Predicting rate of cognitive decline at baseline using a deep neural network with multidata analysis

Sema Candemir,* **Xuan V. Nguyen,*** **Luciano M. Prevedello,**
Matthew T. Bigelow, Richard D. White, Barbaros S. Erdal,
and Alzheimer's Disease Neuroimaging Initiative[†]
The Ohio State University College of Medicine, Laboratory for Augmented Intelligence
in Imaging, Department of Radiology, Columbus, Ohio, United States

Abstract

Purpose: Our study investigates whether a machine-learning-based system can predict the rate of cognitive decline in mildly cognitively impaired patients by processing only the clinical and imaging data collected at the initial visit.

Approach: We built a predictive model based on a supervised hybrid neural network utilizing a three-dimensional convolutional neural network to perform volume analysis of magnetic resonance imaging (MRI) and integration of nonimaging clinical data at the fully connected layer of the architecture. The experiments are conducted on the Alzheimer's Disease Neuroimaging Initiative dataset.

Results: Experimental results confirm that there is a correlation between cognitive decline and the data obtained at the first visit. The system achieved an area under the receiver operator curve of 0.70 for cognitive decline class prediction.

Conclusion: To our knowledge, this is the first study that predicts "slowly deteriorating/stable" or "rapidly deteriorating" classes by processing routinely collected baseline clinical and demographic data [baseline MRI, baseline mini-mental state examination (MMSE), scalar volumetric data, age, gender, education, ethnicity, and race]. The training data are built based on MMSE-rate values. Unlike the studies in the literature that focus on predicting mild cognitive impairment (MCI)-to-Alzheimer's disease conversion and disease classification, we approach the problem as an early prediction of cognitive decline rate in MCI patients.

© 2020 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.7.4.044501]

1 Introduction

Mild cognitive impairment (MCI) is an intermediate stage between cognitively normal (CN) and Alzheimer's disease (AD).[1] The patients in the MCI phase have a varied prognosis such that the cognitive functions of some MCI patients deteriorate, whereas others remain stable or improve.[2,3] Although there has not been any successful treatment to reverse cognitive decline, to date, therapy to decelerate its progression is likely to be most beneficial if it is applied early.[4,5] In this study, we investigate whether a machine learning-based system can predict the "rate of

*Address all correspondence to Sema Candemir, E-mail: sema.candemir@osumc.edu, candemirsema@gmail.com; Xuan V. Nguyen, E-mail: xuan.nguyen@osumc.edu

cognitive decline" in patients with diagnosed MCI by processing only the clinical and imaging data obtained at the initial visit.

Prior studies have reported on biomarkers and the prediction of MCI-to-AD conversion.[4,6–9] Unlike earlier studies, we investigate the feasibility of predicting the "rate of cognitive decline" in MCI patients at the first visit by processing only the baseline MRI and routinely collected clinical data. We built a deep-learning-based predictive model that integrates imaging and non-imaging demographic and clinical data in the same neural network architecture. The system consists of three main inputs: (1) MRI brain images, (2) scalar volumetric features, and (3) demographic and clinical data. MRI brain scans are provided to the network as sequential digital imaging and communications in medicine images and processed through a three-dimensional convolutional neural network (3D-CNN). The scalar volumetric features extracted using FreeSurfer[10] represent selected brain substructures and included total intracranial volume, whole-brain volume, and regional volumes of the hippocampus, entorhinal cortex, fusiform gyrus, and medial temporal lobe. This scalar data are integrated into the system at the fully connected layer of the architecture. The demographic and clinical information included in the neural network architecture are the ones that are routinely collected at the initial clinical visit and include age, gender, years of education, ethnicity, race, and baseline mini-mental state examination (MMSE) score. The proposed predictive model is illustrated in Fig. 2.

We supervised the predictive model with the change in "MMSE scores[11,12]" with the MCI subjects grouped clinically according to (i) slow cognitive decline and (ii) fast cognitive decline. The resulting model processes the clinical data obtained at the baseline visit and predicts the patient's cognitive condition as either "slowly deteriorating/stable" or "rapidly deteriorating." The analysis is performed on a publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (cf., Sec. 2.1).

2 Materials and Methods

2.1 *Data*

The data used in this study were obtained from the ADNI,[13] which is an ongoing multicenter study. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The subjects in the dataset were diagnosed as AD, MCI, subjective memory concern, or CN based on MMSE scores. The enrolled subjects received multiple longitudinal follow-up visits over several years, as specified by the ADNI protocol. For our research, we utilize data on ADNI patients who were clinically diagnosed as MCI at their baseline visits. A total of 569 subjects were included. The demographics and clinical characteristics of the subjects are summarized in Table 1.

Table 1 Characteristics of the study subjects.

Label	MCI
Number of patients	569
Age: mean (range)	76 (55 to 92)
Gender (female and male)	F: 241 and M: 328
Education: mean (range)	16 (6 to 20)
Ethnicity	Not hispanic/hispanic
Race	White, Black, and Asian
Baseline MMSE score: mean (range)	28 (23 to 30)

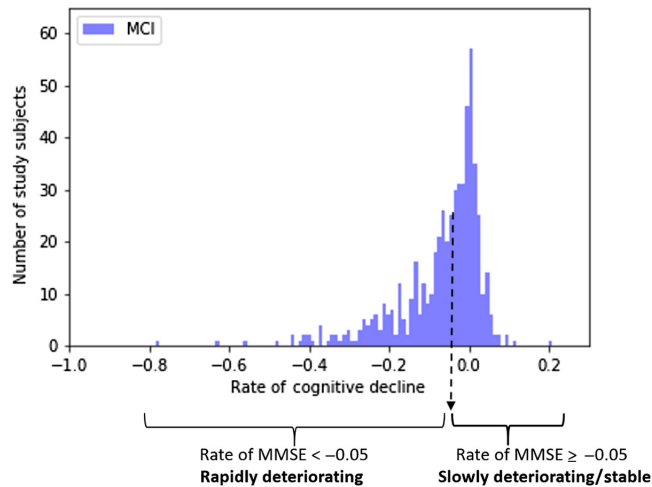MMSE, mini-mental state examination and MCI, mild cognitive impairment.

