Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls with Multilayer Perceptron Neural Network and Neuropsychological Test Data

Ibrahim Almubark[1], Samah Alsegehy[2], Xiong Jiang[3] and Lin-Ching Chang[1,*]
[1]*Department of Electrical Engineering and Computer Science, Catholic University of America, Washington, DC, U.S.A.*
[2]*Department of Computer Science and Engineering, Penn State University, University Park, PA, U.S.A.*
[3]*Department of Neuroscience, Georgetown University Medical Center, Washington, DC, U.S.A.*

Keywords: Multilayer Perceptron Neural Network, Alzheimer's Disease, Mild Cognitive Impairment, Neuropsychological Test.

Abstract: Recent advances in machine learning have shown outstanding performances in biological and medical data analysis to assist for early detection, diagnosis, and treatment of diseases. Alzheimer's disease (AD) is a neurodegenerative disease and the most common cause of dementia in older adults. In this study, multilayer perceptron (MLP) neural networks are developed to classify AD, Mild Cognitive Impairment (MCI), and Cognitive Normal (CN) subjects based upon the data from standard neuropsychological tests. Three neuropsychological tests from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog), Mini-Mental State Examination (MMSE), and Functional Activities Questionnaire (FAQ), were used to train MLP neural networks. We first build three MLP models that can classify AD vs. CN, AD vs. MCI, and MCI vs. CN. We then construct a 3-way MLP classifier to classify AD vs. MCI vs. CN. Finally, we propose a cascade 3-way classification method to further improve the model performance. Using the neuropsychological test data from ADNI database, our result shows the pairwise MLP models (i.e., AD vs. CN, AD vs. MCI, and MCI vs. CN) have the accuracy of 99.76±0.48, 89.64±3.94, and 90.81±2.91, respectively. The multi-class MLP model has an average accuracy of 84.28±3.66, and the proposed cascaded MLP approach further improves the performance of the multi-class classification with an average accuracy of 86.26±3.15.

1 INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease and the most common cause of dementia in older adults. AD pathologies often start 5, 10, or even 20 years before symptoms appear (Alzheimer's Association, 2020). Symptoms usually start with difficulty remembering new information. Since this condition is also common with the normal aging process, distinguishing between early AD and normal aging can be difficult (Holtzman et al., 2011). In advanced stages, symptoms include confusion, mood and behavior changes, and inability to care for one's self and perform basic life tasks. AD is ultimately fatal (Taylor et al., 2017). While significant progress has been made, there are yet no proven effective treatments for AD. As a result, there is increasing pressure to develop techniques to assist in the diagnosis of early AD, as early intervention may be most effective in treating or slowing disease progress. In addition, early diagnosis may provide useful information for the development of more effective treatments (Fiandaca et al., 2014).

Three groups of subjects are included in this machine learning classification study: Cognitively Normal (CN) older adults, Mild Cognitive Impairment (MCI) due to AD, and AD. A CN subject has no signs of cognitive impairment other than age-related normal decline. A subject converts from CN to MCI when symptoms become mild yet noticeable to the patient or close family members/friends. MCI is a transitional stage between CN and AD, and is the earliest clinically detectable stage of progression towards dementia or AD (Sperling et al., 2011). Approximately 15-20% of seniors age 65 or older

[*] For the Alzheimer's Disease Neuroimaging Initiative

Almubark, I., Alsegehy, S., Jiang, X. and Chang, L.
Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls with Multilayer Perceptron Neural Network and Neuropsychological Test Data.
DOI: 10.5220/0010143304390446
In *Proceedings of the 12th International Joint Conference on Computational Intelligence (IJCCI 2020)*, pages 439-446
ISBN: 978-989-758-475-6

have MCI (Alzheimer's Association, 2020). Patients diagnosed with MCI are of higher risk for developing AD or other types of dementia, and are therefore given special attention (Ansart et al., 2020). AD is the last stage in this progression.

