



A 5-min Cognitive Task With Deep Learning Accurately Detects Early Alzheimer's Disease

Ibrahim Alzubair^{1,2*}, Lin-Ching Chang¹, Kyle F. Shattuck³, Thanh Nguyen¹, Raymond Scott Turner⁴ and Xiong Jiang^{3*}

¹ Department of Electrical Engineering and Computer Science, Catholic University of America, Washington, DC, United States, ² Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia, ³ Department of Neuroscience, Georgetown University Medical Center, Washington, DC, United States, ⁴ Department of Neurology, Georgetown University Medical Center, Washington, DC, United States

Introduction: The goal of this study was to investigate and compare the classification performance of machine learning with behavioral data from standard neuropsychological tests, a cognitive task, or both.

Methods: A neuropsychological battery and a simple 5-min cognitive task were administered to eight individuals with mild cognitive impairment (MCI), eight individuals with mild Alzheimer's disease (AD), and 41 demographically match controls (CN). A fully connected multilayer perceptron (MLP) network and four supervised traditional machine learning algorithms were used.

Results: Traditional machine learning algorithms achieved similar classification performances with neuropsychological or cognitive data. MLP outperformed traditional algorithms with the cognitive data (either alone or together with neuropsychological data), but not neuropsychological data. In particular, MLP with a combination of summarized scores from neuropsychological tests and the cognitive task achieved ~90% sensitivity and ~90% specificity. Applying the models to an independent dataset, in which the participants were demographically different from the ones in the main dataset, a high specificity was maintained (100%), but the sensitivity was dropped to 66.67%.

Discussion: Deep learning with data from specific cognitive task(s) holds promise for assisting in the early diagnosis of Alzheimer's disease, but future work with a large and diverse sample is necessary to validate and to improve this approach.

Keywords: Alzheimer's disease, machine learning, artificial neural networks, inhibition of return, neuropsychological test

OPEN ACCESS

Edited by:

Alessandro Martorana,
University of Rome Tor Vergata, Italy

Reviewed by:

Giovanni Augusto Carlesimo,
University of Rome Tor Vergata, Italy

Mingliang Wang,
Nanjing University of Information
Science and Technology, China

*Correspondence:

Ibrahim Alzubair
47almazair@cua.edu
Xiong Jiang
Xiong.Jiang@georgetown.edu

Received: 05 September 2020

Accepted: 13 November 2020

Published: 03 December 2020

Citation:

Alzubair I, Chang L-C, Shattuck KF, Nguyen T, Turner RS and Jiang X (2020) A 5-min Cognitive Task With Deep Learning Accurately Detects Early Alzheimer's Disease. *Front. Aging Neurosci.* 12:603179. doi: 10.3389/fnagi.2020.603179

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia in older adults. Due to significant progress in basic and clinical research, putative disease-modifying treatments for AD may be on the horizon - which may be most effective in early disease stages. As a result, there is increasing impetus to develop techniques that have high sensitivity and specificity to assist in the diagnosis of early AD (Fiandaca et al., 2014).

Machine learning—with the ability to extract features from high dimensional spaces—holds strong promise in assisting disease diagnosis in both translational research and clinical practice

(Weng et al., 2017; Dwyer et al., 2018), especially with recent advances in deep learning techniques (Esteve et al., 2017). Over the past decade, there has been increasing interest in developing machine learning techniques to assist in the diagnosis of AD and mild cognitive impairment (MCI) and to predict disease progression. Most of these studies focus on brain imaging data from magnetic resonance imaging (MRI) or positron emission tomography (PET) scans (Pellegrini et al., 2018), or cerebrospinal fluid (CSF) proteomics to assess CNS amyloid deposition (A), pathologic tau accumulation (T), and neurodegeneration (N) – the A/T/N criteria under the current NIA-AA research framework (Jack et al., 2018). Compared to brain imaging data, behavioral data are feasible and relatively inexpensive to collect. Behavioral data from speech (Fraser et al., 2016; Nagumo et al., 2020), body movement (Khan and Jacobs, 2020), and neuropsychological test scores (Lemos et al., 2012; Williams et al., 2013; Kang et al., 2019; Lee et al., 2019a) may provide useful features to machine learning classifiers for the diagnosis of MCI and AD.

In addition to standard neuropsychological tests that are widely used in both research and clinical environments, cognitive tasks are usually highly specific and customized and are often only used in research studies. Compared to standard neuropsychological tests, cognitive tasks have certain advantages and disadvantages: on the one hand, cognitive tasks are usually limited by a lack of standardized data and/or validation with a large population of participants; on the other hand, cognitive tasks are often based on cutting-edge research hypothesis and may be more sensitive in detecting very specific changes in brain function due to brain disease such as AD (Perry and Hodges, 1999) – which might eventually lead to the development of improved and/or novel neuropsychological tests (or being integrated with existing neuropsychological test battery) that can be used in clinical practice after validation. Machine learning studies have shown that data from certain cognitive tasks may contain useful information to differential AD/MCI patients from healthy controls (Wallert et al., 2018; Valladares-Rodriguez et al., 2019; Hong et al., 2020). Therefore, it is of a high interest to investigate whether a combination of neuropsychological tests and cognitive task(s) may improve machine learning-based classification accuracy in AD (Wallert et al., 2018; He et al., 2019). In a previous study with traditional machine learning models and multivariate feature selection techniques, we investigated the classification performance with data from a standard neuropsychological test battery, a 5-min cognitive task, or both, to distinguish CN from MCI/AD patients (Almubark et al., 2019). The cognitive task was designed to assess the effects of spatial inhibition of return (IOR). Spatial IOR refers to the phenomenon by which individuals are slower to respond to stimuli appearing at a previously cued location compared to uncued locations when the stimuli onset asynchrony (SOA) between the target and cue is long (~300–500 ms or more) (Klein, 2000). First reported by Posner and Cohen (1984), spatial IOR has been extensively studied, including in healthy older adults (Hartley and Kieley, 1995), patients with various neurodegenerative disorders (Possin et al., 2009; Bayer et al., 2014), and non-human subjects (Shariat Torbaghan et al., 2012). In addition

to the superior colliculus (Posner et al., 1985), cortical areas such as the temporoparietal junction (TPJ) and the inferior parietal cortex are important to maintain normal spatial IOR function (Seidel Malkinson and Bartolomeo, 2018; Satel et al., 2019). Both regions have been implicated in AD progression (Besson et al., 2015), suggesting that spatial IOR may be useful to assist MCI and AD diagnosis. While early studies suggest that spatial IOR is relatively preserved in AD (Amieva et al., 2004), recently we (Jiang et al., 2020) and others (Tales et al., 2005, 2011; Bayer et al., 2014) have provided evidence that spatial IOR impairment in MCI/AD, and spatial IOR impairment in MCI patients may be predictive of conversion to dementia (Bayer et al., 2014). Therefore, machine learning with spatial IOR data may be useful in assisting diagnosis of MCI and AD. In addition, spatial IOR have two appealing features: first, the task is simple to understand and easy to implement, thus making it a feasible tool with AD/MCI patients in a typical clinical setting; second, spatial IOR is robust and resistant to practice effect (Pratt and McAuliffe, 1999; Bao et al., 2011), thus making it an ideal tool in longitudinal studies or clinical trials. However, in the previous study, we found that the classification performance with IOR data as well as the NP data had a low sensitivity and combining IOR and neuropsychological data did not significantly improve classification accuracy (Almubark et al., 2019), suggesting a need for further research.

