in mild cognitive impairment and alzheimer disease. *Dement. Geriatr. Cogn. Disord.* 43, 1–14. doi: 10.1159/000452255

Sheth, S. A., Lopez-Rivera, V., Barman, A., Grotta, J. C., Yoo, A. J., Lee, S., et al. (2019). Machine learning-enabled automated determination of acute ischemic core from computed tomography angiography. *Stroke* 50, 3093–3100. doi: 10.1161/STROKEAHA.119.026189

Smith, G., Barret, F., Joo, J. H., Nassery, N., Savonenko, A., Sodums, D., et al. (2017). Molecular imaging of serotonin degeneration in mild cognitive impairment. *Neurobiol. Disord.* 105, 31–41. doi: 10.1016/j.physbeh.2017.03.040

Smith, L. N. (2017). "Cyclical learning rates for training neural networks," in *Procrrdings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (Piscataway, NJ: IEEE).

Spasov, S., Passamonti, L., Duggento, A., Liò, P., and Toschi, N. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *NeuroImage* 189, 276–287. doi: 10.1016/j.neuroimage.2019.01.031

Sun, Z., van de Giessen, M., Lelieveldt, B. P. F., and Staring, M. (2017). Detection of conversion from mild cognitive impairment to Alzheimer's disease using longitudinal brain MRI. *Front. Neuroinformatics* 11:16. doi: 10.3389/fninf.2017.00016

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE), doi: 10.1109/CVPR.2015.7298594

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., et al. (2017). A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Trans. Biomed. Eng.* 64, 155–165. doi: 10.1109/TBME.2016.2549363

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. (2017). "Residual attention network for image classification," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Honolulu, HI: IEEE), doi: 10.1109/CVPR.2017.683

Xie, C., Bai, F., Yu, H., Shi, Y., Yuan, Y., Gang, C., et al. (2012). Abnormal insula functional network is associated with episodic memory decline in amnestic mild cognitive impairment. *NeuroImage* 63, 320–327.

Xu, L., Yao, Z., Li, J., Lv, C., Zhang, H., and Hu, B. (2019). Sparse feature learning with label information for alzheimer's disease classification based on magnetic resonance imaging. *IEEE Spec. Sect. Emerg. Trends Issues Chall. Array Signal Process. Appl. Smart City* 7, 26157–26167. doi: 10.1109/ACCESS.2019.2894530

Yi, H. A., Möller, C., Dieleman, N., Bouwman, F. H., Barkhof, F., Scheltens, P., et al. (2016). Relation between subcortical grey matter atrophy and conversion from mild cognitive impairment to Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 87, 425–432. doi: 10.1136/jnnp-2014-309105

Young, A. L., Oxtoby, N. P., Daga, P., Cash, D. M., Fox, N. C., Ourselin, S., et al. (2014). A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* 137, 2564–2577. doi: 10.1093/brain/awu176