Statistics in Biopharmaceutical Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usbr20

Deep Learning Based Multimodal Progression Modeling for Alzheimer's Disease

Liuqing Yang, Xifeng Wang, Qi Guo, Scott Gladstein, Dustin Wooten, Tengfei Li, Weining Z Robieson, Yan Sun, Xin Huang & for the Alzheimer's Disease Neuroimaging Initiative

To cite this article: Liuqing Yang, Xifeng Wang, Qi Guo, Scott Gladstein, Dustin Wooten, Tengfei Li, Weining Z Robieson, Yan Sun, Xin Huang & for the Alzheimer's Disease Neuroimaging Initiative (2021) Deep Learning Based Multimodal Progression Modeling for Alzheimer's Disease, Statistics in Biopharmaceutical Research, 13:3, 337-343, DOI: 10.1080/19466315.2021.1884129

To link to this article: https://doi.org/10.1080/19466315.2021.1884129

Published online: 10 Mar 2021.

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2021, VOL. 13, NO. 3, 337–343
https://doi.org/10.1080/19466315.2021.1884129

Deep Learning Based Multimodal Progression Modeling for Alzheimer's Disease

Liuqing Yang[a], Xifeng Wang[b], Qi Guo[a], Scott Gladstein[a], Dustin Wooten[a], Tengfei Li[c], Weining Z Robieson[a], Yan Sun[a], Xin Huang[a], and for the Alzheimer's Disease Neuroimaging Initiative*

[a]AbbVie, Inc., North Chicago, IL; [b]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; [c]Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

ABSTRACT

The progression of Alzheimer's disease (AD) is a continuous process in cognitive and biomarker changes, with only a fraction of mild cognitive impairment (MCI) patients eventually advancing to AD. There is no definite biomarker signature to determine this eventual progress. The discovery and development of prognostic biomarker signatures for AD is essential to address the challenges of AD drug discovery and development. The deep learning (DL) technique is a recent breakthrough in data science that enables researchers to discover previously unknown features comprised by complicated patterns learnt from large datasets. It has outperformed many traditional machine learning methods in computer vision tasks. In this article, we evaluated the performance of DL algorithms in differentiating patients with diagnosis of AD, MCI, or no evidence of dementia using baseline MRI data of the brain, and integrated features extracted from the neural network with other baseline biomarkers to develop an AD prognostic signature. This signature can be used to understand the patient heterogeneity in a study cohort and provide enrichment strategies for AD clinical trial design.

ARTICLE HISTORY
Received December 2019
Accepted January 2021

KEYWORDS
Prognostic signature; Structural magnetic resonance imaging; Super Learner; Transfer learning

1. Introduction

As of 2019, Alzheimer's disease (AD) affects over 5.8 million Americans: one out of ten people over 65 years old, and one out of three people over 85 years old have Alzheimer's dementia (Gaugler et al. 2019). As a growing epidemic, the total number of Americans with this disease is estimated to be 13.8 million by 2060 and is projected to nearly double every 20 years worldwide (Matthews et al. 2019; WHO 2019). With scientific progress and better understanding of the disease, the definition of AD has also been changing. In 2011, the National Institute on Aging and Alzheimer's Association created the diagnostic guidelines to separate AD status into three stages: preclinical, mild cognitive impairment (MCI), and Alzheimer's dementia. In 2018, the disease definition was updated as a continuous process in cognitive and biomarker domains, and a research framework was proposed based on the updated definition of AD: abnormal biomarker changes in both β-amyloid (Aβ) deposition and pathologic tau (Jack et al. 2018). It is worth noting that the progression of these biomarkers begins even before symptoms (Bateman et al. 2012). The continued development of biomarkers for AD is expected to better characterize the spectrum of the disease and to facilitate better prevention, diagnosis and treatment of the disease.

Since AD is an irreversible disease, current treatment strategies have been focusing on disease-modifying therapies (DMTs), which aim to slow down the progression of the disease or delay its onset, by interfering with the fundamental pathophysiological progression of the disease (Cummings and Fox 2017). Prognostic biomarkers such as the cerebrospinal fluid (CSF) analytes Aβ 1–42, Total-Tau, and Phospho-Tau, and hippocampal atrophy measured using magnetic resonance imaging (MRI), are investigated to reflect the disease progression for enrichment in AD clinical trials (Counts et al. 2017). However, it remains a challenge to develop prognostic biomarker signatures for early stage patients by integrating biomarkers from different sources.

Imaging biomarkers measuring patient progression to AD, have been contributing to preclinical detection of AD or AD intervention (Cummings, Ritter, and Zhong 2018; Ledig et al. 2018). Recent developments in convolutional neural network algorithms have helped medical researchers optimize their medical imaging applications (Saltz et al. 2018). One of the advantages of this state-of-the-art methodology is its ability to discover previously unknown features comprised of complicated patterns learnt from large datasets. The primary goal of this work is to develop deep learning (DL)-based imaging biomarker signatures from structural MRIs and combine them with demographic, genetic, and cognitive markers to predict the progression of MCI to AD.

2. Materials and Methods

2.1. Participants and Data

Data used in this work was downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database

CONTACT Xin Huang xin.huang@abbvie.com AbbVie, Inc., North Chicago, IL 60064.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: *http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*.

Table 1. Demographic, genetic and cognitive information at baseline.