Deep learning has advantages over machine learning due to its capacity of extracting useful features from highly complex and non-linear datasets (Pedregosa et al., 2011; LeCun et al., 2015), and is gaining popularity in AD research. For example, a PubMed search revealed 8 relevant publications before 2017, 8 in 2017, 26 in 2018, and 65 in 2019. Convolutional-Neural Network (CNN) is the most commonly used deep learning techniques (Gautam and Sharma, 2020). The overwhelming majority of these studies have been focusing on complex and high dimension brain imaging data, especially PET and structural MRI (Jo et al., 2019; Ebrahimighahnavieh et al., 2020; Gautam and Sharma, 2020; Haq et al., 2020). Several recent studies have aimed to integrate multimodal imaging to improve classification performance (Suk et al., 2014; Lu et al., 2018; Huang et al., 2019; Punjabi et al., 2019; Zhou et al., 2019). Deep learning can also help to identify features that are important for disease progression or serve as markers for clinical trials (Ithapu et al., 2015). In addition to harvesting brain imaging data [especially the multimodality imaging data from the public ADNI database (<http://adni.loni.usc.edu/>)], deep learning has been applied to biospecimens (Lee et al., 2019b; Lin et al., 2020), electronic health records (Landi et al., 2020; Nori et al., 2020), speech (Lopez-de-Ipina et al., 2018), neuropsychological data (Choi et al., 2018; Kang et al., 2019), and a combination of MRI and neuropsychological data (Qiu et al., 2018; Duc et al., 2020). By contrast, few studies have applied deep learning to cognitive task data, which – by design – is supposed to be more sensitive to detect early and mild neurocognitive impairment (Locascio et al., 1995; Perry and Hodges, 1999). Highly relevant to the present study, Rutkowski et al. applied various traditional and deep learning models to behavioral data collected from a facial emotion implicit short term memory task (Rutkowski et al., 2020). In their study, Rutkowski et al.

obtained an accuracy close to 90% in distinguishing MCI from normal older adults with either deep learning or logistic regression, supporting a potential of deep learning with cognitive task to aid MCI/AD diagnosis. However, one limitation of their study was that the MCI status was solely defined by the Montreal Cognitive Assessment (MoCA) score rather than a formal clinical evaluation, which is necessary to diagnose MCI (Albert et al., 2011).

In the present study, we further investigated the classification performance of MCI/AD vs. CN using behavioral data from standard neuropsychological tests, a cognitive task (spatial IOR), or both. Both MCI and AD patients were formally diagnosed by clinicians with the consensus guidelines (Albert et al., 2011; McKhann et al., 2011). A variety of machine learning algorithms were tested: four traditional machine learning models and a feed-forward artificial neural network (ANN) model, which has been widely used in AD research (Jo et al., 2019).

MATERIALS AND METHODS

Machine learning was carried out using *Python* and related libraries including *Scikit-learn*, *Pandas*, *Numpy*, *TensorFlow*, and *Keras* (Pedregosa et al., 2011; Chollet, 2015). All experiments were conducted in Google Colaboratory platform (Bisong, 2019).

Data

Participants

Twelve individuals with MCI, 16 individuals with mild AD, and 50 CN participated in the study between 2014 and 2015 (Table 1). The MCI and AD patients were part of the Memory Disorders Program cohort at Georgetown University Medical Center

(<https://memory.georgetown.edu/>). There was no biomarker data from the majority of MCI subjects in this study. All MCI and AD participants were clinically evaluated by clinicians in the Memory Disorders Program with expertise and experience with MCI and AD research. The diagnosis for MCI was based on the clinical interview with the patients and their knowledgeable partners (and also with neuropsychological data when available) using the consensus criteria (Albert et al., 2011) – a Clinical Dementia Rating (CSR) score of 0.5 (https://knightadrc.wustl.edu/cdr/PDFs/CDR_Table.pdf). The AD diagnosis was based on clinical and biomarker data (if available) following the consensus guideline (McKhann et al., 2011). All healthy CN were recruited from the Washington DC metropolitan area. Prior to enrollment, a signed informed consent form approved by the Georgetown University Medical Center's Institutional Review Board was obtained from all participants and their legally authorized representatives (if they had a diagnosis of MCI or mild AD). With the entire study sample, the MCI/AD patients were significantly older than the CN ($p = 0.0001$). As the difference in age could potentially confound the classification results, we identified a subset of demographically comparable subjects, which include 16 MCI/AD patients (8 MCI and 8 AD) and 41 CN. The results from the demographically comparable subset of subjects are included in the main article, and the results from the entire study sample are included in the **Supplementary Materials**.

Neuropsychological Data

The following neuropsychological data were collected from all participants: Mini-Mental State Examination (MMSE); Alzheimer's Disease Assessment Scale – Cognitive subscale (ADAS-Cog); F-A-S Letter Verbal Fluency (LVF) [or Controlled

TABLE 1 | The demographics and neuropsychological test scores of CN and MCI/mild AD participants.

Characteristics	Entire study sample dataset			Demographically comparable dataset		
	CN	MCI/mild AD	p^b	CN	MCI/mild AD	p^b
N (F)	50 (32F ^a)	28 (10F ^a)	n.s. ^c	41 (26F ^a)	16 (6F ^a)	n.s. ^c
Age	65.9 ± 6.2	72.7 ± 7.4	0.0001	67.4 ± 5.3	69.9 ± 5.3	n.s.
Education (years)	18.1 ± 3.9	18.2 ± 3.9	n.s. ^c	18.4 ± 3.4	19.0 ± 4.5	n.s.
%CA	82.0%	89.3%	n.s. ^c	78.1%	81.3%	n.s. ^c
MMSE	29.4 ± 1.0	25.8 ± 4.5	5.3E-06	29.3 ± 1.0	27.3 ± 2.3	3.0E-05
MoCA ^d	25.3 ± 1.8	21.0 ± 4.4	9.7E-05	25.2 ± 1.8	21.4 ± 4.4	0.0035
LM immediate	12.4 ± 3.6	6.6 ± 3.8	4.6E-09	12.0 ± 3.5	7.8 ± 4.2	2.9E-03
LM delayed	9.8 ± 4.3	4.1 ± 3.9	1.2E-07	9.5 ± 4.2	5.4 ± 4.3	0.0019
ADAS-cog	5.9 ± 3.4	18.7 ± 9.7	1.6E-12	6.3 ± 3.4	15.7 ± 7.7	3.1E-08
NPI ^e	2.0 ± 4.5	7.6 ± 8.8	0.0005	2.3 ± 4.9	5.1 ± 5.6	n.s.
LADL ^e	76.4 ± 2.8	67.4 ± 11.9	3.3E-6	76.3 ± 3.0	71.1 ± 8.5	0.0013
LVF	46.8 ± 12.9	37.5 ± 15.5	0.0058	48.2 ± 13.1	39.6 ± 14.0	0.035

CA, Caucasian-Americans; MMSE, Mini-Mental State Exam; MoCA, Montreal Cognitive Assessment; LM, Logical Memory Test; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive subscale; NPI, Neuropsychiatric Inventory; LADL, Lawton Instrumental Activities of Daily Living Scale; LVF, Letter Verbal Fluency (Controlled Oral Word Association Test, COWAT).

^afemale; ^buncorrected p -values for the difference between CN and MCI/mild AD with two-tailed two-sample t -tests (unless otherwise specified); ^cFisher's Exact Test; ^dMoCA were only administered to a subset of participants, including 22 CN and 22 MCI/mild AD patients from the original dataset and 17 CN and 11 MCI/mild AD patients from the demographics matched dataset, and MoCA test scores were not used in classification. ^eNPI and LADL data was missing from one control and one MCI/mild AD patient.

The performance with the optimal hyper-parameter tuning for each dataset is shown in bold and italics font (optimal values for class weight and threshold).

Oral Word Association Test (COWAT)]; Logical Memory subtest of the Wechsler Memory Scale (WMS) – fourth edition (WMS-IV). In addition, data with the Lawton Instrumental Activities of Daily Living Scale (LADL) and Neuropsychiatric Inventory (NPI) were collected and included in this study, as behavioral disturbance and loss of daily functioning are common in AD patients and can be assessed by these two tests (Cipriani et al., 2020; Cummings, 2020).

Cognitive Task [Spatial Inhibition of Return (IOR)]

Experimental details can be found elsewhere (Jiang et al., 2020) and in **Figure 1**. Each trial lasted 2.5 s, and there were 130 trials in total (325 s).