Fig. 1 The rate of cognitive decline distribution of the study subjects.

The MMSE, which is a 30-point test, is a cognitive assessment tool[11,12] and we used the rate of decline in MMSE scores to supervise the system. Changes in MMSE scores in follow-up visits demonstrate the patient's condition in terms of cognitive capabilities. A decrease in MMSE score reflects deterioration in cognitive capabilities; if a patient's cognitive capability is stable, the MMSE scores remain relatively stable.

We model the change in MMSE scores by fitting a line to the scores obtained at follow-up visits. The slope of the line indicates the rate of cognitive loss. A patient who has faster cognitive deterioration would have a higher absolute value of slope. A slope close to zero indicates that the cognitive decline is stable. In this document, the "rate of cognitive decline" term will refer to the slope of the decline. The predictive model is binary. Therefore, the rate of cognitive decline is converted to binary variables using a threshold of −0.05 points/month, such that progressive rapidly deteriorating level of cognition is defined as a rate of decrease exceeding 0.6 points/year. This threshold approximates the mean and the median rate change for this cohort. The rate of cognitive decline distribution of the study subjects is shown in Fig. 1.

2.2 System Pipeline

The predictive model learns the mapping function from input data to the target output. Let V be the imaging sequence, D be the corresponding clinical data, y be the target class, and $f(.)$ represent the mapping function between input data and output labels. The model can be formulated as

$$y_i = f(V_i, D_i) \tag{1}$$

for each subject i in N, where N is the number of patients with MCI in the training data. Clinical data include age, gender, baseline MMSE score, education, ethnicity, and race. We also use brain volumes as supporting scalar features, which are computed with an open-source library (FreeSurfer) that analyzes and visualizes structural and functional neuroimaging data.[10] Specifically used scalar features are whole-brain volume and regional volumes of the hippocampus, entorhinal cortex, fusiform gyrus, and medial temporal lobe. The brain volumes of each subject are available in the ADNI.[13] We pose the problem as a supervised classification task, with training subjects classified into two groups based on the rate of MMSE (cf., Fig. 1). The output variable $y \in (0,1)$ denotes the target classes, 0 represents "slowly deteriorating/ stable" class, and 1 represents "rapidly deteriorating" class. The proposed system is illustrated in Fig. 2.
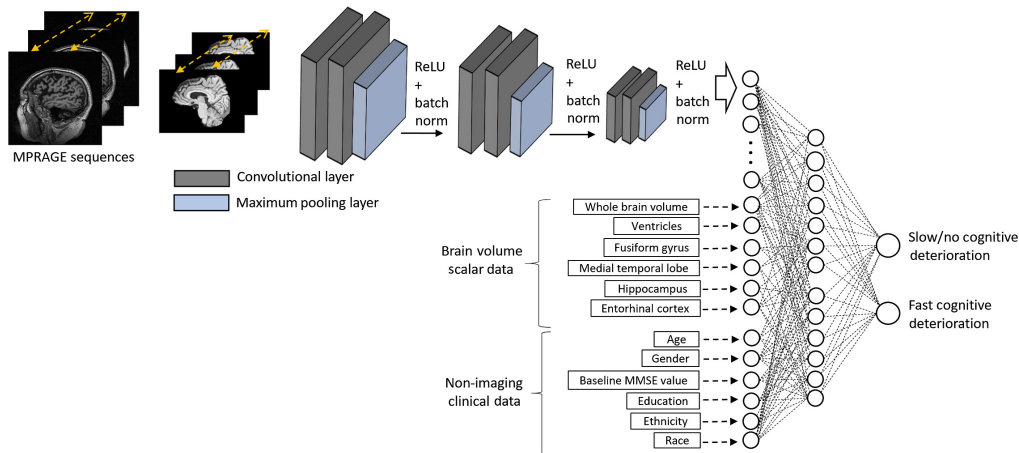
Fig. 2 An illustration of the hybrid prediction system. MPRAGE, magnetization prepared rapid gradient echo; MRI, magnetic resonance imaging; ReLU, rectified linear unit; and batch norm, batch normalization.

2.3 Preprocessing

We apply preprocessing techniques to each MRI volume V and corresponding clinical data D before the training. The MRI sequences are skull-stripped, which includes removal of noncerebral tissue (calvaria, scalp, and dura).[14] The skull-strip algorithm, which is based on U-net architecture[15] trained on skull-stripping datasets,[16] reduces the processing size of volumes, thereby increasingsb computational speed during the training. After the skull-strip, we have applied MRI scale standardization[17] to mitigate the intensity differences between the MRI sequences.

The neural network architectures require the inputs to be scaled in a consistent way for a stable and faster convergence. Therefore, we normalize the images, scalar volumetric features, and demographic and clinical data into the range between 0 and 1. The scalar regional volume features are divided by each subject's whole-brain volume size for normalization. The demographic and clinical data contains categorical values (e.g., gender and ethnicity) that are converted into numeric data. The range for the numeric demographic data is between 0 and 1 to ensure numerical stability.