Study	Diagnosis	Number of subjects	Age	Males/ females	Education (yrs.)	APOE4 (0/1/2)	FAQ	CDRSB	MMSE	ADAS13
ADNI1, ADNI2, ADNIGO	pMCI	230	74.0 ± 7.1	136/94	15.8 ± 2.8	77/114/39	5.6 ± 5.0	2.0 ± 1.0	26.9 ± 1.8	21.2 ± 5.8
	sMCI	335	71.9 ± 7.5	200/135	16.1 ± 2.8	210/105/20	1.7 ± 2.9	1.2 ± 0.7	28.0 ± 1.6	13.4 ± 5.7
	AD	270	74.9 ± 7.7	147/123	15.2 ± 3.0					
	CN	422	74.5 ± 5.8	211/211	16.4 ± 2.7					
J-ADNI	AD	147	73.9 ± 6.6	63/84	12.5 ± 3.2					
	CN	139	68.1 ± 5.4	66/73	13.8 ± 2.8					

NOTE: Subjects at baseline are diagnosed as Alzheimer's disease (AD), control normal (CN), or mild cognitive impairment (MCI). The categorization of progressive MCI (pMCI) and stable MCI (sMCI) depends on whether the subject converts to AD within 36 months after the baseline visit. Means and standard deviations of different biomarkers are summarized in the table. APOE4 represents the APOE $\varepsilon 4$ allele carrier status and four cognitive assessment results measuring cognitive dysfunction are included: Functional Activities Questionnaire (FAQ), Clinical Dementia Rating Sum of Boxes (CDR-SB), Mini-Mental State Examination (MMSE), and Alzheimer's Disease Assessment Scale 13-item Subscale (ADAS13).

(*https://adni.loni.ucla.edu*) and the Japanese ADNI (J-ADNI) data from the National Bioscience Database Center (*https://humandbs.biosciencedbc.jp/en/hum0043-v1*). ADNI was launched in 2004 at 57 clinical sites in the United States. Japanese ADNI(J-ADNI) was launched in 2007–2008 and completed in 2014. Both studies have a goal to develop biomarkers for the early detection and monitoring of AD. Study data include demographics, clinical, psychometric, imaging, and genetic data. The variables we used from these two databases in this study are summarized in Table 1, as well as the raw T1-weighted MRI images for each subject. All of the variables we used are baseline measurements.

An overview of our proposed framework is illustrated in Figure 1. There are two major parts in this framework. In part 1, pretrained DL models are applied on preprocessed structural T1 MR images from AD and control normal (CN) patients to extract AD specific features to serve as direct input for part 2; in part 2, a stacking ensemble learning algorithm is implemented, which integrates the derived AD specific features with other baseline biomarkers to predict if an MCI patient will progress from MCI to AD (pMCI) within 36 months or will remain stable (sMCI). If longitudinal diagnosis information is available for training data, a penalized Cox proportional hazards model is used to predict an MCI subject's risk of progression to AD.

2.2. Image Preprocessing

The image preprocessing procedure uses the ANTs package (Avants et al. 2011). Raw T1 images are first resampled to uniform resolution with voxel dimension 1x1x1 mm followed by an N4 bias field correction. Next, each image goes through a two-step transformation process to co-register the subject brain into a common T1 template OASIS30-Atropos. First, the subject brain undergoes linear (affine) co-registration into standard space which includes translation, rotation, stretching and shearing in all dimensions. Second, the affine co-registered brain undergoes nonlinear co-registration which includes local expansion and contraction of tissue in an attempt to make the subject brain look like the standard brain. Products of the registration relevant to this work that were used in the model framework include the registered T1 image with dimension $216 \times 256 \times 291$, as well as the Log Jacobian (LJ) maps with the same dimensions. LJ maps of the deformational field contain information related to the local tissue expansion and contraction. Specifically, the LJ maps comprise information about differences in local tissue volume relative to the standard brain which is thus directly related to atrophy. For example, a voxel value less than 0 indicates there is local atrophy, whereas a value greater than 0 indicates an increase in tissue volume relative to the standard brain. Registration of all images in this study took approximately 1800 hr of CPU time on a high-performance parallel computing cluster. More details of the processing steps by ANTs were detailed in Tustison et al. (2014).

2.3. DL-Based MRI Biomarker Signature Development

DL has achieved superior results in extracting useful features from large amount of unstructured data such as images compared with traditional machine learning techniques. However, application of this technology in medical images is often limited by the lack of well-annotated data. To address this challenge, one can use transfer learning, which is a technique of tuning pretrained DL models, to solved specific problem of interest based on the limited dataset. Transfer learning has previously been shown to successfully extract useful features from medical images such as MRIs and histopathology images (Hon and Khan 2017; Coudray et al. 2018). In this work, we explored popular pretrained deep neural network architectures including ResNet, Inception and VGG series, and found that tuning the VGG16 network (Simonyan and Zisserman 2014) returns the best classification accuracy for our training task in part 1 (constructing classifier for AD vs. CN). The features extracted by the VGG16 network can be used as inputs for part 2 (developing prognostic signature to predict the progression from MCI to AD) as shown in Figure 1. Due to the continuous spectrum of the AD progression with MCI being an intermediate stage between CN and AD, features extracted from the VGG16 network (trained to classify AD vs. CN) may be important in differentiating pMCI from sMCI. This idea has been tested and shown to be effective in several studies (Moradi et al. 2015; Tong et al. 2017; Lin et al. 2018). Hence, our DL model is designed to be trained with CN and AD images, then applied to MCI images for the prognostic signature development.

