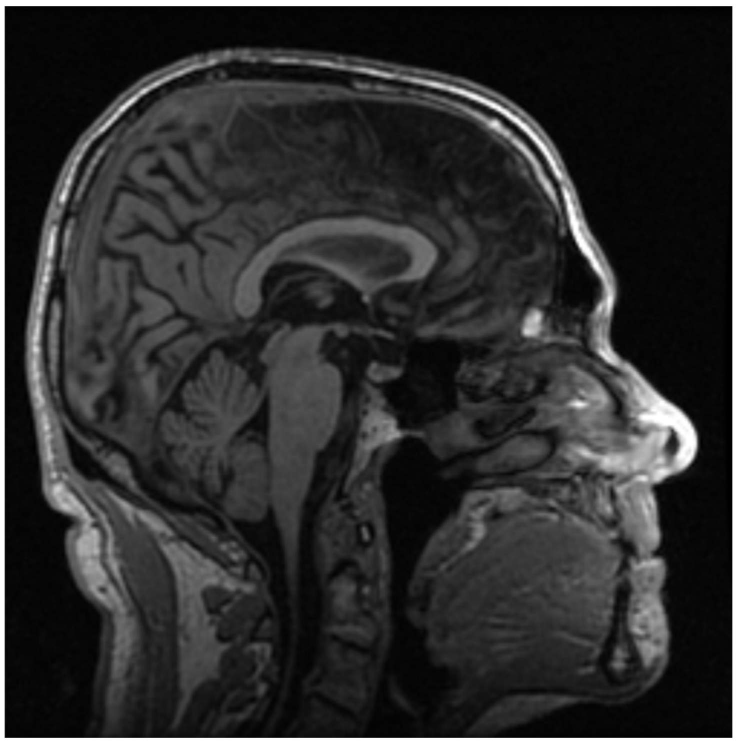
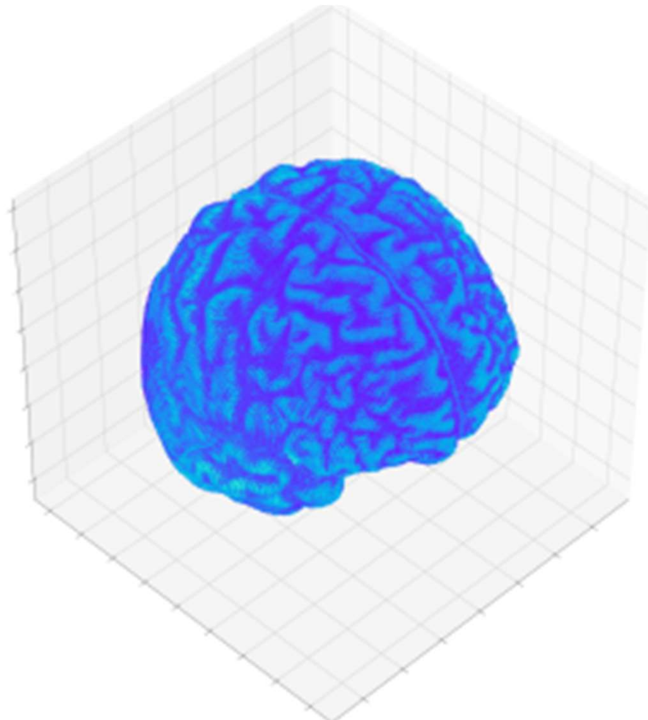
1.Data Preprocessing

We first perform segmentation for Raw MRI. DeepBrain is an open-source tool for segmenting brain raw MRIs. It is a custom U-Net model trained on a variety of manual-verified skull-stripping datasets. We chose DeepBrain because of its high-quality results. It achieves state-of-the art accuracy > 0.97 Dice metric on the test set that is compound with a subset of entries from the CC359 dataset, NFBS dataset and ADNI dataset.



Example of a slice of raw MRI



Example of a segmentation result

After segmentation, the dimensions of segments were different. We calculated the maxima dimensions, so we would need a (157, 213, 217) to hold all samples and it would be the input dimensions for the model. Previous studies would down sample the input dimension, otherwise it would be too computationally intensive for building a deep learning model on giant input data. We did not down sample the brain segments, keeping detail information would more beneficial than reducing the computation cost.