2.4 Model Configuration of the Neural Network

The deep learning algorithm is based on a supervised neural network that has a hybrid architecture with two main components: (i) a 3D-CNN that learns the brain morphology and patterns and (ii) integration of scalar volumetric features and nonimaging data (demographic and clinical information) at the fully connected layer.

2.4.1 Convolutional neural network

The 3D-CNN processes MRI scans models the patterns and structures in brain volume. Earlier layers of the model capture the low-level features of brain details, whereas higher-level layers learn abstract features. The layout of the 3D-CNN architecture is employed from [14] that is proposed for MRI analysis for AD/CN classification. The architecture consists of three batches of convolutional layers with kernels of $3 \times 3 \times 3$ elements. Each batch contains two convolutional layers with 32, 32, 64, 64 and 128, 128 filters, respectively. The batches are then followed by a batch normalization that mitigates the overfitting and improves the system generalization by normalizing the output of the convolutional layer.[18] After batch normalization, max-pooling layers with 2^3, 3^3, and 4^3 sizes are used for feature reduction and spatial invariance. The architecture uses rectified linear unit (ReLU) activation that introduces nonlinearity to the system.[19] The output of the deepest convolutional layer is flattened and fed to the fully connected layer. The architecture parameters are listed in Table 2.

Table 2 The architecture parameters of the proposed model. Each row represents a layer, and the input of a particular layer is the output of the previous layer. There are two input layers: (1) processing MRI sequences and (2) processing meta data. The meta-data input layer is added to the architecture at the dense layer through a concatenate function.

Layer (type)	Input size (output shape)	# of filters	Level of pooling	# of parameters	
Input (MRI)	$130 \times 116 \times 83$			0	
Conv3-D	$130 \times 116 \times 83$	32	—	896	
Conv3-D	$130 \times 116 \times 83$	32	—	27,680	
Batch norm	$130 \times 116 \times 83$	32	—	128	
Max pooling 3-D	$65 \times 58 \times 41$	32	2	0	
Conv3-D	$65 \times 58 \times 41$	64	—	55,360	
Conv3-D	$65 \times 58 \times 41$	64	—	110,656	
Batch norm	$65 \times 58 \times 41$	64	–	256	
Max pooling 3-D	$21 \times 19 \times 13$	64	3	0	
Conv3-D	$21 \times 19 \times 13$	128	—	221,312	
Conv3-D	$21 \times 19 \times 13$	128	—	442,496	
Batch norm	$21 \times 19 \times 13$	128	—	512	
Max pooling 3-D	$5 \times 4 \times 3$	128	4	0	
Flatten	7680			0	**Input (metadata)**
Dense	512			3,932,672	12
Concatenate	524			0	
Dropout					
Dense	256			134,400	
Dropout					
Dense (out)	2			514	

2.4.2 Integration of scalar volumetric features, nonimaging demographic and clinical data with CNN

The clinical and demographic information presumably contains additional information that would help the classification decision. To incorporate the nonimaging data for assessment of its impact, we have changed the standard CNN architecture. The convolutional part of CNN is the feature extraction component of the architecture, and the fully connected layer part is the classifier component. The output of the final pooling layer, which holds the imaging features, is flattened and fed into the fully connected layer. The flattened imaging features and nonimaging features create a concatenated vector as an input to the dense layer. The remaining part of the architecture is the classifier component of the hybrid system, trained with this vector to form the final prediction model. The concatenated dense layer is then followed by a dropout layer,[20] in which the system temporarily ignores randomly selected neurons during the training to prevent the system from memorizing the training data with the intent to decrease overfitting. The final layer is another dense layer with a softmax activation function with two nodes that provide

Table 3 Implementation details: parameters.

Parameter	Value
Processing dimension of each MRI volume	$116 \times 130 \times 83$ voxels
Optimizer	Adam[24]
Learning rate	0.00005
β_1	0.9
β_2	0.999
ϵ	10^{-8}
Loss function	Categorical cross entropy
Batch size	16
Dropout keep rate	0.5
$L2$ weight regularizer[23] kernel coefficient	0.5
$L2$ weight regularizer bias coefficient	1
Early stopping max epoch	400
Early stopping patience epoch	20

probabilities for "slowly deteriorating/stable" class and "rapidly deteriorating" class. The architecture parameters are listed in Table 2.

2.5 Addressing Overfitting

The voxel-based CNNs are prone to overfitting due to high-dimensional data, large number of parameters, and relatively small number of cases to optimally train the system.[14,21,22] To address the relatively low number of patients, we utilized augmentation strategies. We flipped MRI volumes such that left and right hemispheres are reversed[14] and randomly tilted at <5 deg. We have also employed the regularization techniques of dropout[20] and weight decays[23] in order to increase the generalization capacity of the model. The parameters of dropout and weight decays are listed in Table 3.

3 Experiments

3.1 Implementation Details

The dataset used in the study consists of 569 subjects with MPRAGE (MRI) scans and corresponding clinical data (cf., Sec. 2.1). We perform fivefold cross validation to reduce the performance difference due to relatively small size datasets and provide more robust generalization performance. At each fold, 60% of the dataset is used to train the model, 20% is used for model validation, and 20% of the dataset is used to test the model.