Remark. The VGG16 network was originally trained for object recognition on two-dimensional natural images from the ImageNet dataset. After feature extraction, to implement the transfer learning, two fully connected layers with ReLu activation

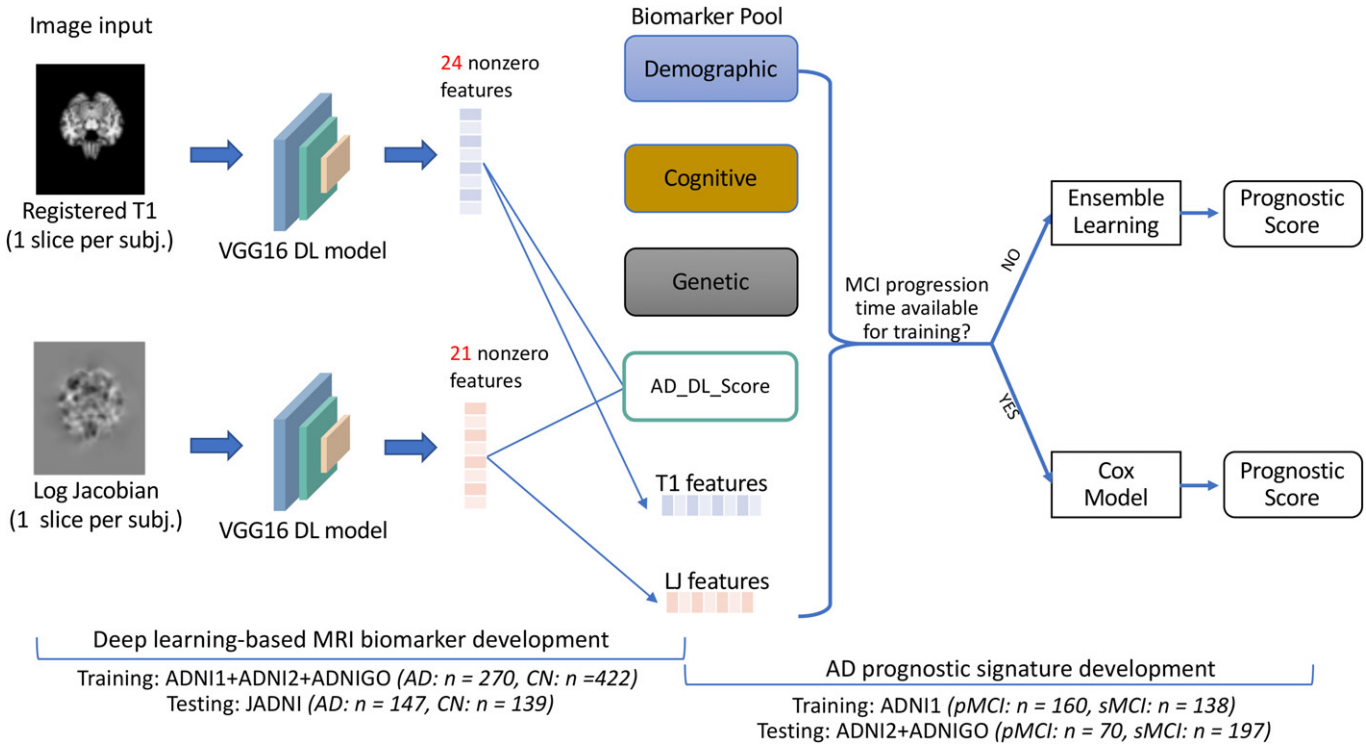

Figure 1. Framework overview. Part 1 (left panel) of the framework takes processed MRI maps from MCI patients as the inputs for two VGG16 deep learning models to generate the DL-based biomarker signatures. Parameters in this part were trained using AD versus CN images from the ADNI studies. Part 2 (right panel) of the framework combines the features extracted from deep learning models with other clinical biomarkers to generate the final prognostic score for MCI patients. Parameters in this part were trained using data from pMCI and sMCI patients in ADNI1, which is the first completed ADNI study. Two modeling choices are implemented depending on the availability of MCI to AD time-to-progression outcome in the training data.