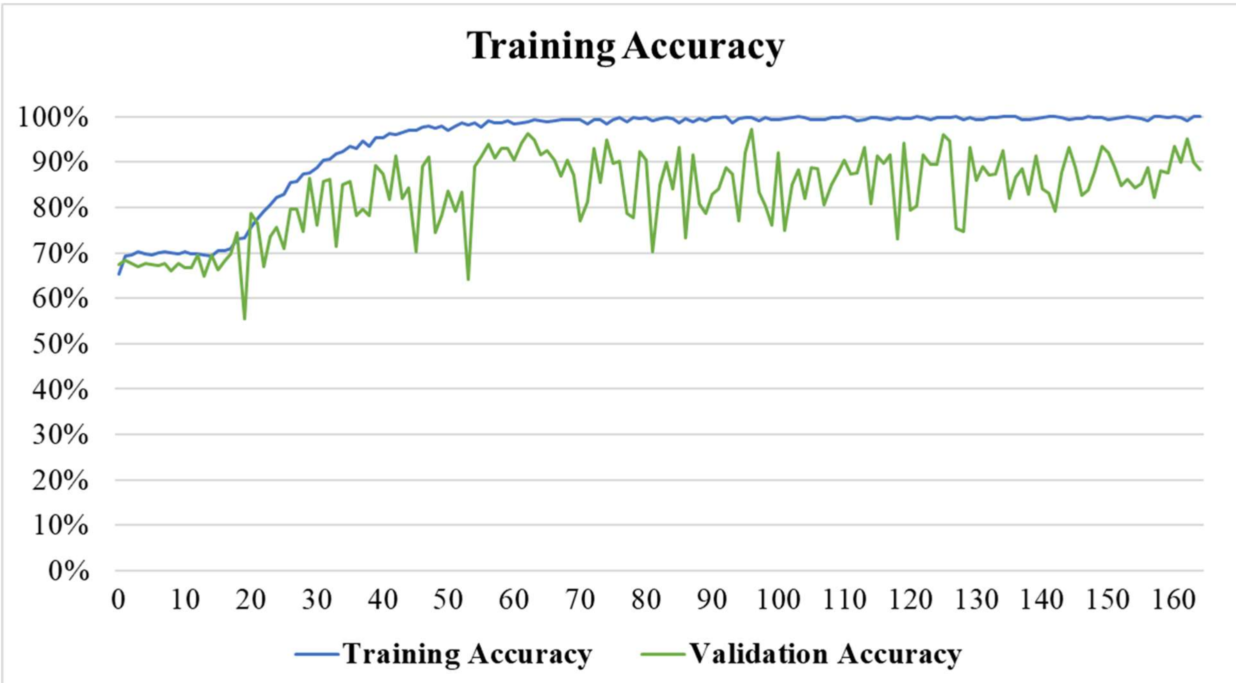
We applied Z-score normalization to all MRIs. The standard variance and mean were calculated from all MRI, the normalization scope was the whole dataset. Although we leveraged batch normalization (BN) layers in our model, we found it's still necessary to apply normalization before training, otherwise the model would not perform well.