In recent years, the world has seen many major breakthroughs in the field of healthcare because of the rapid proliferation of large biomedical datasets, concurrent with advances in machine learning, especially in deep learning (Esteva et al., 2017). These advances have opened new avenues for the development of diagnostic tools to assist early detection of AD. Recently, several studies have focused on the detection of different cognitive groups by utilizing various types of biomedical data including brain imaging data (Pellegrini et al., 2018), cerebrospinal fluid (CSF) specimens (Jack et al., 2018; Shaffer et al., 2013), and behavioral data from speech (Fraser et al., 2016; Nagumo et al., 2020), body movement (Khan & Jacobs, 2020), and neuropsychological tests (Grassi et al., 2019; Kang et al., 2019; Lee et al., 2019).

Standard neuropsychological tests are typically used in the diagnosis of cognitive impairment in individuals with MCI, AD, or other neurological conditions (Seo, 2018). These tests are less expensive, easy to conduct, and widely available compared to medical imaging. The scores from these tests can measure normal and abnormal cognitive and behavioral functions and provide useful features to machine learning methods for the early detection of AD (Anastasi & Urbina, 1997). Repeated assessment with these tests is frequently used to evaluate changes in a treated person's condition over time (Harvey, 2012). The existence of multiple cognitive deficits indicates that a combination of neuropsychological tests from different domains may improve clinical diagnosis accuracy (Harvey, 2012; Storey & Kinsella, 2007; Yeatts et al., 2018).

In this paper, we present fully connected multilayer perceptron (MLP) networks to perform binary classification between different cognitive groups (i.e., AD vs. CN, AD vs. MCI, and MCI vs. CN) using the baseline visit data from three neuropsychological tests in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). A direct MLP based 3-way classification (i.e., AD vs. MCI vs. CN) is also developed. Additionally, we propose a MLP based cascading approach to further improve the multi-class classification performance.

The paper is organized as follows. Section 2 includes descriptions of data used in this study, data pre-processing, experimental design, and proposed methods. Section 3 includes results from several multilayer perceptron models. Section 4 includes a summary of results, discussion, and future research directions.

2 MATERIALS AND METHODS

2.1 Data

2.1.1 ADNI Database

The data used in this study was obtained from the ADNI database. ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD progression. The ADNI project draws on a broad range of academic institutions and private corporations, with subjects recruited from over 50 sites across the U.S. and Canada. The project began in 2003 and has been extended to different phases. The first phase of ADNI (ADNI-1) was completed in 2010 and has been followed by ADNI-GO, ADNI-2, and ADNI-3. These four protocols have recruited over 1900 adults, with ages from 55 to 90, and consist of elderly CN controls, people with MCI, and people with AD. The follow-up duration of each group is described in the protocols for ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. For detailed information, please see (www.adni-info.org).

2.1.2 Subjects

In this study, we used the baseline visit data from a total of 808 subjects at the initial project period (ADNI-1), including 188 AD, 391 MCI, and 229 CN. The enrolled subjects were between 55-90 (inclusive) years of age, in good general health, having a partner who is able to provide an independent assessment of the subject's functioning, having at least 6 grades of education or work history, and were fluent in English or Spanish. All subjects and their study partners completed the informed consent process, and study protocols were reviewed and approved by the Institutional Review Board at each ADNI data collection site (Petersen et al., 2010). Table 1 shows the characteristics of the AD, MCI, and CN subjects included in this study. The mean test score was computed by averaging the scores from all the questions in one test.

Table 1: Characteristics of subjects at their baseline visit recruited during ADNI-1.

Characteristic	AD (n=188)	MCI (n=391)	CN (n=229)	*p*-value
Age, years	74.9±7.4	74.4±7.3	75.4±5.0	0.167
Education, years	14.7±3.1	15.6±3.0	16.1±2.8	3.98×10^{-5}
Sex, male/female	97/91	255/136	119/110	5.43×10^{-4}
ADAS-Cog score	28.9±7.6	18.6±6.3	9.5±4.2	4.95×10^{-145}
MMSE score	23.4±2.0	27.0±1.8	29.1±0.9	3.08×10^{-164}
FAQ score	13.0±6.8	3.8±4.5	0.1±0.6	4.07×10^{-28}

Values are shown as mean ± standard deviation or gender ratios. The *p*-values for differences between AD, MCI, and CN are based on t-test. ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognitive subscale; MMSE = Mini-Mental State Examination; FAQ = Functional Activities Questionnaire.