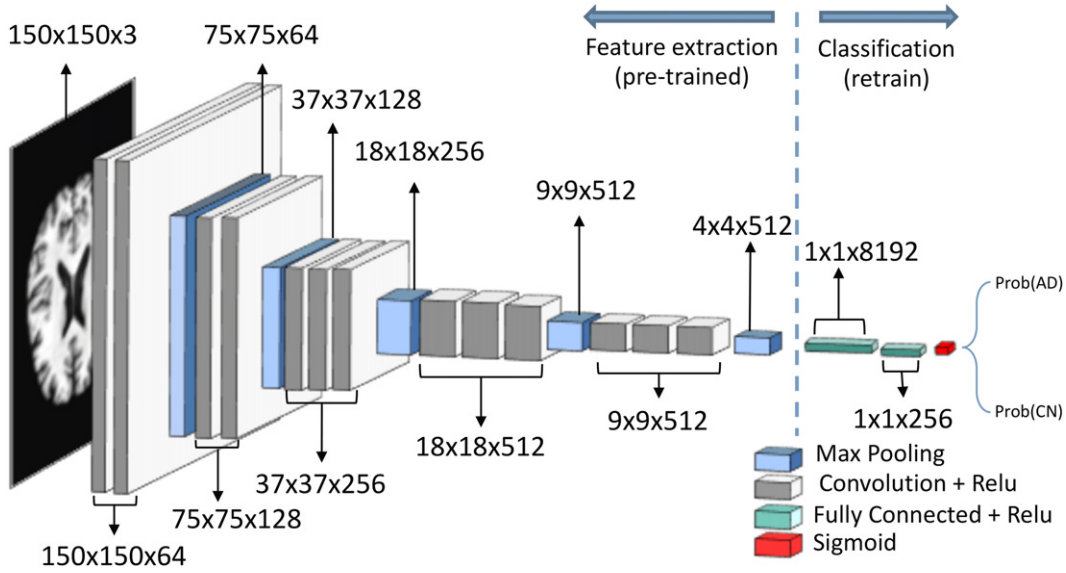

Figure 2. Transfer learning framework based on VGG16.

function were added to the network, followed by a dropout layer with rate 0.5 to prevent overfitting and finally output the probability of AD with a sigmoid classifier, with binary cross-entropy as the loss function (Figure 2). These newly added layers were retrained specifically to our classification task in part 1 based on the training dataset. The computation was performed using the RMSProp optimization technique (Hinton, Srivastava, and Swersky 2012) and the model was trained on NVIDIA V100 GPUs using Tensorflow.

2.4. Selection of Axial Two-Dimensional Slice

In our preliminary study, 50 consecutive axial two-dimensional slices taken from the middle part of each subject's three-dimensional MRI scans were used to train the DL model. This part has been shown to contain the most useful brain structures in differentiating AD versus CN such as the ventricles, hippocampus, and amygdala (Ledig et al. 2018). ADNI AD and CN images were used with an 80/20 training/testing split at

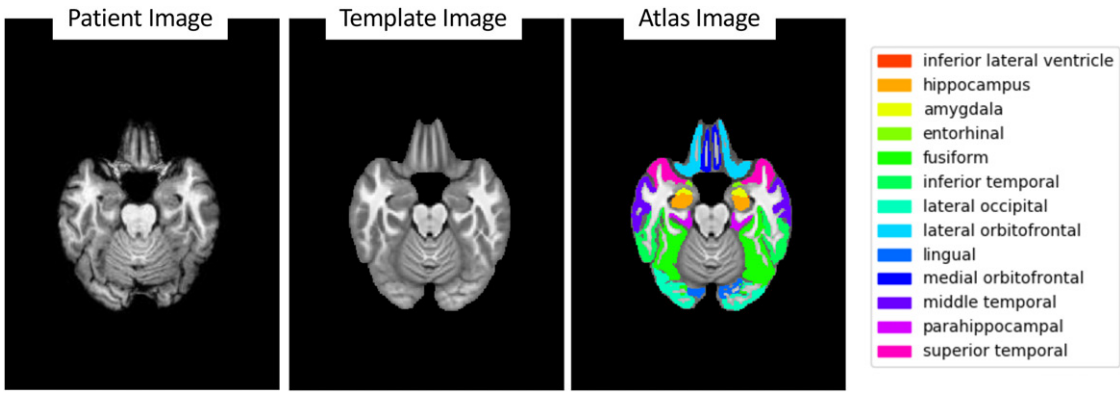

Figure 3. From left to right: (1) an example registered T1 image from an ADNI AD patient shows the slice selected as the input for our deep learning model; (2) the corresponding slice from the registration template OASIS30-Atropos; (3) atlas-based segmentation labels on the template image shows the brain regions present in the selected slice. OASIS-TRT-20 joint fusion atlas was used for labeling. Details can be found in *https://mindboggle.info/data.html*.

the subject level to avoid any data leakage. Results from this preliminary study (not shown) suggest the prediction probabilities from these 50 slices are highly correlated. Together with the work in Valliani and Soni (2017), where the authors used a single slice from the median of axial brain per subject to train a transfer learning neural network and received reasonable accuracy (78.8%) in differentiating AD versus CN, we decide to construct a single-slice DL model instead of using 50 slices per subject and directly use the prediction for the single slice as the prediction for the subject. The optimal slice was selected by training 50 single-slice VGG16 models and choosing the slice corresponding to the model with the highest validation accuracy. The nonlinear transformation we used to register the three-dimensional scans ensures the chosen slice contains the same anatomical tissues across all subjects. Valliani and Soni (2017) implemented a pretrained ResNet model with 417 axial brain slices, gaining 78.8% prediction accuracy on the testing set. In comparison, our work uses a pretrained VGG16 model with 692 single axial brain slices and obtains 82.9% accuracy on the testing set. The difference of performance may be affected by the choice of DL models, sample sizes and/or the choice of slice. The slice we selected through modeling not only gives reasonable prediction accuracy, but also contains structures related to the disease progression (Figure 3). The segmentation results in Figure 3 suggest the selected slice captures brain regions such as hippocampus and amygdala, whose atrophy has shown to be associated with AD (Jack et al. 2010; Klein-Koerkamp et al. 2014).