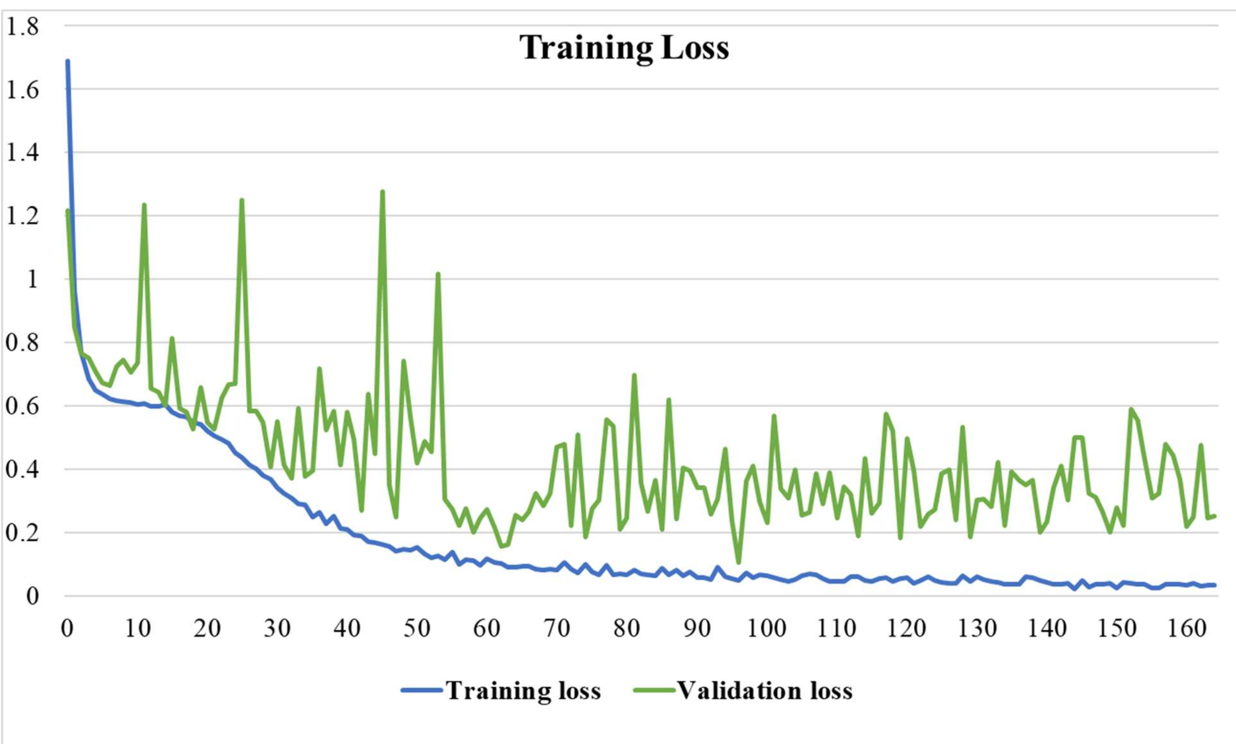
We did not perform other preprocessing such as registration, since we found it not necessary for our classification task. We even found brain segmentation has minor effects. The model can still reach nearly 90% accuracy on validation dataset without segmentation. It could be due to our advanced model design and large-scale training.

2.FPN training

We train Feature Pyramid Network (FPN) for 100-200 epochs for AD vs CN task and image feature extraction. We shifted the preprocessed images along x,y,z axis in 6 directions as our data augmentation. Time for 150 epochs of FPN training are around 25 hours with data augmentation. If we use DenseNet it would be 75-100 hours of training and it required down sampling, since limitation of the GPU memory. The PFN we developed is much efficient than DenseNet.

We show the FPN training curve blow. The accuracy on validation dataset on are well above 90%, the training accuracy was above 99.5%. Our FPN achieved of an accuracy of 96.5%, sensitivity of 95.8%, specificity of 97.0% on AD vs CN test set.





3.Related Results

3.1 Results of different fusion methods

We compared self-attention-based fusion with other commonly used fusion methods: fully connected layer and bi-linear pooling. We found self-attention is the most effective method. Bilinear pooling has positive effective if combined with fully connected layer, but it is just not effective as self-attentions. In our case, fully connected layer just did not work for all-stage fusion. Some non-deep-learning machine learning could be used, but there are just too many of them and comparing all of them would be beyond of the scope and aim of our paper.

	Accuracy	Precision	Recall	AUC
Attention	**90.50%**	**88.00%**	**95.70%**	**91.47%**
Attention + bilinear pooling	88.10%	88.00%	91.67%	92.47%
Fully connected layer + bilinear pooling	83.33%	84.00%	87.50%	92.47%
Fully connected layer	40.00%	0	0	37.88%