The training parameters are listed in Table 3. We train the model using Adam optimizer,[24] which provides faster convergence due to the velocity and acceleration components. As a training strategy, we monitor the model performance and use two early stopping callbacks to stop the training before the model begins to overfit.[25] We set a large epoch value (cf., Table 3, max epoch 400) as an upper bound iteration. The number of training iterations is decided automatically based on the model performance on the validation and training set. If the validation loss has started to increase during the training process, the system triggers the early stopping callback.

If the validation loss continues to increase for another 20 iterations, then the system stops the training. The continuous increase in validation loss is an indication of overfitting. The second callback is monitoring the training accuracy. If the training accuracy reaches the maximum value, the early stopping callback stops the training due to an indication of no further improvement in the model. The weights are randomly initialized from scratch.

The model is developed in Python (version 3.6.8) using Tensorflow Keras API (version 2.1.6-tf) and trained on an Nvidia Quadro GV100 system with 32 GB graphics cards with CUDA/CuDNN v9 dependencies for GPU acceleration.

3.2 Evaluation

We built three models: (i) an imaging model based on a 3D-CNN that processes brain MRI, (ii) a hybrid model that combines the 3D-CNN component with brain-volume scalar data demographic and clinical information, and (iii) a model that processes brain-volume scalar data demographic and clinical information. We assess the models' prediction performance in terms of accurately classifying the cognitive decline on a test dataset at each test fold and average the evaluation metric scores across all the models. The performance metrics used in the study are sensitivity, specificity, accuracy, PPV, NPV, and AUC. Table 4 lists the performance metrics.

3.2.1 Imaging module prediction performance

The correlation between the morphological changes in the brain (e.g., parenchymal volume loss) and AD is known.[26,27] Based on a prior study,[28] (i) MCI subjects have medium atrophy of hippocampus; (ii) the brain morphology in nonconverters is similar to brain morphology in CN, and converters are more similar to AD; and (iii) converters have more severe deterioration of neuropathology than nonconverters. Due to the correlation between the pathological changes in brain morphology and the AD stages, we first measured how much we could predict the pace of the cognitive decline of patients by processing only the baseline MRI scans through a 3D-CNN. The system achieved 0.67 AUC for predicting the cognitive-decline class by processing only baseline MRI sequences. The receiver operator characteristic (ROC) curve for this experiment is shown in Fig. 3(a).

3.2.2 Hybrid model prediction performance

The hybrid model processes the MRI sequences, brain volume scalar data, and demographic information (age, gender, years of education, ethnicity, and race). Table 4 lists the performance scores obtained with the proposed system in terms of mean and standard deviation across the cross-validated folds. The system achieved an accuracy of 63.3%, with a PPV of 56.9%, sensitivity of 60.8%, specificity of 65.2%, and NPV of 69% at threshold 0.5. The average AUC is 0.67. Adding the brain volume and demographic information as scalar values to the system increased the system performance from 0.67 AUC to 0.70 AUC as shown in Fig. 3(b).

3.2.3 Brain volume scalar data and nonimaging clinical data prediction performance

The voxel-based convolutional neural networks are prone to overfitting due to high-dimensional data, large number of parameters, but relatively low number of subject to optimally train the system.[14,21,22] Although we utilize several regularization techniques, we still observed overfitting due to the 3D-CNN module of the hybrid system. In this experiment, we remove the 3D-CNN module of the hybrid model and run the experiments only using brain-volume scalar data with nonimaging clinical data. The system achieved 0.70 average AUC for cognitive decline class prediction as shown in Fig. 4(a).

Table 4 The hybrid model prediction performance. Training: 60%, validation: 20%, test: 20% of ADNI baseline set.

	Formula	Metric th = 0.2	Metric th = 0.3	Metric th = 0.4	Metric th = 0.5	Metric th = 0.6	Metric th = 0.7	Metric th = 0.8	Metric th = 0.9
Accuracy	$(TP + TN)/(TP + FP + FN + TN)$	0.630 ± 0.024	0.644 ± 0.026	0.646 ± 0.008	0.633 ± 0.024	0.630 ± 0.034	0.628 ± 0.047	0.595 ± 0.047	0.549 ± 0.052
PPV	$TP/(TP + FP)$	0.619 ± 0.045	0.611 ± 0.046	0.597 ± 0.022	0.569 ± 0.026	0.560 ± 0.035	0.550 ± 0.042	0.520 ± 0.039	0.488 ± 0.036
Sensitivity	$TP/(TP + FN)$	0.380 ± 0.103	0.490 ± 0.069	0.551 ± 0.072	0.608 ± 0.077	0.657 ± 0.068	0.755 ± 0.072	0.820 ± 0.088	0.890 ± 0.065
Specificity	$TN/(FP + TN)$	0.818 ± 0.076	0.760 ± 0.071	0.717 ± 0.059	0.652 ± 0.053	0.609 ± 0.061	0.532 ± 0.073	0.425 ± 0.106	0.292 ± 0.095
NPV	$TN/(FN + TN)$	0.638 ± 0.023	0.665 ± 0.019	0.681 ± 0.018	0.690 ± 0.032	0.703 ± 0.037	0.744 ± 0.060	0.766 ± 0.068	0.787 ± 0.099