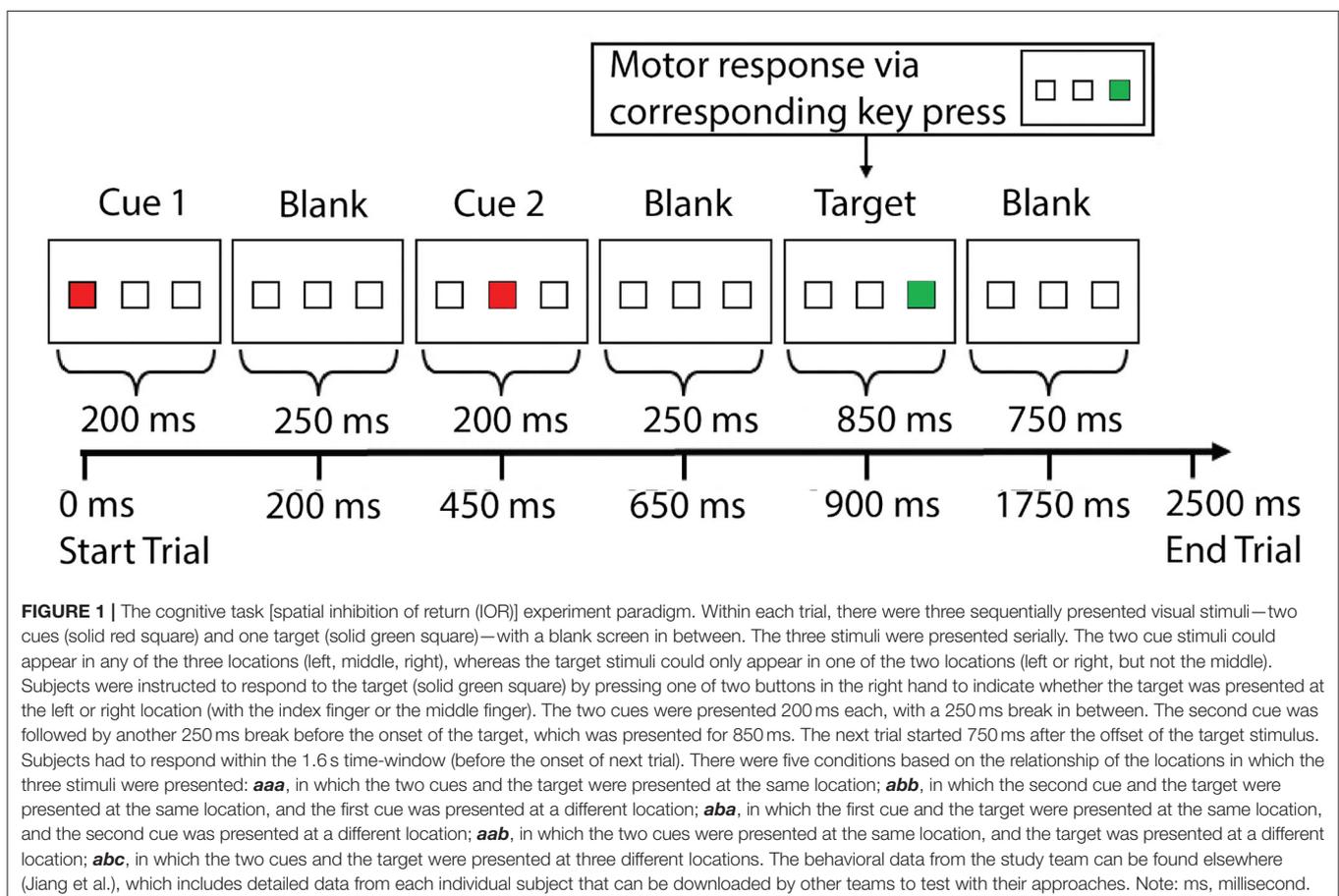
Experimental Design for Classification

The data described in the previous section was arranged into five datasets: Dataset 1 (NP), the scores of nine standard neuropsychological tests (including general cognitive assessments, learning and memory, language, and activities of daily living); Dataset 2 (IOR_{trial}), the responses and reaction time of each trial from the 5.5-min spatial attention IOR task (for a trial without response, the values were set to 0 for response and 10,000 for reaction time); Dataset 3 (IOR_{cond}), the overall accuracy of all trials (with non-responding trials included or excluded), the accuracy (with non-responding trials included

or excluded), and the mean reaction time (correct trials only) of five experimental conditions (**Figure 1**); Dataset 4 (NP + IOR_{trial}), a combination of Datasets 1 (NP) and 2 (IOR_{trial}); Dataset 5 (NP + IOR_{cond}), a combination of Datasets 1 (NP) and 3 (IOR_{cond}). There were a total of 9 features in NP Dataset (7 neuropsychological tests, 3 scores for LVF test (the number of words generated from the three letters, F, A, and S, within 60 s for each letter, respectively, i.e., F: fruit, fog, fun, figure, etc.), 260 features for IOR_{trial} Dataset (responses and reaction time of 130 trials), 17 features for IOR_{cond} Dataset (accuracy and reaction time of experimental conditions, **Figure 1**), 269 features for NP + IOR_{trial} Dataset, and 26 features for NP + IOR_{cond} Dataset.

Data Pre-processing

There were 4 missing values in Dataset 1 (NP) (see **Table 1**), 0 missing value in Dataset 2 (IOR_{trial}), 0 missing value in Dataset 3 (IOR_{cond}), 4 missing values in Dataset 4 (NP + IOR_{trial}), and 4 missing values in Dataset 5 (NP + IOR_{cond}). Missing values in data were handled first by using the multivariate imputation methods available in the fancyimpute library (Rubinsteyn, 2020). Four different missing value imputation techniques were tested: (1) Mean imputation fills a missing value with the mean value of the respective feature from the same group. (2) Nearest neighbor imputations weights samples using the mean squared difference on features for which observed data is contained



in both rows. (3) Softimpute, fast and effective for datasets with high dimensionality, completes matrices through iterative soft thresholding of SVD decompositions (Mazumder et al., 2010). (4) Nuclear norm minimization adopts the *cvxpy* library (Diamond and Boyd, 2016) to provide a simple implementation of Exact Matrix Completion via Convex Optimization (Candès and Recht, 2009). Analyses were carried out using each of the four imputation methods, and nuclear norm minimization was chosen as it consistently provided superior results across all classification algorithms.

Imbalanced classes are common but may result in biased classifiers with poor accuracy on the minority class. To control for class imbalance in the present study (41 CN vs. 16 MCI/AD), an over-sampling technique called Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) was used. Through over-sampling the minority class, the SMOTE technique has shown to improve classification accuracy of imbalanced datasets.

Feature scaling was also performed on both neuropsychological tests and cognitive task datasets. We used the standard scalar, which transforms the data to have the mean of zero with a standard deviation of one. Before the training, a grid search was conducted to obtain an optimal set of hyper-parameters for each algorithm. The grid search works by exhaustive searching through a specified subset of hyper-parameters, and find the best combination of parameters for each algorithm (Bergstra and Bengio, 2012).

Traditional Machine Learning Algorithms

There are many supervised machine learning algorithms for classification problems. Based on the size, quality, and nature of our data, four machine learning algorithms were investigated; Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), and AdaBoost (AB) classifiers.

Support Vector Machines (SVM) are supervised machine learning algorithms that analyze data used for classification, regression and outlier detection (Cristianini and Shawe-Taylor, 2000). Linear SVM seeks a hyperplane that best separates two classes. SVM trains data to find multiple support vectors, which define the hyperplane. The prediction only relies on the support vectors. In addition to linear classification, SVM can use kernels to perform a non-linear classification by mapping their inputs into higher dimensional feature spaces.

Random Forest (RF) algorithm is an ensemble classifier consisting of many decision tree classifiers. Output is determined by the majority vote among all the decision trees for each sample (Breiman, 2001). The RF algorithm combines bootstrap aggregation (bagging) (Breiman, 1996) and random feature (Amit and Geman, 1997) to construct a collection of decision trees exhibiting controlled variation. Because the classification is not based on one tree alone, RF is thought to be more robust than a single decision tree classifier in performance.

Gradient Boosting (GB) is a machine learning technique for classification and regression problems that produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees (Mayr et al., 2014). It trains many

models sequentially. Each new model gradually minimizes the loss function of the whole system using gradient descent. The learning procedure consecutively fits new models to provide a more accurate estimate of the response variable. We selected decision trees with tunable hyper-parameter as base learners for our GB classifier. The GB is used to construct new base learners with maximum correlation with negative gradient of the loss function, associated with the whole ensemble.