2.1.3 Neuropsychological Data

The itemized scores of three neuropsychological tests were used, which include Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) (Rosen et al., 1984), Mini-Mental State Examination (MMSE) (Folstein et al., 1975), and Functional Activities Questionnaire (FAQ) (Pfeffer et al., 1982). Table 2 details the cognitive/daily functions associated with individual tests. In total, there are 13, 30, and 10 questions from ADAS-Cog (note that Q13 was not used by default), MMSE, and FAQ, respectively. The score of each individual question is treated as a feature in our machine learning task, resulting in a total of 13, 30, and 10 features for the ADAS-Cog, MMSE, and FAQ datasets. The three neuropsychological tests are widely used to assist cognitive impairment in AD. The most prominent feature of AD is memory impairment. Therefore, ADAS-Cog and MMSE tests were used in this study. ADAS-Cog and MMSE also test global cognitive function as well as several domains other than memory (Casanova et al., 2013). Information on function from the FAQ test was also included as functional changes begin to appear earlier in the dementia process (John et al., 2016).

Table 2: Neuropsychological tests used in this study.

Neuropsychological Tests	
ADAS-Cog	Registration (3)
Q1. Word recall	Attention and calculation (5)
Q2. Word recognition	Recall (3)
Q3. Object naming	Language (8)
Q4. Recall test instructions	Visual construction (1)
Q5. Orientation	**FAQ**
Q6. Commands	Q1. Manage finances
Q7. Clarity of language	Q2. Complete forms
Q8. Comprehension	Q3. Shop
Q9. Word finding	Q4. Perform games of skill or hobbies
Q10. Ideational praxis	Q5. Prepare hot beverages
Q11. Constructional praxis	Q6. Prepare a balanced meal
Q12. Delayed word recall	Q7. Follow current events
Q14. Number cancellation	Q8. Attend to TV, books, or magazines
MMSE	Q9. Remember appointments
Orientation (10)	Q10. Travel out of the neighborhood

2.2 Experimental Design for Classification

Our multilayer perceptron (MLP) neural networks were developed using (1) the original set of features from each neuropsychological test (i.e., 13 from ADAS-Cog, 30 from MMSE, and 10 from FAQ), and (2) the combined-test of 53 features from three neuropsychological tests. We first trained three binary MLP models to perform binary classification between different cognitive groups for both the individual tests and the combined-test: (Case 1) AD vs. CN, (Case 2) AD vs. MCI, and (Case 3) MCI vs. CN. We also trained a MLP network with multi-class classification network, i.e., AD vs. MCI vs. CN (case 4). Furthermore, we proposed a MLP based cascading approach to further improve the multi-class classification performance (case 5).

The proposed cascade MLP method is composed of 2 steps. In the first step, we classified CN vs. (AD + MCI) with a MLP network. Then we trained another MLP network with the predicted AD or MCI samples from the first step to classify AD vs. MCI (step 2). Note that the true CN samples misclassified in step 1 as (AD + MCI) will participate in the second step and they will be counted as misclassified CN regardless of their predicted results in step 2. On the other hand, the true MCI or AD samples misclassified as CN will not participate in the training in step 2, but will be counted in the final 3×3 confusion matrix along with the correctly classified CN samples after step 1.

The implementation was carried out using *Python* and related libraries including *Scikit-learn*, *Pandas*, *Numpy*, *TensorFlow*, and *Keras* (Chollet, 2015; Pedregosa et al., 2011).

2.3 Data Pre-processing

The original baseline visit dataset in ADNI-1 consists of 200 AD, 400 MCI, and 229 CN. We excluded 12 subjects from AD and 9 subjects from MCI before our data analysis since answers to some questions were

not recorded (i.e., missing values in some features). Therefore, the dataset (n=808, AD=188, MCI=391, CN=229) used in our study does not have any missing values. In statistics it is a common practice to drop cases with missing values as long as the sample size is sufficiently large and the number of dropped cases does not exceed 5% of the overall sample. In this study, the number of missing value subjects was relatively small (12 out of 200 for AD and 9 out of 400 for MCI). Therefore, we elected to drop without replacement the twenty-one samples with missing value. Feature normalization was performed by standard scaling with a zero mean and standard deviation equal to one.

2.4 Data Partitioning

Cross-validation (CV) with stratified K-Fold was used to evaluate the predictive model. Data was divided into 5 disjoint subsets with consistent ratios between classes in each fold. Eighty percent of the data was used in training and 20% of the data was used for testing in each fold.