FN, false negatives; FP, false positives; NPV, negative predictive value; PPV, positive predictive value; TN, true negatives; and TP, true positives.
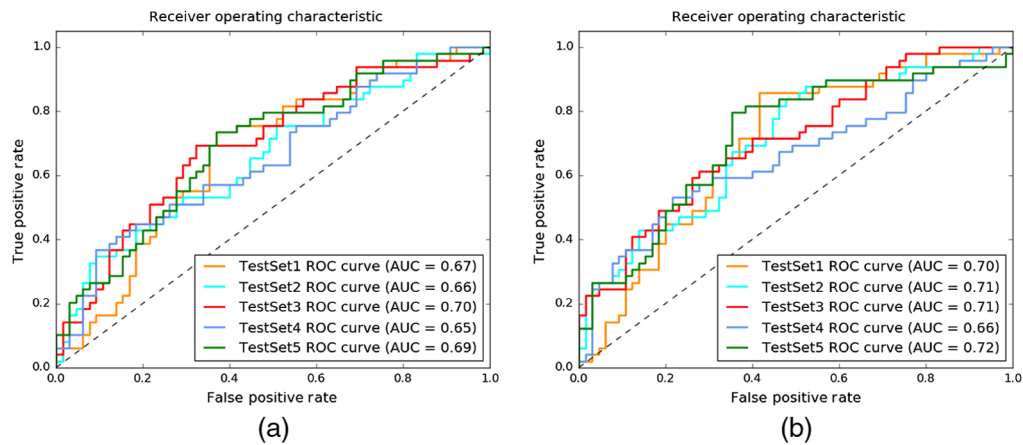
Fig. 3 The plots depict the system performance for predicting cognitive-decline class. (a) The predictive model processed only MRI sequences with 3D-CNN; average AUC is 0.67. (b) The hybrid predictive model is based on MRI sequences with 3D-CNN, brain-volume scalar data and nonimaging clinical data; the average AUC = 0.70.
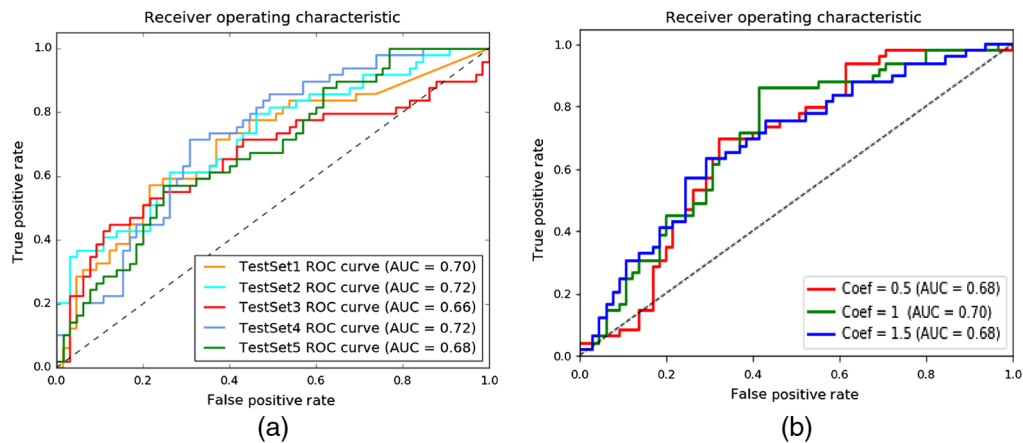

Fig. 4 (a) The predictive model processes only scalar data (brain-volume and nonimaging clinical data); average AUC is 0.70. ROC, receiver operator characteristic and AUC, area under the curve. (b) An illustration of the hybrid prediction system. MPRAGE, magnetization prepared rapid gradient echo; MRI, magnetic resonance imaging; ReLU, rectified linear unit; and batch norm, batch normalization.

3.3 Integration of Components

The literature has several techniques to combine data from different resources. We summarized these studies in Sec. 3.4. We utilized CNN-based architecture as a classifier and combined the imaging features with nonimaging data at a dense layer in a straightforward way, as in Refs. 14 and 29. Note that imaging features or nonimaging features should not dominate the training. To our knowledge, there is not any CNN-based study that adjusts the effects of modules on the prediction results. External weights can be used to adjust the contribution of one module over the other modules. However, these weights are additional hyperparameters of the system and can be decided on train/validation subsets. In order to observe how different weights affect the final decision, we have conducted an additional experiment. We have multiplied scalar input values (scalar brain volumes and demographic data) with external weight (coefficient value) and kept the MRI imaging weights the same. The external weight adjusts the contribution of the nodes to the classifier decision. The coefficient set is 0.5, 1, and 1.5. Figure 4(b) shows the performance of the model with different coefficient values.

3.4 *Comparison with the Literature*

3.4.1 *AD-MCI-CN classification*

The correlation between the morphological changes in the brain (e.g., parenchymal volume loss) and AD has been known for years.[26,27] The literature has several studies with quantitative analysis of brain MRI to assess AD (e.g., classification of AD versus CN).[14,30] These studies measured the volumes, cortical thickness, or shape of various structures such as the hippocampus[28] or the whole brain[14] to assess the disease and the severity of the disease as a percentage of volume. For example, in Ref. 14, a 3D-CNN-based framework is proposed to learn the imaging characteristics of AD and CN through convolutional layers. The model is further modified to diagnose MCI, the prodromal stage of AD. In our study, instead of anatomy-disease correlation, we focus on anatomy-function correlation. A comprehensive review of AD detection/classification can be found in Ref. 31.