AdaBoost (AB) is a meta-learning algorithm used to build a weak classifier iteratively on others according to the performance of the previous weak classifiers (Feng et al., 2005). The AB algorithm can be used for both classification and regression problems. The AB fits a sequence of weak learners on differently weighted training data. The process begins with prediction of the original data set and gives equal weight to each observation. If prediction is incorrect using the first learner, a higher weight is given to observation. Continuing its iterative process, the AB adds learners until a limit is reached in the number of models or accuracy. We used decision stumps (1-layer decision trees) as base learners for our AB classifier, however any machine learning algorithm can be used as base learner if it accepts weight on training data set.

Each of the dataset described in the previous section was used to train each machine learning algorithm using stratified K-Fold cross validation (CV) [see next section Cross Validation (CV)]. Grid search was conducted to determine an optimal set of hyper-parameters using the training data only. Here is the list of hyper-parameters we tuned for each of the four traditional machine learning algorithms. **SVM:** C (1e-3, 1e-2, 0.1, 1, 10), kernel (linear, rbf, poly), and gamma (1e-3, 1e-2, 0.1, 1, 10); **RF:** n_estimators (10, 20, 30, 50, 80, 100), max_depth:np.arange (1, 6, 8), min_samples_split (2, 3, 4, 5), min_samples_leaf (1, 2, 3), and max_features (0.5, log2, auto, 1.0); **GB:** n_estimators (20, 50, 80, 100), learning_rate (0.01, 0.1, 1.), max_depth:np.arange (1, 4), min_samples_split (2, 3, 4), min_samples_leaf (1, 2), and max_features (0.5, log2, auto, 1.0); **AB:** n_estimators (20, 50, 80, 100), and learning_rate (0.01, 0.1, 1.). For the definition of each parameter, the readers can refer to *Scikit-learn* (Pedregosa et al., 2011).

Feature Extraction and Selection

While machine learning algorithms can be developed to deal with a large number of features, such classifiers tend to have lower discriminative power and lower generalization capabilities. Both principal component analysis (PCA) (García-Gil et al., 2018) and feature selection (Li et al., 2017) were used to reduce the number of features in the dataset. In this study, we tested a few feature selection methods including the SelectKBest (SKB) module from *Scikit-learn* with *f_classif* as the metric (Pedregosa et al., 2011), Sequential Forward Selection (SFS), and Sequential Backward Selection (SBS).

Cross Validation (CV)

Cross validation (CV) with stratified K-Fold was used to evaluate the predictive models. For the stratified K-Fold CV, data was divided into 5 disjoint subsets with consistent ratios between patient and control in each fold. Eighty percent (80%) of the data

was used in training and 20% of the data was used for testing in each fold. Note that SMOTE over-sampling, PCA, feature selection, and hyper-parameter grid search were all performed in each fold on the training data only. Leave-one-out CV with default parameters from each algorithms (without grid search) produced similar results (data not shown).

Artificial Neural Networks (ANNs)

A perceptron is a mathematical model of a biological neuron, which is the basic computing unit for artificial neural networks (ANNs). An ANN in its simplest form has only three layers: an input layer, an output layer, and a hidden layer. Deep learning is an ANN with multiple hidden layers. A perceptron can take in a few inputs, each of which has a weight to signify how important it is, and generate an output. A deep learning system can self-teach to learn from data by filtering information through multiple hidden layers which mimicking the human brain in several ways. Deep learning networks have been developed to solve many real-world complex problems (LeCun et al., 2015).

A multilayer perceptron (MLP) is a class of feedforward ANN which often applied to supervised learning problems and uses back-propagation to adjust the weights for training. MLP is commonly used in situations where no analytical solution exists. MLP is very popular in pattern recognition systems and for interpolation and processing massive digital images. It has been used for AD detection with structural MRI images and other medical images from various types of imaging modalities (Santos et al., 2008; Joshi et al., 2010; Tufail et al., 2012).

In MLP, learning involves updating parameters including weights and biases. Training can be broken into three main steps: forward-propagation, error/loss calculation, and back-propagation. In the forward-propagation, we take an input x and multiply it with a weight w and add bias b as in equation (1):

$$y = \sum_{i=1}^n x_i w_i + b \quad (1)$$

Where n denotes the input count, x denotes the vector of inputs, w denotes weights, and b denotes the bias. Then, the activation function is applied on the output result y . The target output (label) is known, and therefore can be compared against the predicted output to compute the loss. A common choice for a loss function in a binary classification task is a binary cross-entropy that can be represented as:

$$Loss = -y \log(p) + (1-y) \log(1-p) \quad (2)$$

where y denotes the actual value and p denotes the predicted probability. Lastly, after calculating the loss, we back-propagate the loss and update the weights by using an optimizer that seeks to minimize the loss function. Weight update is performed according to the formula:

$$w \leftarrow w - \eta \frac{\partial E}{\partial w} \quad (3)$$

where w is the weight, η is the learning rate, and $\frac{\partial E}{\partial w}$ represents the partial derivatives of the error function E . The partial derivative

of the error function with respect to the weight can be calculated using the chain rule as follows:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_i} \cdot \frac{\partial x_i}{\partial w} \quad (4)$$

where y_i is the i -th neuron in the output layer and x_i is the i -th neuron in the input layer. Depending on the utilized loss function, the process may also involve calculating the partial derivatives at each node, adding them to the chain rule, and calculating the product of partial derivatives at each node to obtain the value of $\frac{\partial E}{\partial w}$. In implementation, a training is stopped once a convergence criteria is reached. This means a minima is reached (though this may be a local minima rather than a global minimum). A training may also be stopped after a number of epochs, or a number of passes over the training data.

In this study, we developed a fully connected MLP network for AD classification using data from neuropsychological tests and a simple 5.5-min cognitive task. We used Rectified Linear Unit (ReLU) as the activation function for the input and hidden layers. The function can be written as:

$$f(x) = \max(0, x) \quad (5)$$

We used the sigmoid function as the activation function for the output layer to obtain output between 0 and 1 for prediction of probabilities. The function can be written as:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

Adaptive moment estimation (Adam) (Kingma and Ba, 2015), an adaptive learning rate optimization algorithm designed specifically for training deep neural networks, was used as an optimizer. Binary cross entropy tunable with class weights was used as the loss function, in order to penalize more on the Type II errors. Due to the limited number of samples, our optimization was performed in a stochastic fashion to get the best performance: the batch size was set to 1 and the samples were shuffled before each epoch began. The maximum number of iterations was set to 250. Both the L1 and L2 regularization were added to each layer to constrain overfitting. Early stopping and learning rate shrinkage (with a minimum learning rate of 5×10^{-4}) were performed on monitoring the validation loss function. We performed hyper-parameter tuning with different class weight ratios and probability threshold to improve the sensitivity of the proposed model. A 5-folds CV was constructed by using 20% as the testing set; the rest 80% split into the training and validation sets. The hyper-parameter tuning was performed on the training dataset using the sensitivity on the validation set as the metric. Similar results can be obtained by using the SMOTE over-sampling with more balanced class weights and probability thresholds.

Performance Evaluation

Three performance measurements were used to evaluate the performance of each model: sensitivity, specificity, and accuracy

(equations 7, 8, and 9, respectively).

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where: TP, True Positives; TN, True Negatives; FN, False Negatives; FP, False Positives. Based on the predicted probabilities of each subject, we also plotted the Receiver Operating Characteristic (ROC) curve and calculated the corresponding Area Under the Curve (AUC).

RESULTS

Traditional Machine Learning Algorithms

Supplementary Table 1 summarizes the classification performance for four machine learning algorithms in terms of sensitivity, specificity, and accuracy using each dataset (NP, IOR_{trial}, IOR_{cond}, NP + IOR_{trial}, and NP + IOR_{cond}) and a 5-folds CV. The result from using either all features in each dataset or PCA with 90% of total variation were compared. **Supplementary Table 2** is similar to **Supplementary Table 1** but, in addition, applying SMOTE over-sampling technique to each dataset before training.

Supplementary Table 3 shows the best model performance for each machine learning algorithm with a feature selection technique (SKB or SFS or SBS). The numbers of features used in each algorithm were also reported which were selected based on the highest sensitivities in each model.