2.5 Multilayer Perceptron (MLP) Neural Network

Multilayer perceptron (MLP) is a feed-forward artificial neural networks that uses back-propagation to update weights (Marsland, 2015). The neurons are connected to later layers in a way that pushes information from the input, through hidden layer(s), to the output layer. MLP leverages a layered architecture of stacked perceptrons to solve complex, often supervised, problems. MLP can approximate non-linear functions for both classification and regression (Joshi, 2020).

In this paper, we developed multiple fully connected MLP networks to classify different cognitive groups (AD, MCI, and CN) using data from three neuropsychological tests. Figure 1 shows a 3-layer MLP network that can be used to classify AD and CN subjects. As an example, if the combined-test data was used, the resulted MLP model will have 53 nodes in the input layer, 6 nodes in the hidden layer, and 1 node in the output layer. The Rectified Linear Unit (ReLU) was selected as the activation function for the input and hidden layers (Xu et al., 2015). The sigmoid function and the binary cross-entropy loss function were used for the binary classifications. The softmax function and the multi-class cross-entropy were used for the multi-class classification (Sharma, 2017). The Adaptive Movement Estimation (Adam) was used as our optimizer to tune the network during the training (Kingma & Ba, 2014).

The cross-entropy loss function defined in equation (1) is used to quantify MLP model errors:

$$\mathcal{L} = -\sum_{ic} w_c p(c|x_i) \log q(c|x_i) + \lambda_k \sum_{k} ||w_{k,k-1}||^2 \quad (1)$$

where c, i, k are indices for classes, samples, and layers, respectively; w_c is the class weight for class c; $p(c|x_i)$ is the true probability for sample x_i to be assigned to class c; $q(c|x_i)$ is the predicted probability for sample x_i to be assigned to class c; λ_k is the regularization strength for layer k; and $w_{k,k-1}$ are the weights between the $(k-1)^{th}$ and the k^{th} layer. A larger class weight on class c will penalize more if samples in class c are misclassified. We also considered the probability threshold as another hyper-parameter. For example, a sample can be predicted to have a probability of 0.45 to be class 1 and 0.55 to be class 0, if we set this probability at default 0.5, this sample will be classified to class 0. However, one could set the probability threshold to 0.4 instead, the sample will then be classified to class 1. In our training, we also included the class weights and probability thresholds as hyper-parameters to avoid the imbalance issue between the model sensitivity and specificity.

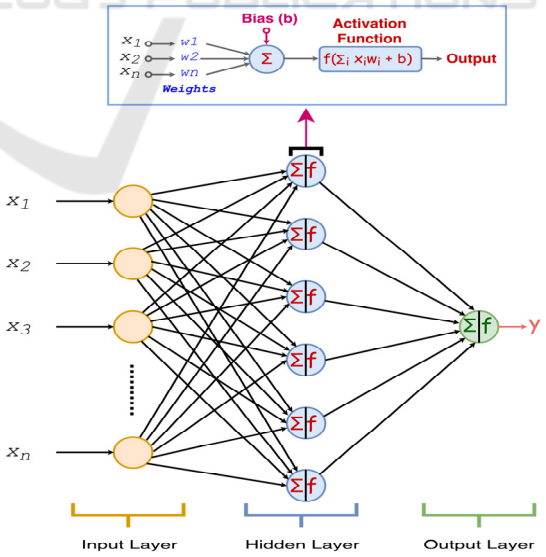

Figure 1: A 3-layer MLP network. Circular nodes represent artificial neurons. Arrows represent connection from a neuron output to a neuron input.

To avoid overfitting, we have employed L2 regularization in the hidden layers as shown in equation (1). The regularization strength for each layer λ_k was also tuned as a hyper-parameter. We also applied the *EarlyStopping* function in *Keras* by observing the loss function on the validation set. If the loss function reduction was less than (1×10^{-4}) for 5 consecutive epochs, the training will be stopped. Learning rate (or shrinkage factor) was adjusted based on the reduction rate of loss function to train the network more efficiently, and the *ReduceLROnPlateau* function in *Keras* was used to observe the loss function reduction. If the loss function reduction was less than (1×10^{-3}) for 10 consecutive epochs, the learning rate would be reduced to one tenth of the previous learning rate or the minimum learning rate. The initial learning rate was set to 0.01 and the minimum value is (5×10^{-4}). The early stopping and the learning rate shrinkage helped our training process with a positive impact in the classification performance.