3.4.2 *MCI-to-AD conversion*

Many researchers have attempted to predict the conversion of MCI to AD using the correlation between the morphological changes in the brain and disease progression.[28] The volumetric analysis of the brain (especially the hippocampus and entorhinal cortex) produces satisfactory results in predicting conversion to AD.[32,33] One of the most commonly employed classifiers is the support vector machine[8,34] Recent studies utilize deep-learning-based approaches using neural network classifiers[28,35] they show that predicting progressive MCI or detecting MCI patients who later progress to AD is still a goal of ongoing research.[4,9]

3.4.3 *Hybrid models*

The clinical and demographic information contains additional data that contributes to the algorithm decision. To our knowledge, the algorithms that incorporate clinical data into MRI data results are limited.[36] In Ref. 14, age and gender information are concatenated with imaging features in a 3D-CNN architecture through additional nodes at the fully connected layer. In Ref. 28, a CNN was trained with local patches extracted from the hippocampus and combined with FreeSurfer brain data. The algorithm extracted imaging features through a CNN architecture and processed imaging features and FreeSurfer brain data using principal component analysis following by the Lasso regression algorithm. The processed features were provided as input to a NN algorithm that combined these features. Qiu et al.[36] proposed a multimodal fusion model to classify MCI and CN cases. The study employed two multilayer perceptron architectures to train nonimaging data and 2D-CNN to train the imaging data. The predictive model processed the test scores of MMSEs, the Wechsler memory scale for logical memory, and MRI sequences. The predictions from each NN block were then combined using majority voting. Another interesting study that combined baseline MRI with baseline cognitive test scores was proposed in Ref. 8. The cognitive scores used in the study were Rey's auditory verbal learning test, Alzheimer's disease assessment scale cognitive subtest, MMSE, clinical dementia rating sum of boxes, and functional activities questionnaire. The MRI, age, and cognitive measurements were integrated as input features to a random forest classifier. Another hybrid method was proposed in Ref. 37 that combined MRI and FDG-PET images at multiple scales within a NN framework. Six independent deep NNs processed different scales of image sequences. Another NN fused the features extracted from these first 6 DNN. The algorithm was proposed to classify AD and NC cases. One of the most recent studies is[35] that combined MRI sequences, demographic, neuropsychological, and APOe4 genetic data to predict MCI patients who have a likelihood of developing AD within 3 years. The study combined the imaging data with nonimaging data at the fully connected layer of their proposed deep-learning-based architecture. We list the recent studies that combined different sources of information in Table 5. Several other detailed comparison tables can be found in Refs. 8, 31, and 35.

Table 5 Comparison with other studies that use multimodal data.

Study	Method	Data source	Brain region	Objective	Performance
This study	3D-CNN multimodal	Baseline MRI +	Whole brain	Predicting fast decliners	AUC: 0.70
		Baseline MMSE +			Acc: 63.3%
		Demographic data +			Sens: 60.8%
		Baseline scalar volume			Spec: 65.2%
Lee et al.[4]	Recurrent NN multimodal	Baseline MRI +	Regional hippocampus	MCI-to-AD conversion	AUC: 0.86
		Demographic data +			Acc: 81%
		Long. CSF biomarkers +			Sens: 84%
		Long. cognitive performance +			Spec: 80%
Lin et al.[28]	2.5D-CNN multimodal PCA + Lasso + NN	Baseline MRI +	Regional hippocampus	MCI-to-AD conversion	AUC: 0.86
		325 free surfer feature			Acc: 79.9%
					Sens:84.0%
					Spec:74.8%
Lu et al.[37]	Multimodal multiscale NN	MRI + FDG-PET	Whole brain	Stable MCI versus progressive MCI	Acc: 82.9%
		Long. time-points			Sens:79.7%
					Spec:83.8%
Esmaeilzadeh et al.[14]	Multimodal	MRI +	Whole brain	AD-NC-MCI classification	Acc: 94.1%
		Age + Gender			Sens:94%
					Sens:91%
Moradi et al.[8]	Random forest multimodal	Baseline MRI + Age	Whole brain	MCI-to-AD conversion	AUC: 0.9
		Baseline cognitive measurements + (RAVLT + ADAS-cog)			Acc: 82%
					Sens: 87%
		(MMSE + CRD-SB + FAQ)			Spec:74%
Spasov et al.[35]	3D-CNN multimodal	MRI +	Whole brain	MCI-to-AD conversion	AUC: 0.925
		Demographic data +			Acc: 86%
		Neuropsychological data +			Sens: 87.5%
		APOe4 genetic data			Spec:85%

ML, machine learning; 2-D: two-dimension, 3-D: three-dimension; NN, neural network; CNN, convolutional neural network; R-CNN, recurrent convolutional neural network; Acc, accuracy; Sens, sensitivity; Spec, specificity; AUC, area under curve; CSF, cerebrospinal fluid; Long, longitudinal; RAVLT, Rey's auditory verbal learning test; ADAS-cog, Alzheimer's disease assessment scale cognitive subtest; MMSE, mini-mental state examination; CDR-SB, clinical dementia rating sum of boxes; FAQ, functional activities questionnaire.

3.4.4 *This study*

Unlike prior studies, we investigate the feasibility of predicting the "rate of cognitive decline" in MCI patients at the first visit by processing only the baseline MRI and routinely collected clinical data. The training data are separated into two classes based on MMSE-rate values. We train our

model with "slowly deteriorating/stable" or "rapidly deteriorating" classes formed based on MMSE-rate values. Therefore, we do not predict patients that convert to AD. However, some MCI cases deteriorated faster than the others. We investigate the prediction performance of our multimodality architecture to predict the rapidly deteriorating cases based on information available at only the baseline visit. The proposed hybrid architecture jointly learns brain patterns and morphology from MRI sequences and additional information from the demographic data. To our knowledge, this is the first research study that investigates the feasibility of predicting the rate of cognitive decline by processing routine data collected at the first visit.