Figure 2 shows the ROC curves for the best classifiers for each dataset in which the best traditional machine learning algorithm without SMOTE and with SMOTE were plotted in **Figures 2A,B**, respectively. Note that the best classifier selected for each dataset can be obtained from **Supplementary Tables 1-3**.

The classification performance is comparable between the five datasets across algorithms and techniques (**Supplementary Table 4**). Combining neuropsychological and cognitive data together (NP + IOR_{trial}, and NP + IOR_{cond}) slightly improved classification accuracy over NP only, but the difference was not significant (**Supplementary Tables 1-4**).

Multilayer Perceptron (MLP)

The MLP classification performance with each dataset is shown in **Table 2**. The mean classification accuracy using the conventional choice of the class weight (1:1) and threshold (0.5) is shown, along with the model performance with an optimal setting for each dataset (shown in **bold font**). Each optimal setting was determined by tuning class weight and binarizing

TABLE 2 | Multilayer perceptron (MLP) classification performance using the demographically comparable dataset.

Datasets	Probability threshold	Class weight	SEN%	SPE%	ACC ± std%
NP	0.5	1:1	62.50	87.8	80.7 ± 7.76
	0.4	1:1.17	75.00	87.8	84.21 ± 4.91
IOR _{trial}	0.5	1:1	75.00	95.12	89.47 ± 6.17
	0.5	1:1.5	81.25	95.12	91.23 ± 4.28
IOR _{cond}	0.5	1:1	68.75	97.56	89.47 ± 3.12
	0.4	1:1.5	81.25	90.24	87.72 ± 3.12
NP + IOR _{trial}	0.5	1:1	75.00	95.12	89.47 ± 7.5
	0.5	1:1.1	87.50	95.12	92.98 ± 4.28
NP + IOR _{cond}	0.5	1:1	87.50	97.56	94.74 ± 6.75
	0.5	1:1.5	93.75	92.68	92.98 ± 6.33

The sensitivity (SEN), specificity (SPE), accuracy (ACC), and standard deviation of the accuracy (std) for each dataset were calculated from 5-fold CV using the default setting for class weight (1:1) and threshold (0.5). The performance with the optimal hyper-parameter tuning for each dataset is shown in **bold font** (optimal values for class weight and threshold). IOR, inhibition of return; IOR_{trial}, the reaction time and responses of each trial; IOR_{cond}, the mean accuracy of all trials, mean accuracy of responded trials, and mean reaction time of each condition (**Figure 1**); NP, neuropsychological test scores (**Table 1**).

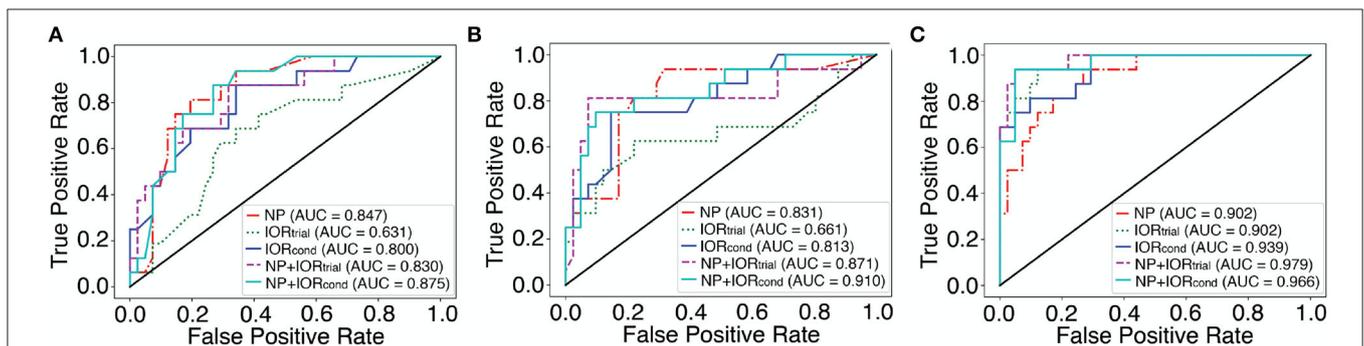


FIGURE 2 | The ROC curves for the best classifiers selected by the highest sensitivity for each dataset with traditional machine learning algorithms and with MLP. See **Table 2, Supplementary Tables 1-3** for the specific algorithms and parameters used for these “best” classifiers (shown in **bold and italics font**). **(A)** Traditional machine learning algorithms with all features and PCA (without and with SMOTE over-sampling). **(B)** Traditional machine learning algorithms with features selection (without and with SMOTE over-sampling). **(C)** The ROC curves for each dataset with MLP using the demographically comparable dataset. The AUC score is shown in the legend box.

threshold in neural network to improve the sensitivity while maintaining accuracy. The comparative ROC curves for each dataset in **Table 2** are plotted in **Figure 2C**.

The MLP classification performance with the 5-min cognitive task (IOR) data is noteworthy, especially with the NP + IOR_{cond} dataset (a combination of summarized scores from neuropsychological tests and the IOR task), which resulted in a high performance, suggesting that deep learning with behavioral data from certain cognitive task(s) (such as spatial IOR task tested here) may be useful to assist early AD detection/diagnosis.

The MLP classification performance using the entire study sample dataset is shown in **Supplementary Table 5**. A direct comparison between the best traditional machine learning and deep learning methods is concluded in **Table 3** (a summary of **Table 2, Supplementary Tables 1-3**).

In addition, using the scores from the ADAS-Cog test – one of most commonly used neuropsychological tests in MCI/AD research and clinical trials (as a primary cognitive outcome), the specificity and sensitivity to distinguish MCI/AD patients from controls were 95.1 and 56.3% with a cut-off score of 12 (Chu et al., 2000), respectively, or 78.1 and 66.8% if a cut-off score of 10 was used (Nogueira et al., 2018).

DISCUSSION

Using traditional machine learning algorithms, the classification based on the 5-min cognitive task (IOR) achieved a performance comparable to the classification based on a neuropsychological test battery (which takes ~75–85 min to administer), and the classification performance was slightly improved when both sets of data (IOR and neuropsychological tests) were used. Deep learning (MLP) outperformed traditional machine learning algorithms with IOR data (either alone or together with neuropsychological data), in particular, a high performance (~90% sensitivity and ~90% specificity) was obtained when MLP was applied to a combination of summarized scores from neuropsychological tests and the 5-min IOR task.

Using standard neuropsychological data (NP Dataset), the classification performance is comparable to other previous

studies with traditional machine learning algorithms and standard neuropsychological data (Lemos et al., 2012; Williams et al., 2013; Grassi et al., 2019; Lee et al., 2019a). Similar performance was also obtained using the summarized performance data of each condition from the 5-min spatial IOR cognitive task (IOR_{cond} Dataset), suggesting that cognitive tasks such as spatial IOR might provide useful information to assist MCI/AD diagnosis. However, when the raw IOR data (i.e., response and reaction time of each individual trial) (IOR_{trial} Dataset) was used, traditional machine learning algorithms struggled with low sensitivities, suggesting that traditional machine learning algorithms might not be the best tool to extract “diagnostic” information from individual trials of a cognitive task—likely due to a high variance in responses to individual trials—and traditional machine learning algorithms might benefit greatly from “guided” feature reductions [i.e., by averaging responses from multiple trials of the same experimental condition (e.g., IOR_{cond} Dataset)].

Deep learning is gaining popularity in medical application, including AD. In the present study, deep learning (MLP) performed comparably to the four traditional machine learning algorithms when the standard neuropsychological data was used. Moreover, MLP with either of the two IOR datasets (IOR_{trial} or IOR_{cond}) consistently outperformed traditional machine learning algorithms with either IOR dataset as well as deep learning with standard neuropsychological data, suggesting that: (i) cognitive tasks such as spatial IOR could produce rich information useful for MCI/AD diagnosis; and (ii) the rich but complex and non-linear information can be reliably extracted/captured by the deep learning algorithms (but might be difficult for traditional machine learning algorithms). It is noteworthy to point out that spatial IOR effects are robust and resistant to practice (Müller and von Mühlénen, 1996; Pratt and McAuliffe, 1999), making it an ideal tool in evaluating disease progression in longitudinal studies and effects of novel interventional treatments; and the spatial IOR task is simple and easy to understand, making it an ideal task to implement in clinical practice and research. Furthermore, combining IOR data and standard neuropsychological data further improved classification performance of MLP, with a sensitivity around 90% and a specificity above 90%, which— if independently verified in larger cohorts—has potential clinical implication.