2.6 Performance Evaluation

To evaluate the performance of the classifiers, sensitivity, specificity, and accuracy were calculated for each model. Sensitivity measures the ratio of actual positive subjects to the total numbers of subjects identified by the test as being positive, i.e., true positive rate. Specificity measures the ratio of actual negative subjects to the total number of subjects testing negative, i.e., true negative rate. Accuracy is the ratio of correctly classified subjects to the entire set of subjects. In other words, sensitivity, specificity, and accuracy are described in terms of TP (True Positives), TN (True Negatives), FN (False Negatives), and FP (False Positives), and defined in equations (2), (3), and (4), respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

To illustrate the diagnostic ability of a classifier, we also calculated the Area Under the Curve (AUC) from the Receiver Operating Characteristic (ROC) curve. The AUC was also used as the target metric during hyper-parameter tuning. An algorithm with an AUC closer to 1, indicating a near perfect performance, is considered as the more reliable predictive model.

3 RESULTS AND DISCUSSION

Table 3 summarizes the performance of binary and 3-way classification using each individual neuropsychological test (i.e., ADAS-Cog, MMSE, and FAQ) as well as a combination of these three tests to discriminate different cognitive groups. The default class weights was set to 1:1 for binary classifier or 1:1:1 for 3-way classifier, and the default probability threshold was set to 0.5. As shown in Table 3, the model using the combined features from three tests outperformed the models using each single test. The classification of AD vs. CN subjects had very high accuracies (98%~100%) in all models. This indicates the proposed MLP method using neuropsychological test data is very effective in classifying AD and CN. The classification accuracy of MCI vs. CN was 77%~82% when using a single test, but reached 90% when using the combined tests. Classification between AD and MCI was our most challenging task. Although the overall accuracy was acceptable (80%~84%), the sensitivity was very poor (47%~69%) when using a single neuropsychological test. However, the sensitivity significantly improved, to 81.38%, when using the combined-test.

In Table 4, we demonstrated that class weights and probability thresholds could be used to improve the model performance and obtain a more balanced sensitivity and specificity ratio. By tuning the class weights and probability thresholds in training of MLP networks, the sensitivity as well as the accuracy can be further improved. For example, the sensitivity of AD vs. MCI improved from 46.81% to 79.26% with ADAS-Cog test (data not shown) and from 81.38% to 91.49% with the combined-test. The accuracy of AD vs. MCI vs. CN was improved from 82.43% to 84.28%. While this MLP 3-way classification accuracy is notably lower than the binary classifications, it outperformed other existing methods. For example, its accuracy was 22% higher than Lee's model (Lee et al., 2019).

Table 5 shows the performance of the MLP model with the cascade approach using the combined-test. The class weights and probability thresholds were tuned to obtain optimal model performance. This new model further improved the results compared to the direct 3-way classification (Tables 3 and 4) in terms of sensitivity, specificity, and accuracy.

Standard neuropsychological tests are often incorporated into regular physical examinations for

seniors, this study demonstrated that early screening of AD is possible when using these tests with neural networks. The proposed methods are effective to classify different cognitive groups, and do not require the medical procedures that are presently more expensive, invasive, or not offered in many clinical settings. These medical procedures include neuroimaging, cerebrospinal fluid (CSF), and genetic testing. This study also showed that a combination of a variety of neuropsychological tests and assessments used for AD diagnosis improved the accuracy of clinical diagnosis. Lastly, this study points to the potential for MLP neural network enabled classifiers in discriminating between AD progression classes.

Table 3: The classification performance of binary and 3-way multilayer perceptron (MLP) networks using data from a single neuropsychological test and the combined-test. The default class weights (1:1 or 1:1:1) and probability thresholds (0.5) were used in each model.