2.5. LJ Maps

All raw MRI scans in our dataset went through a nonlinear registration process as mentioned in the imaging preprocessing section. The nonlinear registration in our raw MRI scan preprocess ensures all MRI images are registered to a common template, but it also subjects to a concern about information loss as raised by Tong et al. (2017). The lost information is captured by LJ maps which contains information of the deformation field of the nonlinear image registration. Hence, adding LJ maps as the image input of the DL model may improve the model performance. To achieve this, we built three different DL models taking registered T1, LJ maps, or both types as the input images.

Table 2. Prediction performance of DL models using different image sources as input.

Image source	AUC	ACC*	SEN*	SPE*
T1	0.887	0.822	0.762	0.885
LJ	0.814	0.741	0.728	0.755
T1 and LJ	0.908	0.825	0.755	0.899

NOTE: Performance comparison of AD scores generated from different image sources: registered T1, LJ maps or both in differentiating AD versus CN using J-ADNI data. (ACC = accuracy, SEN = sensitivity, SPE = specificity).
*ACC, SEN, and SPE were computed based on a 0.5 probability cutoff.

These models were trained and validated on ADNI's AD/CN data and tested on an independent study J-ADNI, which has similar inclusion criteria and image acquisition parameters. Results in Table 2 suggests the adding LJ map does improve the overall model performance (AUC improved from 0.887 to 0.908). Thus, our final model combines both image types as the input.

As shown in Figure 1, slices of registered T1 images and LJ maps are fed into two separate VGG16 transfer learning models. Due to the dropout regularization and the fact that part of the model is fixed and pretrained on a nonmedical image dataset, many features from the last fully connected layer have zero or small weights. As a result, 24 nonzero features from T1 images and 21 nonzero features from LJ images are selected from the last dense layer. These features can be directly combined with other biomarkers or be used to develop the "AD_DL_Score." To generate the AD_DL_Score, we use the ADNI data, 270 AD subjects and 422 CN subjects, to train a penalized logistic regression model, where the AD/CN labels are the responses and the VGG16 extracted T1 and LJ features are the predictors. A 10-fold cross-validation is applied to tune the model parameters. A larger AD_DL_Score value indicates a higher probability of being an AD MRI. This AD_DL_Score can also be combined with other biomarkers to develop the final AD prognostic signature as discussed in detail in the next section.

2.6. AD Prognostic Signature Development

It has been shown in several studies that combining data from different modalities has the potential to improve pMCI versus sMCI prediction accuracy (Korolev et al. 2016; Tong et al. 2017;

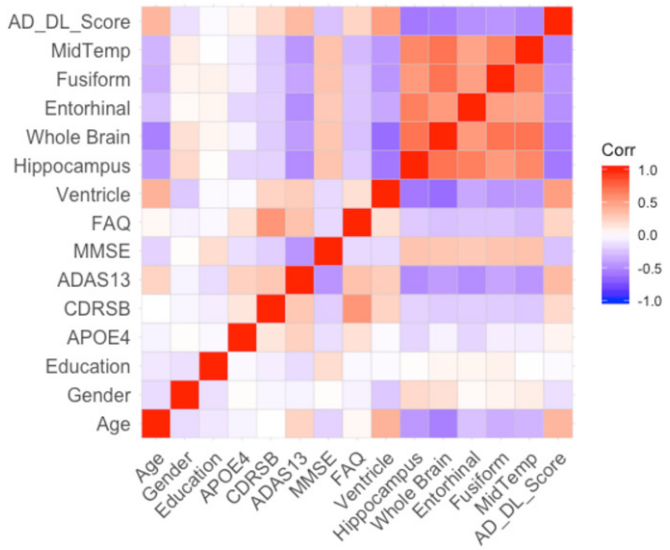

Figure 4. Pairwise Pearson correlation map of biomarkers. Featuring the AD_DL_Score with other eight biomarkers and six structural volume measures. The structural volumes are proportionally normalized by intracranial volume.

Spasov et al. 2019). To develop the AD prognostic signature, we plan to combine newly discovered features extracted by the trained DL model with biomarkers from other sources. We first explore the association between the AD_DL_Score and other biomarkers. We expect this image biomarker signature to correlate with the volume estimates provided by ADNI using the FreeSurfer image analysis suite (Reuter et al. 2012). Figure 4 shows the correlation between the AD_DL_Score and other available biomarkers for MCI patients in the ADNI dataset. The correlation map shows the score is well correlated with the volume of several regions of interest related to AD progression (Ledig et al. 2018). In particular, it is negatively correlated with the hippocampus volume (Pearson $\rho = -0.58$) and positively correlated with ventricle volume ($\rho = 0.5$). All these volumes were normalized by intracranial volume. The DL derived score is also positively correlated with the Alzheimer's Disease Assessment Scale-Cognitive subscale with 13 items (ADAS13) ($\rho = 0.36$), a composite score measuring overall cognition. Note that higher ADAS13 scores represent worse cognitive function. These associations validate the interpretability of the AD_DL_Score. The fact that AD_DL_Score is not highly correlated with any nonimage biomarkers (all absolute correlation coefficient values are less than 0.38), suggests its potential of providing additional information to the progression prediction.