3.2 Results of different inputs

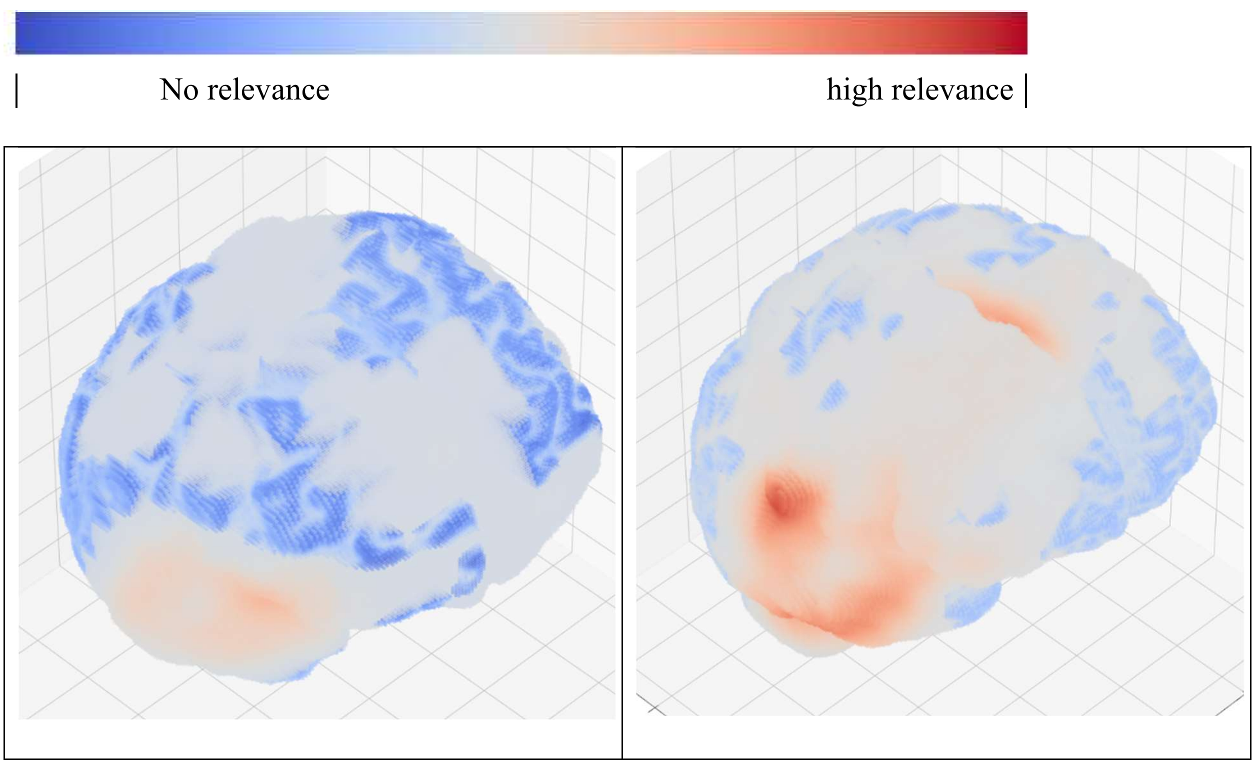
We also used different combination of inputs. Since we found blood test had lower attention weights in visualization, we removed blood test input to see the impact on the predication results.
The accuracy is 85.7% if we remove plasma-tau181, which is good compared with only using image features. Using all kinds of input had better results for just using a subset of available kinds of inputs. The results also showed combine different modality is necessary since image only method had poor results.

	Accuracy	Precision	Recall	AUC
Image feature + P-Tau181, Age, gender, APOE	**90.50%**	**88.00%**	**95.70%**	**91.47%**
Image feature + Age, gender, APOE	85.7%	88.0%	82.4%	92.7%
Image feature + P-Tau181	85.7%	88.0%	82.4%	93.0%
IF(image feature)	69.7%	47.9%	86.9%	79.3%

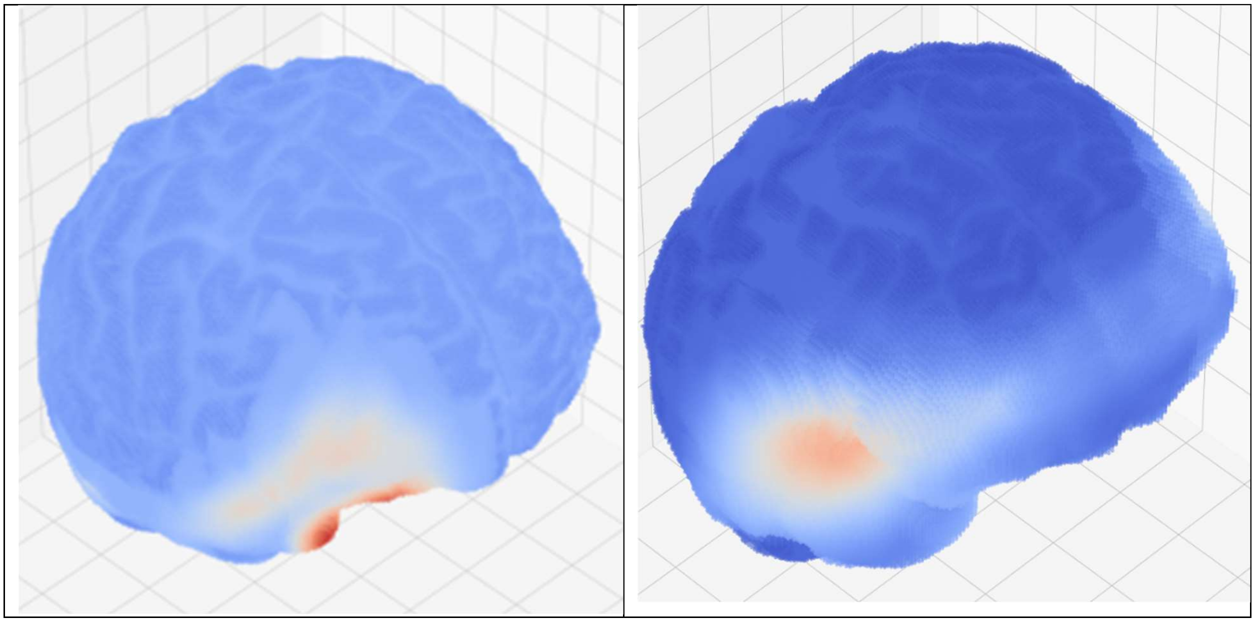
3.3 AD activation map:

Below shows 6 activation heatmaps of AD, MCI, CN case. We did not perform quantitively analysis of the activation map, the CN cases had much less AD related area than AD in general. MCI activation map could be look like AD or CN, so MCI are hard sample for image model. We can also find AD relevance area in most CN MRIs; it could be brain atrophy due to normal aging and samples in ADNI dataset are most elderly people.

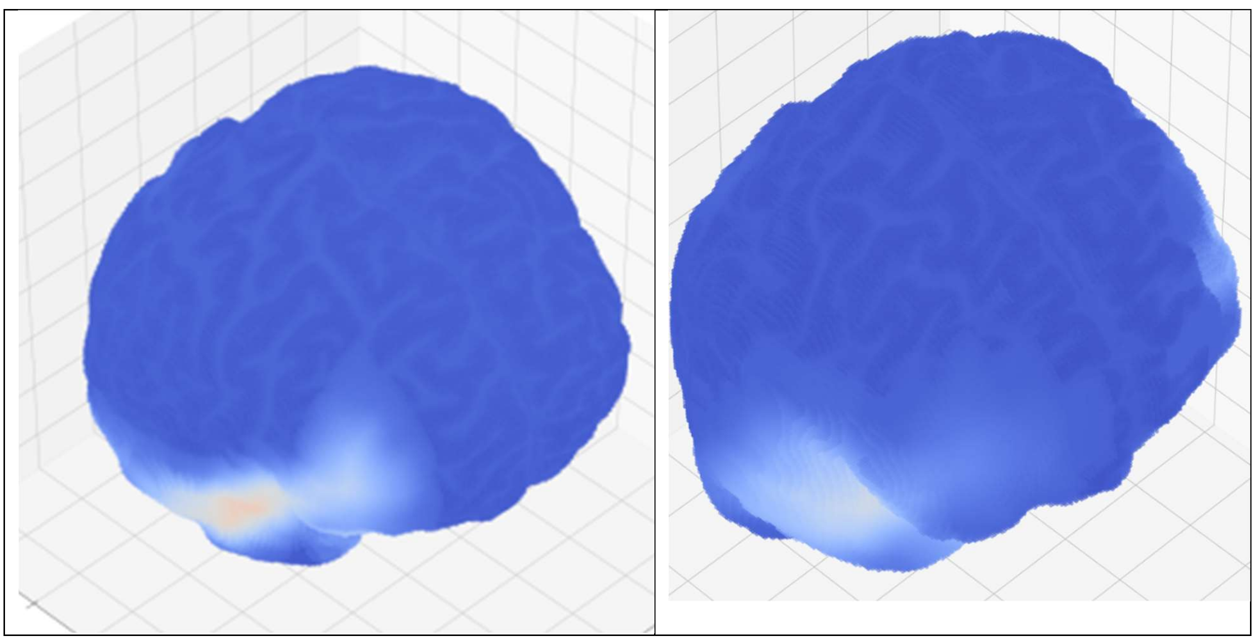
Activation map could explain why image feature only method is hard to predict MCI from CN. And it supports our idea that using other modalities are necessary.



Activation map on two AD MRIs



Activation map on two MCI MRIs



Activation map on two CN MRIs

4. Code availability

Model implementation is publicly available on GitHub. We only uploaded the core model code, since other code such as preprocessing, merging data from different modality, MRI data loader are complex, lack of readability and hard to configure environment.

Link: https://github.com/Qinyong-Wang/AANet-Attentive-All-level-Fusion-for-Medical-Images