We follow the same concatenation approach as in Ref. 14, which is proposed for disease detection. Another similar study proposed in Ref. 28 trained a convolutional neural network with local patches extracted from the hippocampus and combined the extracted information with FreeSurfer brain data. Our results are similar in that combining CNN features with scalar brain data features obtained with FreeSurfer increases the prediction performance. However, our study has differences since our model (i) does not predict the MCI-to-AD conversion probability, but instead predicts the rate of cognition deterioration in MCI patients by utilizing only the first-visit data; (ii) identifies patterns within the whole brain MRI instead of only the hippocampus; and (iii) uses limited FreeSurfer brain data (six additional volume elements) compared with the brain data used in Ref. 28 (325 additional data).

Although, we roughly compare our study with MC-to-AD conversion studies, note that there are differences on approaching the problem. To our knowledge, this is the first study that predicts "slowly deteriorating/stable" or "rapidly deteriorating" classes by processing routinely collected baseline clinical and demographic data (baseline MRI, baseline MMSE, scalar volumetric data, age, gender, education, ethnicity, and race). The training data are built based on MMSE-rate values. Therefore, how our study approach predicting progressive MCI is different than the previous studies. Also note that our method does not process data not routinely collected during visits (e.g., APOe4 genetic data) or longitudinal data.

4 Conclusions and Discussion

In this study, we investigate whether a machine learning-based system can predict cognitive decline in MCI patients at the initial visit by processing routinely collected clinical data. Unlike other studies that focus on predicting MCI-to-AD conversion or AD/CN/MCI classification, we approach the problem as an early prediction of cognitive decline rate in MCI patients. The ability to identify an individual's cognitive decline rate potentially helps the clinician to develop early preventive treatment strategies.

We observed the performances of three models for the prediction of cognitive-decline class. Our results confirm that there is a correlation between the cognitive decline and the clinical data obtained at the first visit. The imaging model achieved 0.67 AUC. By adding brain volume and demographic information as scalar values to the system, the performance increased to 0.70 AUC. Processing brain volumes (from FreeSurfer brain data) and demographic information as scalar values provide similar results as the hybrid module performance. Even though patient's cognitive condition is mostly decided based on nonimaging clinical data (e.g., MMSE score and patient age) at the clinical visit, and MRI scans are generally collected to exclude other brain pathology, our results show that the structural MRI provides useful information related to the patients cognitive condition and may further contribute to the clinical evaluation and follow-up of patients with MCI. We have conducted experiments on ADNI dataset (cf., Sec. 2.1) due to the availability of longitudinal MMSE scores and baseline clinical data. To our knowledge, there is not any available dataset that has a rich source of information regarding the Alzheimer's disease and its progression. However, the system needs to be further investigated and validated on an independent dataset.

Our system performance is lower compared to the published studies that investigate MCI-to-AD conversion or AD/CN classification that model the disease progression by processing longitudinal data obtained at several visits or by processing additional data that is not routinely obtained during visits (e.g., APOe4 genetic data). Note that predicting cognitive decline is more challenging than AD/CN classification due to the subtle nature of pathological changes.[28]

Moreover, our system processed only data that is routinely collected at the first visit, and thus makes predictions based on much less information compared to studies that incorporate follow-up data through time-sequence analysis.

The clinical and demographic information contains additional data that contributes to the algorithm decision. In this study, we utilized CNN-based architecture as a classifier and combined the imaging features with nonimaging data at a dense layer in a straightforward way. To our knowledge, there is not any comprehensive study that investigates the best merging methods of different sources of information in CNN-based architecture, and it is an open research area.

Disclosures

The authors declare that they have no conflicts of interest.

Acknowledgments

References

1. W. R. Markesbery, "Neuropathologic alterations in mild cognitive impairment: a review," *J. Alzheimer's Disease* **19**(1), 221–228 (2010).
2. R. C. Petersen et al., "Mild cognitive impairment: clinical characterization and outcome," *Arch. Neurol.* **56**(3), 303–308 (1999).
3. T. Qarni and A. Salardini, "A multifactor approach to mild cognitive impairment," *Semin. Neurol.* **39**(02), 179–187 (2019).
4. G. Lee et al., "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Sci. Rep.* **9**(1), 1952 (2019).
5. J. L. Cummings, R. Doody, and C. Clark, "Disease-modifying therapies for Alzheimer's disease: challenges to early intervention," *Neurology* **69**(16), 1622–1634 (2007).

6. H.-I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," *Lect. Notes Comput. Sci.* **8150**, 583–590 (2013).

7. S. F. Eskildsen et al., "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiol. Aging* **36**, S23–S31 (2015).

8. E. Moradi et al., "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage* **104**, 398–412 (2015).

9. M. Tábuas-Pereira et al., "Prognosis of early-onset vs. late-onset mild cognitive impairment: comparison of conversion rates and its predictors," *Geriatrics* **1**(2), 11 (2016).

10. "Freesurfer," https://surfer.nmr.mgh.harvard.edu// (accessed 30 Nov 2019).

11. I. Arevalo-Rodriguez et al., "Mini-mental state examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)," *Cochrane Database Syst. Rev.* **2015**(3), CD010783 (2015).

12. L. E. Harrell et al., "The severe mini-mental state examination: a new neuropsychologic instrument for the bedside assessment of severely impaired patients with Alzheimer disease," *Alzheimer Disease Assoc. Disord.* **14**(3), 168–175 (2000).