With different types of available biomarker signatures, we propose two approaches for AD prognostic signature development based on different choice of endpoints. In the first approach, we aim at predicting the probability of an MCI patient progressing from MCI to AD (pMCI) within the next 3 years, while in the second approach, we aim at developing a prognostic score that is concordant with the time to AD progression from baseline. For the first approach, since it is a binary classification problem, we propose to model it with Super Learner, an ensemble learning method developed by van der Laan, Polley, and Hubbard (2007). Super Learner is a cross-validation based optimization framework, which finds the optimal weights of a collection of base learner algorithms by minimizing the overall cross-validation errors:

$$\text{argmin}_{\vec{w}} \sum_{v=1}^{V} \sum_{i=1}^{N_v} l\left(y_i, \sum_{j=1}^{m} w_j \widehat{y_{ij}}^{-v}\right),$$

where w_j is the weight of the jth base learner out of a total of m algorithms and subjects to a standard simplex:

$$\sum_{j=1}^{m} w_j = 1, w_j \geq 0, \quad \text{for } j = 1, \ldots, m$$

and V is the V-fold cross-validation, N_v is the number of subjects in the vth fold, y_i is the true pMCI/sMCI label of the ith subject and $\widehat{y_{ij}}^{-v}$ is the prediction of the ith subject by the jth base learner trained on the whole data except for the vth fold. The loss function l evaluates the cross-validation error between the prediction and the truth. In our study, l uses the rank loss, whose minimization is a proxy to maximizing the prediction AUC.

The ensemble learning model is first trained on 298 ADNI1 MCI subjects using all available biomarker signatures including the AD_DL_Score and then tested on 267 MCI subjects from ADNI2 and ADNI-GO. Four base learner algorithms are used in our dataset: gradient boosting machine (XGBoost), elastic net (GLMNET), recursive partitioning (RPART), and generalized additive models (GAM), whose weights are shown in Figure 5(a). XGBoost receives the highest weight under the 10-fold cross-validation run. Figure 5(b) shows the important features selected by this base algorithm using "gain" measure in the xgboost R package (Chen et al. 2019). Gain represents the approximate loss function gain based on each feature's splits. Higher value in the plot indicates a more important predictive feature in building the decision tree. The ADAS13 cognitive score ranks the highest and the image biomarker signature we developed (AD_DL_Score) ranks the fourth among all nine biomarkers, higher than some of the cognitive scores (CDRSB, MMSE).

For the second approach, we fit a penalized Cox proportional hazards model when longitudinal diagnosis information is available for MCI patients. (For ADNI dataset, MCI patients have follow-up visits scheduled every 6 months for 3 years.) Each conversion to AD from MCI is considered as an event. Similar to the ensemble learning model, all the baseline biomarker signatures are used as the covariates. The relative risk score predicted by the Cox model is regarded as the prognostic score, which describes an MCI subject's risk of progression to AD. This prognostic score derived from the Cox model is evaluated by AUC in differentiating pMCI versus sMCI summarized in Table 3.

We are interested in knowing how different subsets of biomarkers contribute to the progression prediction. AUC performance using different combinations of biomarkers as predictors are explored under two modeling strategies. The second row of Table 3 suggests adding the DL-derived features improves prediction performance from the range of 0.6–0.7, which is a promising gain by adding an objective measure. Adding cognitive biomarkers boosts the performance further from the range of 0.7–0.9. However, this strong enhancement contributed by the cognitive biomarkers is expected because the final diagnosis of dementia is based on these cognitive measures. The last row

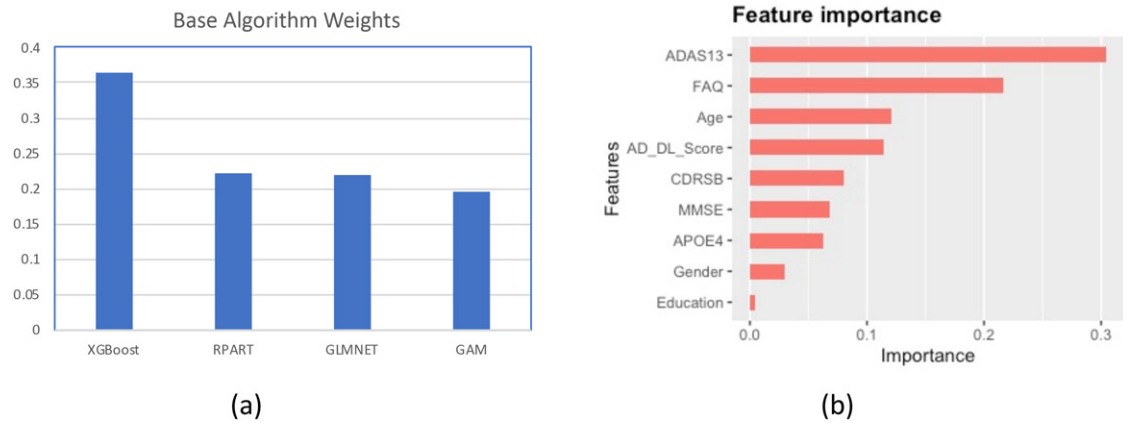

Figure 5. (a) Weight distribution among four base learner algorithms in ensemble learning using all biomarker signatures. (b) Feature importance selected by XGBoost.