The present study has several limitations. First, the sample size is small and there are much fewer MCI/AD patients than CN, which might contribute to the low sensitivity as well as the unsatisfying performance with traditional machine learning algorithms. To control for the imbalanced classes in deep learning analysis, we examined the performance with higher class weights for AD samples as well as lower probability thresholds that favor the MCI/AD class to improve sensitivity. The standard deviations from the 5-folds CV are low, showing a low possibility of overfitting of our artificial neural network. This is a result of using appropriate regularization and other previously described methods. Second, due to the small sample size, MCI and AD patients were combined together; this is

TABLE 3 | A direct comparison between the best traditional machine learning and deep learning methods (a summary of **Table 2, Supplementary Tables 1-3**).

Datasets	Method	SEN%	SPE%	ACC ± std%
NP	RF	87.5	70.73	75.61 ± 5.76
	MLP	75.00	87.8	84.21 ± 4.91
IOR _{trial}	AB	62.5	78.05	73.48 ± 9.75
	MLP	81.25	95.12	91.23 ± 4.28
IOR _{cond}	RF	75	80.49	78.79 ± 9.25
	MLP	81.25	90.24	87.72 ± 3.12
NP + IOR _{trial}	GB	81.25	92.68	89.55 ± 3.12
	MLP	87.50	95.12	92.98 ± 4.28
NP + IOR _{cond}	SVM	87.5	82.93	84.09 ± 3.91
	MLP	93.75	92.68	92.98 ± 6.33

suboptimal due to known differences between MCI and AD as well as the clinical significance of distinguishing between MCI and AD. Third, deep learning performs better with large training data sets, thus the classification performance of deep learning with the NP + IOR_{cond} (or NP + IOR_{trial}) must be verified and validated by independent and larger studies. We ran an exploratory analysis using the data from the 21 excluded subjects (12 MCI/AD and 9 CN). That is, during each fold in MLP, the classifiers were also applied to the excluded subjects. There was a hit in classification performance. For instance, for NP + IOR_{cond} Dataset, the specificity remained high at 100%, but the sensitivity, however, was dropped to 66.67%. This is interesting as the excluded MCI/AD patients were either too old (>80 y.o.) or at more advanced stage of disease (i.e., failure to perform the simple IOR task due to time constraints). Indeed, a close inspection revealed that the low sensitivity was driven by four MCI/AD patients (AD, $n = 2$; MCI, $n = 2$), who were consistently misclassified as CN. All of them were older than 80 y.o. and performed high on the spatial IOR cognitive task, or the neuropsychological tests (except memory, for which three of them performed poorly), or both, suggesting that behavioral markers/features in the sixteen MCI/AD patients included in the main analysis (only three of them were older than 74 y.o.) might be different from the markers in the four relatively high performance older MCI/AD patients (Koedam et al., 2010; Ye et al., 2012). That is, visuospatial attention as measured by spatial IOR is more likely being affected in relatively young MCI/AD patients, whereas memory is more likely affected in MCI/AD patients who are at more advanced stage of age (Koedam et al., 2010; Ye et al., 2012). This is confirmed by a high classification performance when the training sample included some of the oldest MCI/AD patients (>80 y.o.) (**Supplementary Table 5**). In addition, we tested our methods to a large and independent dataset from ADNI (<http://adni.loni.usc.edu/>) and obtained encouraging classification performance using neuropsychological data (Almubark et al., 2020). However, spatial IOR data is not available in the ADNI database, thus the prediction power of double-cue spatial IOR task needs to be validated in future studies with a diverse and large sample.

In summary, most previous machine learning studies have focused on brain imaging data. With the readily availability and lower cost of behavioral data, here we investigated the classification performance of MCI/AD vs. CN using traditional machine learning algorithms and deep learning with behavioral data (from standard neuropsychological tests, a specific cognitive task, or both). Deep learning with a combination of standard neuropsychological data and cognitive task (IOR) produced a classification performance with ~90% sensitivity and ~90% specificity, which may be clinically meaningful and supports the collection of simple cognitive task(s) data in future clinical studies. However, due to the small sample size in this study, the conclusion should be taken with caution, and future studies with larger samples are needed to verify, validate, and improve upon the techniques presented in the study, i.e., to exploit public database from the large cohorts in

which both standard neuropsychological tests and cognitive task(s) were administered, or to develop a mobile app to collect spatial IOR data from a large and diverse sample of participants.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Georgetown University IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

IA and L-CC analyzed the data. IA, L-CC, and XJ wrote the manuscript. KS and RT revised the manuscript. RT provided clinical diagnosis of MCI and AD patients. XJ conceptualized the study, designed the behavioral experiment, and supervised the behavioral data collection. All authors reviewed the final manuscript.

FUNDING

This study was funded by the Alzheimer's Drug Discovery Foundation (XJ) and the BrightFocus Foundation (XJ). The authors would also like to thank Qassim University and the Deanship of Scientific Research for their support and funding the publication.

ACKNOWLEDGMENTS

We wish to thank all participants for their time and participation, the Memory Disorders Program at Georgetown University Medical Center, and the assistance for patient care from the Georgetown University Clinical Research Unit (GU-CRU), which has been funded in whole or in part with Federal funds (Grant # UL1TR000101 previously UL1RR031975) from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through the Clinical and Translational Science Awards Program (CTSA), a trademark of DHHS, part of the Roadmap Initiative, Re-Engineering the Clinical Research Enterprise.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2020.603179/full#supplementary-material>