13. "Alzheimer's disease neuroimaging initiative," http://adni.loni.usc.edu// (accessed 24 Nov 2018).

14. S. Esmaeilzadeh et al., "End-to-end Alzheimer's disease diagnosis and biomarker identification," in *Int. Workshop Mach. Learn. Med. Imaging*, Springer, pp. 337–345 (2018).

15. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).

16. "Deepbrain—skull strip algorithm," https://github.com/iitzco/deepbrain// (accessed 10 Oct. 2019).

17. L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging* **19**(2), 143–150 (2000).

18. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Machine Learn.*, Vol. **37**, pp. 448–456 (2015).

19. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436 (2015).

20. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).

21. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).

22. M. R. Arbabshirani et al., "Single subject prediction of brain disorders in neuroimaging: promises and pitfalls," *Neuroimage* **145**, 137–165 (2017).

23. A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization," in *Adv. Neural Inf. Process. Syst.*, pp. 950–957 (1992).

24. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent.*, San Diego, Y. Bengio and Y. LeCun, Eds., (2015).

25. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, London, England (2016).

26. L. A. van de Pol et al., "Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease," *J. Neurol. Neurosurg. Psychiatry* **77**(4), 439–442 (2006).

27. G. Karas et al., "Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease," *Neuroimage* **23**(2), 708–716 (2004).

28. W. Lin et al., "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Front. Neurosci.* **12**, 777 (2018).

29. S. I. Dimitriadis et al., "Random forest feature selection, fusion and ensemble strategy: combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, CMCI and Alzheimer's disease patients: from the Alzheimer's disease neuroimaging initiative (ADNI) database," *J. Neurosci. Methods* **302**, 14–23 (2018).

30. X. Long et al., "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLoS One* **12**(3), e0173372 (2017).

31. S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications," *Neurosci. Biobehav. Rev.* **74**, 58–75 (2017).

32. G. B. Frisoni et al., "The clinical use of structural MRI in Alzheimer disease," *Nat. Rev. Neurol.* **6**(2), 67–77 (2010).
33. L. K. McEvoy et al., "Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis," *Radiology* **259**(3), 834–843 (2011).
34. S. Rathore et al., "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage* **155**, 530–548 (2017).
35. S. Spasov et al., "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *Neuroimage* **189**, 276–287 (2019).
36. S. Qiu et al., "Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's Dementia: Diagn. Assess. Disease Monitor.* **10**, 737–749 (2018).
37. D. Lu et al., "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-pet images," *Sci. Rep.* **8**(1), 1–13 (2018).
38. http://www.fnih.org.

Sema Candemir is a research scientist at the Ohio State University Wexner Medical Center. Her research interest focuses on medical image and video analysis. From 2013 to 2018, she was a postdoctoral research scientist at the U.S. National Library of Medicine (NLM). She has received the Best Paper Award in IEEE CBMS in 2017, the NLM Honor Awards in 2016 to 2017, and the U.S. Department of HHS Ignited Pathway Award for Automatic X-ray Screening for Rural Areas in 2014.

Xuan V. Nguyen is a clinical neuroradiologist and an assistant professor in the Department of Radiology at the Ohio State University College of Medicine. He also has a doctoral degree in anatomy and neurobiology and a bachelor's degree in chemical engineering. He has diverse research interests in radiologic imaging, including recent work on artificial intelligence applications in clinical neuroimaging.

Luciano M Prevedello is an associate professor of radiology at the Ohio State University Medical Center. He is the vice-chair for Medical Informatics and Augmented Intelligence in Imaging. He is a member of the Board of Directors of the Society for Imaging Informatics in Medicine. He also chairs the Machine Learning Steering committee at the Radiological Society of North America and is an associate editor of the Radiology: *Artificial Intelligence Journal*.

Matthew T. Bigelow is ARRT(N)(CT) certified with additional background in imaging informatics. He attended the Ohio State University for undergrad and is a MBA graduate from Fisher School of Business at the Ohio State University. His role in the lab is compiling all data for each use case, extracting them from PACS, and sorting for labeling/tagging. He manages the lab databases works closely with radiologists in developing the GUIs used in tagging/labeling data and displaying/manipulating results.

Richard D. White is medical director of the Program for Augmented Intelligence in Imaging at Mayo Clinic, Florida (2020 to present), succeeding OSU (2010 to 2020) and UF, Jacksonville (2006 to 2010) Radiology Chairmanships; these followed Cleveland Clinic cardiovascular-imaging leadership (1989 to 2006). He received MD (1978 to 1981) and Sarnoff Foundation fellowship at Duke (1981 to 1982). At UCSF, he completed residency with ABR-certification (1982 to 1986) and cardiovascular-imaging fellowship (1985 to 1987). Post-training, he had cardiovascular-imaging leaderships at Georgetown (1987 to 1988) and CWRU (1988 to 1989). He recently refocused toward imaging informatics with MS-Heath Informatics from Northwestern (2016 to 1918).

Barbaros S. Erdal is technical director of the Program for Augmented Intelligence in Imaging at Mayo Clinic, Florida (2020 to present). He received his PhD in electrical and computer engineering from the Ohio State University where he also served as an associate professor of radiology; assistant chief of Medical Imaging Informatics and director of Scholarly Activities (2012 to 2020). Prior to joining Mayo Clinic, he served as the director of Laboratory for Augmented Intelligence in Imaging at OSU Medical Center, Department of Radiology (2018 to 2020).