Table 3. Prognostic score performance in differentiating pMCI versus sMCI in ADNI2 and ADNI-GO under different modeling or biomarker choices

Biomarker set	AUC	
	Ensemble learning	Cox model
Demographic + genetic	0.665	0.657
Demographic + genetic + AD_DL_Score/45 DL features	0.750/0.762	0.733/0.767
Demographic + genetic + cognitive	0.904	0.913
Demographic + genetic + cognitive + AD_DL_Score/45 DL features	0.915/0.921	0.922/0.921

of Table 3 shows the added value of image biomarker signatures with the presence of all other biomarkers, improving prediction performance from 0.90 to 0.92 in the ensemble learning framework. Spasov et al. (2019) developed a similar multimodal DL framework using three-dimensional MRI images, Jacobian images and clinical data for pMCI versus sMCI discrimination. Our results (last row of Table 3) are very similar to theirs (AUC = 0.917), though we only use a single two-dimensional slice transfer learning model, which is more computationally efficient. The comparison of performance between using the original 45 T1 and LJ combined DL features and using a single AD_DL_Score suggests that the former set may provide richer information for modeling progression. Specifically, combining the 45 DL-features directly into the ensemble learning may provide flexibility beyond linear modeling and enable the algorithm to search for nonlinear association between predictors and response.

3. Discussion

The growing number of AD patients creates a burden to both the patients' families and the society. The development of effective AD treatment is challenging as we have seen many failed clinical trials (Mehta et al. 2017). One of the most important factors in a successful DMT drug design is enrichment with patients of fast progression, so that the clinical trial can detect the treatment slow-down on disease progression within short period of time (typically 24–36 months). The key element of this enrichment design is to construct an effective prognostic signature reflecting the disease progression. In this article, we have used DL to extract the AD features from structural MRI images and combined them with other biomarkers to construct an effective prognostic signature. We demonstrate a good performance using this prognostic signature in predicting patient who will progress from MCI to AD within 36 months.

Considering the continuous spectrum of the disease, there can be a caveat to use sMCI and pMCI as the endpoint to develop the prognostic score. In the ADNI dataset, MCI patients' progression time to AD were recorded in a 6-month unit as an approximation to the exact time of progression to AD. We observed an improvement of prediction performance when using a Cox model on the time-to-event data in Table 3. This result suggests dichotomizing the continuous disease may cause information loss.

We have also observed that cognitive scores play an important role in the prediction performance of prognostic scores. However, there are caveats for a prognostic score developed with the component of cognitive scores. First, in clinical practice the disease is diagnosed and staged by cognitive scores (Sheehan 2012). Second, the scoring methods in the clinic are subjective and vary from physician to physician (Tractenberg, Schafer, and Morris 2001). To develop an objective prognostic signature, one of our future pieces of work is to train machine learning models using baseline cognitive scores or change from baseline as the endpoint instead of predictors. Another advantage of using cognitive scores as endpoints is that it may reflect the continuous spectrum of the AD disease progression, which may enhance the prediction performance.

Other future work is to explore different neural networks, and test 2.5-dimensional (Lin et al. 2018) or three-dimensional images as the image input. The combination of three-dimensional images and longitudinal changes can also be explored. Studies (Ewers et al. 2011; Hinrichs et al. 2011) have also shown the improvement of prediction by including other imaging modalities, such as fMRI, FDG-PET, DTI into the modeling.

With the different combinations of input and endpoints mentioned above, we can use the proposed framework described in this article to explore the optimal prognostic signatures for AD progression. The final prognostic signature can be integrated to the clinical trial development for patient enrichment design, and it also can be integrated to physician clinical workflow to facilitate decision making in the clinical practice.

Acknowledgments

We thank Dr. Yanping Luo, Dr. Xiaomeng Zhang, Dr. Deli Wang, and Dr. Hui Zheng from AbbVie Inc. for useful discussions about this work.

References

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011), "A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration," *Neuroimage*, 54, 2033–2044. [338]

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., and Holtzman, D. M. (2012), "Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease," *New England Journal of Medicine*, 367, 795–804. [337]

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., and Li, Y. (2019), "xgboost: Extreme Gradient Boosting," R Package Version 0.90.0.2. [341]

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A. L., Razavian, N., and Tsirigos, A. (2018), "Classification and Mutation Prediction From Non-Small Cell Lung Cancer Histopathology Images Using Deep Learning," *Nature Medicine*, 24, 1559–1567. [338]

Counts, S. E., Ikonomovic, M. D., Mercado, N., Vega, I. E., and Mufson, E. J. (2017), "Biomarkers for the Early Detection and Progression of Alzheimer's Disease," *Neurotherapeutics*, 14, 35–53. [337]

Cummings, J., and Fox, N. (2017), "Defining Disease Modifying Therapy for Alzheimer's Disease," *The Journal of Prevention of Alzheimer's Disease*, 4, 109. [337]

Cummings, J., Ritter, A., and Zhong, K. (2018), "Clinical Trials for Disease-Modifying Therapies in Alzheimer's Disease: A Primer, Lessons Learned, and a Blueprint for the Future," *Journal of Alzheimer's Disease*, 64, S3–S22. [337]