REFERENCES

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008
- Almubark, I., Alsegehy, S., Jiang, X., and Chang, L. (2020). "Classification of Alzheimer's disease, mild cognitive impairment, and normal controls with multilayer perceptron neural network and neuropsychological test data," in *Proceedings of the 12th International Joint Conference on Computational Intelligence - Vol. 1: NCTA*. 439–446. doi: 10.5220/0010143304390446
- Almubark, I., Chang, L.-C., Nguyen, T., Turner, R. S., and Jiang, X. (2019). "Early detection of Alzheimer's disease using patient neuropsychological and cognitive data and machine learning techniques," in *2019 IEEE International Conference on Big Data (Big Data)* (Los Angeles, CA), 5971–5973. doi: 10.1109/BigData47090.2019.9006583
- Amieva, H., Phillips, L. H., Della Sala, S., and Henry, J. D. (2004). Inhibitory functioning in AlzheimerData (Big DaBrain 127, 949–964. doi: 10.1093/brain/awh045
- Amit, Y., and Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Comput.* 9, 1545–1588. doi: 10.1162/neco.1997.9.7.1545
- Bao, Y., Sander, T., Trahms, L., Pöppel, E., Lei, Q., and Zhou, B. (2011). The eccentricity effect of inhibition of return is resistant to practice. *Neurosci. Lett.* 500, 47–51. doi: 10.1016/j.neulet.2011.06.003
- Bayer, A., Phillips, M., Porter, G., Leonards, U., Bompas, A., and Tales, A. (2014). Abnormal inhibition of return in mild cognitive impairment: is it specific to the presence of prodromal dementia? *J. Alzheimers. Dis.* 40, 177–189. doi: 10.3233/JAD-131934
- Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 281–305. doi: 10.5555/2188385.2188395
- Besson, F. L., La Joie, R., Doevre, L., Gaubert, M., Mézenge, F., Egret, S., et al. (2015). Cognitive and brain profiles associated with current neuroimaging biomarkers of preclinical Alzheimer's disease. *J. Neurosci.* 35, 10402–10411. doi: 10.1523/JNEUROSCI.0150-15.2015
- Bisong, E. (2019). "Google cloud machine learning engine (Cloud MLE)," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (Berkeley, CA: Springer), 545–579. doi: 10.1007/978-1-4842-4470-8_41
- Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Candès, E. J., and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* 9:717. doi: 10.1007/s10208-009-9045-5
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Choi, H.-S., Choe, J. Y., Kim, H., Han, J. W., Chi, Y. K., Kim, K., et al. (2018). Deep learning based low-cost high-accuracy diagnostic framework for dementia using comprehensive neuropsychological assessment profiles. *BMC Geriatr.* 18:234. doi: 10.1186/s12877-018-0915-z
- Chollet, F. (2015). *Keras: Deep Learning Library for Theano and Tensorflow*. Available online at: <https://keras.io/k>
- Chu, L. W., Chiu, K. C., Hui, S. L., Yu, G. K., Tsui, W. J., and Lee, P. W. (2000). The reliability and validity of the Alzheimer's disease assessment scale cognitive subscale (ADAS-Cog) among the elderly Chinese in Hong Kong. *Ann. Acad. Med. Singap.* 29, 474–485. doi: 10.1016/S0197-4580(00)83371-0
- Cipriani, G., Danti, S., Picchi, L., Nuti, A., and Fiorino, M. D. (2020). Daily functioning and dementia. *Dement. Neuropsychol.* 14, 93–102. doi: 10.1590/1980-57642020dn14-020001
- Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511801389
- Cummings, J. (2020). The neuropsychiatric inventory: development and applications. *J. Geriatr. Psychiatr. Neurol.* 33, 73–84. doi: 10.1177/0891988719882102
- Diamond, S., and Boyd, S. (2016). CVXPY: A python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* 17, 2909–2913. doi: 10.5555/2946645.3007036
- Duc, N. T., Ryu, S., Qureshi, M. N. I., Choi, M., Lee, K. H., and Lee, B. (2020). 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. *Neuroinformatics* 18, 71–86. doi: 10.1007/s12021-019-09419-w
- Dwyer, D. B., Falkai, P., and Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* 14, 91–118. doi: 10.1146/annurev-clinpsy-032816-045037
- Ebrahimighannavieh, M. A., Luo, S., and Chiong, R. (2020). Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. *Comput. Methods Programs Biomed.* 187:105242. doi: 10.1016/j.cmpb.2019.105242
- Esteve, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056
- Feng, K.-Y., Cai, Y.-D., and Chou, K.-C. (2005). Boosting classifier for predicting protein domain structural class. *Biochem. Biophys. Res. Commun.* 334, 213–217. doi: 10.1016/j.bbrc.2005.06.075
- Fiandaca, M. S., Mapstone, M. E., Cheema, A. K., and Federoff, H. J. (2014). The critical need for defining preclinical biomarkers in Alzheimer. *etworks. iAlzheimers Dement.* 10, S196–S212. doi: 10.1016/j.jalz.2014.04.015
- Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *J. Alzheimers Dis.* 49, 407–422. doi: 10.3233/JAD-150520
- García-Gil, D., Ramírez-Gallego, S., García, S., and Herrera, F. (2018). Principal components analysis random discretization ensemble for big data. *Knowl. Based Syst.* 150, 166–174. doi: 10.1016/j.knsys.2018.03.012
- Gautam, R., and Sharma, M. (2020). Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. *J. Med. Syst.* 44:49. doi: 10.1007/s10916-019-1519-7
- Grassi, M., Rouleaux, N., Caldirola, 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
- Haq, E. U., Huang, J., Kang, L., Haq, H. U., and Zhan, T. (2020). Image-based state-of-the-art techniques for the identification and classification of brain diseases: a review. *Med. Biol. Eng. Comput.* 58, 2603–2620. doi: 10.1007/s11517-020-02256-z
- Hartley, A. A., and Kieley, J. M. (1995). Adult age differences in the inhibition of return of visual attention. *Psychol. Aging* 10, 670–683. doi: 10.1037/0882-7974.10.4.670
- He, H., Xu, P., Wu, T., Chen, Y., Wang, J., Qiu, Y., et al. (2019). Reduced capacity of cognitive control in older adults with mild cognitive impairment. *J. Alzheimers Dis.* 71, 185–200. doi: 10.3233/JAD-181006
- Hong, Y., Alvarado, R. L., Jog, A., Greve, D. N., and Salat, D. H. (2020). Serial reaction time task performance in older adults with neuropsychologically defined mild cognitive impairment. *J. Alzheimers. Dis.* 74, 491–500. doi: 10.3233/JAD-191323
- Huang, W., Luo, M., Liu, X., Zhang, P., Ding, H., Xue, W., et al. (2019). Arterial spin labeling images synthesis from sMRI using unbalanced deep discriminant learning. *IEEE Trans. Med. Imaging* 38, 2338–2351. doi: 10.1109/TMI.2019.2906677
- Ithapu, V. K., Singh, V., Okonkwo, O. C., Chappell, R. J., Dowling, N. M., and Johnson, S. C. (2015). Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimer Dement.* 11, 1489–1499. doi: 10.1016/j.jalz.2015.01.010
- Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Aeberlein, S. B., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer deep learniAlzheimers Dement. 14, 535–562. doi: 10.1016/j.jalz.2018.02.018
- Jiang, X., Howard, J. H. Jr, Rebeck, G. W., and Turner, R. S. (2020). Spatial inhibition of return is impaired in mild cognitive impairment and mild Alzheimer's disease. *bioRxiv*. Available online at: <https://www.biorxiv.org/content/10.1101/2020.05.11.089383v1> (accessed May 16, 2020).

- Jo, T., Nho, K., and Saykin, A. J. (2019). Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front. Aging Neurosci.* 11:220. doi: 10.3389/fnagi.2019.00220
- Joshi, S., Simha, V., Shenoy, D., Venugopal, K. R., and Patnaik, L. M. (2010). Classification and treatment of different stages of Alzheimer's disease using various machine learning methods. *Int. J. Bioinform. Res.* 2, 44–52. doi: 10.9735/0975-3087.2.1.44-52
- Kang, M. J., Kim, S. Y., Na, D. L., Kim, B. C., Yang, D. W., Kim, E.-J., et al. (2019). Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. *BMC Med. Inform. Decis. Mak.* 19:231. doi: 10.1186/s12911-019-0974-x
- Khan, T., and Jacobs, P. (2020). Prediction of mild cognitive impairment using movement complexity. *IEEE J. Biomed. Health Inform.* 24:1. doi: 10.1109/JBHI.2020.2985907
- Kingma, D. P., and Ba, J. (2015). Adam: a method for stochastic optimization. *arXiv [Preprint]*. arXiv:1412.6980.
- Klein (2000). Inhibition of return. *Trends Cogn. Sci.* 4, 138–147. doi: 10.1016/S1364-6613(00)01452-2
- Koedam, E. L. G. E., Lauffer, V., van der Vlies, A. E., van der Flier, W. M., Scheltens, P., and Pijnenburg, Y. A. L. (2010). Early-versus late-onset Alzheimer's disease. *Alzheimers Dis.* 19, 1401–1408. doi: 10.3233/JAD-2010-1337
- Landi, I., Glicksberg, B. S., Lee, H.-C., Cherng, S., Landi, G., Danielelto, M., et al. (2020). Deep representation learning of electronic health records to unlock patient stratification at scale. *NPJ Digit. Med.* 3:96. doi: 10.1038/s41746-020-0301-z
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Lee, G., Nho, K., Kang, B., Sohn, K.-A., Kim, D., and for Alzheimer's Disease Neuroimaging Initiative (2019b). Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* 9:1952. doi: 10.1038/s41598-018-37769-z
- Lee, G. C., Huang, P.-W., Xie, Y.-R., and Pai, M.-C. (2019a). "Classification of Alzheimer's disease, mild cognitive impairment, and cognitively normal based on neuropsychological data via supervised learning," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (Kochi: IEEE), 1808–1812. doi: 10.1109/TENCON.2019.8929443
- Lemos, L., Silva, D., Guerreiro, M., Santana, I., de Mendonça, A., Tomás, P., et al. (2012). Discriminating Alzheimer's disease from mild cognitive impairment using neuropsychological data. *Age* 70:73.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., et al. (2017). Feature selection: a data perspective. *ACM Comput. Surv.* 50, 1–45. doi: 10.1145/3136625
- Lin, C.-H., Chiu, S.-I., Chen, T.-F., Jang, J.-S. R., and Chiu, M.-J. (2020). Classifications of neurodegenerative disorders using a multiplex blood biomarkers-based machine learning model. *Int. J. Mol. Sci.* 21:6914. doi: 10.3390/ijms21186914
- Locascio, J. J., Growdon, J. H., and Corkin, S. (1995). Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. *Arch. Neurol.* 52, 1087–1099. doi: 10.1001/archneur.1995.00540350081020
- Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Calvo, P. M., Mekyska, J., Beitia, B., Barroso, N., et al. (2018). Advances on automatic speech analysis for early detection of Alzheimer's disease: a non-linear multi-task approach. *Curr. Alzheimer Res.* 15, 139–148. doi: 10.2174/1567205014666171120143800
- Lu, D., Popuri, K., Ding, G. W., Balachandar, R., Beg, M. F., and Alzheimer's Disease Neuroimaging Initiative (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci. Rep.* 8:5697. doi: 10.1038/s41598-018-22871-z
- Mayr, A., Binder, H., Gefeller, O., and Schmid, M. (2014). The evolution of boosting algorithms. *Methods Inf. Med.* 53, 419–427. doi: 10.3414/ME13-01-0122
- Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* 11, 2287–2322. doi: 10.5555/1756006.1859931
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr, Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- Müller, H. J., and von Mühlenen, A. (1996). Attentional tracking and inhibition of return in dynamic displays. *Percept. Psychophys.* 58, 224–249. doi: 10.3758/BF03211877
- Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T., et al. (2020). Automatic detection of cognitive impairments through acoustic analysis of speech. *Curr. Alzheimer Res.* 17, 60–68. doi: 10.2174/1567205017666200213094513
- Nogueira, J., Freitas, S., Duro, D., Almeida, J., and Santana, I. (2018). Validation study of the Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog) for the Portuguese patients with mild cognitive impairment and Alzheimer's disease. *Clin. Neuropsychol.* 32, 46–59. doi: 10.1080/13854046.2018.1454511
- Nori, V. S., Hane, C. A., Sun, Y., Crown, W. H., and Bleicher, P. A. (2020). Deep neural network models for identifying incident dementia using claims and EHR datasets. *PLoS ONE* 15:e0236400. doi: 10.1371/journal.pone.0236400
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195
- Pellegrini, E., Ballerini, L., Hernandez, M. D. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., et al. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement.* 10, 519–535. doi: 10.1016/j.jad.2018.07.004
- Perry, R. J., and Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease. A critical review. *Brain* 122, 383–404. doi: 10.1093/brain/122.3.383
- Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cogn. Neuropsychol.* 2, 211–228. doi: 10.1080/02643298508252866
- Posner, M. I., and Cohen, Y. (1984). "Components of visual orienting," in *Attention and performance X*, eds H. Bouma, and D. G. Bouwhuis (Hillsdale, NJ: Lawrence Erlbaum), 531–556.
- Possin, K. L., Filoteo, J. V., Song, D. D., and Salmon, D. P. (2009). Space-based but not object-based inhibition of return is impaired in Parkinson's disease. *Neuropsychologia* 47, 1694–1700. doi: 10.1016/j.neuropsychologia.2009.02.006
- Pratt, J., and McAuliffe, J. (1999). Examining the effect of practice on inhibition of return in static displays. *Percept. Psychophys.* 61, 756–765. doi: 10.3758/BF03205543
- Punjabi, A., Martersteck, A., Wang, Y., Parrish, T. B., Katsaggelos, A. K., and Alzheimer's Disease Neuroimaging Initiative (2019). Neuroimaging modality fusion in Alzheimer's classification using convolutional neural networks. *PLoS ONE* 14:e0225759. doi: 10.1371/journal.pone.0225759
- Qiu, S., Chang, G. H., Panagia, M., Gopal, D. M., Au, R., and Kolachalama, V. B. (2018). Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimers Dement.* 10, 737–749. doi: 10.1016/j.jad.2018.08.013
- Rubinsteyn, A. (2020). *Multivariate Imputation and Matrix Completion Algorithms Implemented in Python*. Available online at: <https://github.com/iskandr/fancyimpute> (accessed January 20, 2020).
- Rutkowski, T. M., Abe, M. S., Koculak, M., and Otake-Matsuura, M. (2020). Classifying mild cognitive impairment from behavioral responses in emotional arousal and valence evaluation task - ai approach for early dementia biomarker in aging societies. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2020, 5537–5543. doi: 10.1109/EMBC44109.2020.9175805
- Santos, W. P., Souza, R. E., Silva, A. F. D., and Santos Filho, P. B. (2008). Evaluation of Alzheimer's disease by analysis of MR images using multilayer perceptrons and committee machines. *Comput. Med. Imaging Graphics* 32, 17–21. doi: 10.1016/j.compmedimag.2007.08.004
- Satel, J., Wilson, N. R., and Klein, R. M. (2019). What neuroscientific studies tell us about inhibition of return. *Vision* 3:58. doi: 10.3390/vision3040058
- Seidel Malkinson, T., and Bartolomeo, P. (2018). Fronto-parietal organization for response times in inhibition of return: the fortior model. *Cortex* 102, 176–192. doi: 10.1016/j.cortex.2017.11.005
- Shariat Torbaghan, S., Yazdi, D., Mirpour, K., and Bisley, J. W. (2012). Inhibition of return in a visual foraging task in non-human subjects. *Vision Res.* 74, 2–9. doi: 10.1016/j.visres.2012.03.022
- Suk, H.-I., Lee, S.-W., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2014). Hierarchical feature representation and multimodal fusion

- with deep learning for AD/MCI diagnosis. *Neuroimage* 101, 569–582. doi: 10.1016/j.neuroimage.2014.06.077
- Tales, A., Snowden, R. J., Haworth, J., and Wilcock, G. (2005). Abnormal spatial and non-spatial cueing effects in mild cognitive impairment and Alzheimer's disease. *Neurocase* 11, 85–92. doi: 10.1080/13554790490896983
- Tales, A., Snowden, R. J., Phillips, M., Haworth, J., Porter, G., Wilcock, G., et al. (2011). Exogenous phasic alerting and spatial orienting in mild cognitive impairment compared to healthy ageing: study outcome is related to target response. *Cortex* 47, 180–190. doi: 10.1016/j.cortex.2009.09.007
- Tufail, A. B., Abidi, A., Siddiqui, A. M., and Younis, M. S. (2012). Automatic classification of initial categories of Alzheimer's disease from structural MRI phase images: a comparison of PSVM, KNN and ANN methods. *Age* 2012:1731. doi: 10.5281/zenodo.1084608
- Valladares-Rodríguez, S., Fernández-Iglesias, M. J., Anido-Rifón, L., Facal, D., Rivas-Costa, C., and Pérez-Rodríguez, R. (2019). Touchscreen games to detect cognitive impairment in senior adults. A user-interaction pilot study. *Int. J. Med. Inform.* 127, 52–62. doi: 10.1016/j.ijmedinf.2019.04.012
- Wallert, J., Westman, E., Ulinder, J., Annerstedt, M., Terzis, B., and Ekman, U. (2018). Differentiating patients at the memory clinic with simple reaction time variables: a predictive modeling approach using support vector machines and bayesian optimization. *Front. Aging Neurosci.* 10:144. doi: 10.3389/fnagi.2018.00144
- Weng, S. F., Reys, J., Kai, J., Garibaldi, J. M., and Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE* 12:e0174944. doi: 10.1371/journal.pone.0174944
- Williams, J. A., Weakley, A., Cook, D. J., and Schmitter-Edgecombe, M. (2013). "Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia," in *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence* (Bellevue, WA), 71–76.
- Ye, B. S., Seo, S. W., Lee, Y., Kim, S. Y., Choi, S. H., Lee, Y. M., et al. (2012). Neuropsychological performance and conversion to Alzheimer's disease in early- compared to late-onset amnesic mild cognitive impairment: CREDOS study. *Dement. Geriatr. Cogn. Disord.* 34, 156–166. doi: 10.1159/000342973
- Zhou, T., Thung, K.-H., Zhu, X., and Shen, D. (2019). Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Hum. Brain Mapp.* 40, 1001–1016. doi: 10.1002/hbm.24428

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

Copyright © 2020 Almubark, Chang, Shattuck, Nguyen, Turner and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.