Ewers, M., Frisoni, G. B., Teipel, S. J., Grinberg, L. T., Amaro, E., Jr, Heinsen, H., Thompson, P. M., and Hampel, H. (2011), "Staging Alzheimer's Disease Progression With Multimodality Neuroimaging," *Progress in Neurobiology*, 95, 535–546. [342]

Gaugler, J., James, B., Johnson, T., Marin, A., and Weuve, J. (2019), "2019 Alzheimer's Disease Facts and Figures," *Alzheimers & Dementia*, 15, 321–387. [337]

Hinrichs, C., Singh, V., Xu, G., Johnson, S. C., and Alzheimers Disease Neuroimaging Initiative (2011), "Predictive Markers for AD in a Multi-Modality Framework: An Analysis of MCI Progression in the ADNI Population," *Neuroimage*, 55, 574–589. [342]

Hinton, G., Srivastava, N., and Swersky, K. (2012), "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, 4, 26–31. [339]

Hon, M., and Khan, N. M. (2017), "Towards Alzheimer's Disease Classification Through Transfer Learning," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 1166–1169. [338]

Jack, C. R., Jr, Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., and Liu, E. (2018), "NIA-AA Research Framework: Toward a Biological Definition of Alzheimer's Disease," *Alzheimer's & Dementia*, 14, 535–562. [337]

Jack, C. R., Jr, Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L., Zeng, G., Bernstein, M. A., Gunter, J. L., Pankratz, V. S., Aisen, P. S., and Weiner, M. W. (2010), "Brain Beta-Amyloid Measures and Magnetic Resonance Imaging Atrophy Both Predict Time-to-Progression From Mild Cognitive Impairment to Alzheimer's Disease," *Brain*, 133, 3336–3348. [340]

Klein-Koerkamp, Y., A Heckemann, R., T Ramdeen, K., Moreaud, O., Keignart, S., Krainik, A., Hammers, A., Baciu, M., Hot, P., and Alzheimer's Disease Neuroimaging Initiative (2014), "Amygdalar Atrophy in Early Alzheimer's Disease," *Current Alzheimer Research*, 11, 239–252. [340]

Korolev, I. O., Symonds, L. L., Bozoki, A. C., and Alzheimer's Disease Neuroimaging Initiative (2016), "Predicting Progression From Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification," *PLOS One*, 11, e0138866. [340]

Ledig, C., Schuh, A., Guerrero, R., Heckemann, R. A., and Rueckert, D. (2018), "Structural Brain Imaging in Alzheimer's Disease and Mild Cognitive Impairment: Biomarker Analysis and Shared Morphometry Database," *Scientific Reports*, 8, 11258. [337,339,341]

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X., and Alzheimer's Disease Neuroimaging Initiative (2018), "Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment," *Frontiers in Neuroscience*, 12, 777. [338,342]

Matthews, K. A., Xu, W., Gaglioti, A. H., Holt, J. B., Croft, J. B., Mack, D., and McGuire, L. C. (2019), "Racial and Ethnic Estimates of Alzheimer's Disease and Related Dementias in the United States (2015–2060) in Adults Aged ≥ 65 Years," *Alzheimer's & Dementia*, 15, 17–24. [337]

Mehta, D., Jackson, R., Paul, G., Shi, J., and Sabbagh, M. (2017), "Why Do Trials for Alzheimer's Disease Drugs Keep Failing? A Discontinued Drug Perspective for 2010–2015," *Expert Opinion on Investigational Drugs*, 26, 735–739. [342]

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and Alzheimer's Disease Neuroimaging Initiative (2015), "Machine Learning Framework for Early MRI-Based Alzheimer's Conversion Prediction in MCI Subjects," *Neuroimage*, 104, 398–412. [338]

Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012), "Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis," *Neuroimage*, 61, 1402–1418. [341]

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K. R., Zhao, T., Batiste, R., and Van Arnam, J. (2018), "Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images," *Cell Reports*, 23, 181–193. [337]

Sheehan, B. (2012), "Assessment Scales in Dementia," *Therapeutic Advances in Neurological Disorders*, 5, 349–358. [342]

Simonyan, K., and Zisserman, A. (2014), "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv no. 1409.1556. [338]

Spasov, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., and Alzheimer's Disease Neuroimaging Initiative (2019), "A Parameter-Efficient Deep Learning Approach to Predict Conversion From Mild Cognitive Impairment to Alzheimer's Disease," *Neuroimage*, 189, 276–287. [341,342]

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., and Alzheimer's Disease Neuroimaging Initiative (2017), "A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease," *IEEE Transactions on Biomedical Engineering*, 64, 155–165. [338,340]

Tractenberg, R. E., Schafer, K., and Morris, J. C. (2001), "Interobserver Disagreements on Clinical Dementia Rating Assessment: Interpretation and Implications for Training," *Alzheimer Disease & Associated Disorders*, 15, 155–161. [342]

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., van Strien, N., Stone, J. R., Gee, J. C., and Avants, B. B. (2014), "Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements," *Neuroimage*, 99, 166–179. [338]

Valliani, A., and Soni, A. (2017), "Deep Residual Nets for Improved Alzheimer's Diagnosis," in *BCB*, p. 615. [340]

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007), "Super Learner," *Statistical Applications in Genetics and Molecular Biology*, 6, 25. [341]

World Health Organization (2019), "Key Facts," available at *https://www.who.int/news-room/fact-sheets/detail/dementia*. [337]