Dataset	Classification Case	SEN%	SPE%	ACC%	AUC
ADAS-Cog (13)	AD vs. CN	93.62	98.69	96.04 ± 1.71	0.994
	AD vs. MCI	46.81	95.43	79.73 ± 2.29	0.874
	MCI vs. CN	82.23	80.35	81.54 ± 5.24	0.905
	AD vs. MCI vs. CN	72.75	80.90	72.75 ± 2.82	0.887
MMSE (30)	AD vs. CN	95.85	97.82	96.92 ± 1.62	0.998
	AD vs. MCI	69.43	92.19	84.75 ± 3.75	0.914
	MCI vs CN	78.34	75.98	77.48 ± 3.43	0.857
	AD vs. MCI vs. CN	70.21	80.28	70.21 ± 5.89	0.873
FAQ (10)	AD vs. CN	90.16	99.56	95.26 ± 1.52	0.982
	AD vs. MCI	54.40	92.89	80.24 ± 3.50	0.868
	MCI vs. CN	71.83	92.14	79.29 ± 2.26	0.845
	AD vs. MCI vs. CN	71.32	85.00	71.33 ± 3.74	0.853
Combined-Test (53)	**AD vs. CN**	**98.04**	**100.00**	**99.28 ± 0.59**	**1.0**
	AD vs. MCI	**81.38**	**93.35**	**89.46 ± 3.93**	**0.964**
	MCI vs. CN	**90.79**	**89.08**	**90.16 ± 3.55**	**0.960**
	AD vs. MCI vs. CN	**82.43**	**88.62**	**82.43 ± 3.92**	**0.946**

Table 4: The classification performance of binary and 3-way multilayer perceptron (MLP) networks using data from the combined three neuropsychological tests. The class weights and probability thresholds were tuned during the training to obtain a balanced sensitivity and specificity ratio.

Classification Case	Probability Threshold	Class Weight	SEN%	SPE%	ACC%	AUC
AD vs. CN	0.5	1:1.5	99.47	100.00	99.76 ± 0.48	1.0
AD vs. MCI	0.4	1:1.5	91.49	88.75	89.64 ± 3.94	0.965
MCI vs. CN	0.5	1:1.5	92.07	88.65	90.81 ± 2.91	0.964
AD vs. MCI vs. CN	0.5	1.5:1.5:1	84.28	90.36	84.28 ± 3.66	0.954

Table 5: The multilayer perceptron (MLP) cascading classification performance. The classification performance of the tuned class weights and probability thresholds are shown.

Classification steps	Probability Threshold	Class Weight	SEN%	SPE%	ACC%	AUC
Step1: CN vs. (AD + MCI)	0.5	1:1	93.27	92.14	92.95 ± 2.33	0.973
Step2: AD vs. MCI using the (AD + MCI) from step 1	0.6	1:1	86.26	91.15	86.26 ± 3.15	0.957

4 CONCLUSIONS AND FUTURE WORK

Most previous studies on AD detection using machine learning techniques have been focusing on utilizing brain imaging data. With the rich availability and low cost of standard neuropsychological tests, this paper investigated the classification performance for detecting different cognitive groups (AD, MCI, and CN) using MLP neural networks. Several important conclusions can be drawn from this study. First, using a single neuropsychological test to classify AD and MCI yielded a very poor sensitivity. Second, the combination of three neuropsychological test data with MLP networks showed good potential for early AD detection. The MLP classifiers performed well on all three binary cases with the combined-test as well as for the 3-way classification. Finally, the proposed cascade MLP approach can further improve the performance of multi-class classification.

The proposed method is not only reliable but also cost effective, and therefore it can support large-scale cognitive screening. Our future work will include identifying individuals with MCI who would be more likely to develop AD within a defined period of time. Additionally, we will investigate other artificial neural networks on diagnostics and prediction of AD. We also plan to study the combination of brain imaging and behavioral data with both machine learning and deep learning techniques that may offer additional insights into the progression of various stages of AD.

ACKNOWLEDGEMENTS

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

REFERENCES

Alzheimer's Association. (2020). 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *16*(3), 391–460.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.

Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronne, R., Faouzi, J., Koval, I., & Louis, M. (2020). Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. *MedRxiv*.

Casanova, R., Hsu, F.-C., Sink, K. M., Rapp, S. R., Williamson, J. D., Resnick, S. M., Espeland, M. A., & Initiative, A. D. N. (2013). Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One*, *8*(11), e77949.

Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. *URL: Https://Keras. Io/k*, *7*(8), T1.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Fiandaca, M. S., Mapstone, M. E., Cheema, A. K., & Federoff, H. J. (2014). The critical need for defining preclinical biomarkers in Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *10*(3 Suppl), S196-212. https://doi.org/10.1016/j.jalz.2014.04.015

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease: JAD*, *49*(2), 407–422. https://doi.org/10.3233/JAD-150520

Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., & Dumontier, M. (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. *Frontiers in Neurology*, *10*, 756.

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, *14*(1), 91.

Holtzman, D. M., Morris, J. C., & Goate, A. M. (2011). Alzheimer's disease: The challenge of the second century. *Science Translational Medicine*, *3*(77), 77sr1–77sr1.

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Contributors. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

John, S. E., Gurnani, A. S., Bussell, C., Saurman, J. L., Griffin, J. W., & Gavett, B. E. (2016). The effectiveness and unique contribution of neuropsychological tests and the δ latent phenotype in the differential diagnosis of dementia in the uniform data set. *Neuropsychology*, *30*(8), 946.

Joshi, A. V. (2020). Perceptron and Neural Networks. In *Machine Learning and Artificial Intelligence* (pp. 43–51). Springer, Cham.

Kang, M. J., Kim, S. Y., Na, D. L., Kim, B. C., Yang, D. W., Kim, E.-J., Na, H. R., Han, H. J., Lee, J.-H., Kim, J. H., Park, K. H., Park, K. W., Han, S.-H., Kim, S. Y., Yoon, S. J., Yoon, B., Seo, S. W., Moon, S. Y., Yang, Y., … Youn, Y. C. (2019). Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. *BMC Medical Informatics and Decision Making*, *19*(1), 231. https://doi.org/10.1186/s12911-019-0974-x

Khan, T., & Jacobs, P. (2020). Prediction of mild cognitive impairment using movement complexity. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2020.2985907

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

Lee, G. G. C., Huang, P.-W., Xie, Y.-R., & Pai, M.-C. (2019). Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Normal Based on Neuropsychological Data via Supervised Learning. *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, 1808–1812.

Marsland, S. (2015). *Machine learning: An algorithmic perspective* (2nd ed). CRC press.

Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T., Sumi, S., Kurita, S., Nakakubo, S., Lee, S., Doi, T., & Shimada, H. (2020). Automatic Detection of Cognitive Impairments through Acoustic Analysis of Speech. *Current Alzheimer Research*, *17*(1), 60–68. https://doi.org/10.2174/1567205017666200213094513

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pellegrini, E., Ballerini, L., Hernandez, M. D. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T. J., Trucco, E., & Wardlaw, J. M. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, *10*, 519–535. https://doi.org/10.1016/j.dadm.2018.07.004

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., & Harvey, D. J. (2010). Alzheimer's disease neuroimaging initiative (ADNI). *Neurology*, *74*(3), 201–209.

Pfeffer, R. I., Kurosaki, T. T., Harrah Jr, C. H., Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, *37*(3), 323–329.

Rosen, W., Mohs, R., & Davis, K. (1984). Alzheimer's Disease Assessment Scale—Cognitive and Non-Cognitive Sections (ADAS-Cog, ADAS Non-Cog). *Journal of Psychiatry*, *141*, 1356–1364.

Seo, E. H. (2018). Neuropsychological assessment of dementia and cognitive disorders. *Journal of Korean Neuropsychiatric Association*, *57*(1), 2–11.

Shaffer, J. L., Petrella, J. R., Sheldon, F. C., Choudhury, K. R., Calhoun, V. D., Coleman, R. E., Doraiswamy, P. M., & Initiative, A. D. N. (2013). Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology*, *266*(2), 583–591.

Sharma, S. (2017). Activation functions in neural networks. *Towards Data Science*, *6*(12), 310–316.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr, C. R., Kaye, J., & Montine, T. J. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 280–292.

Storey, E., & Kinsella, G. J. (2007). Principles of neuropsychometric assessment. In *Neurology and Clinical Neuroscience* (pp. 22–30). Mosby.

Taylor, C. A., Greenlund, S. F., McGuire, L. C., Lu, H., & Croft, J. B. (2017). Deaths from Alzheimer's Disease—United States, 1999–2014. *MMWR. Morbidity and Mortality Weekly Report*, *66*(20), 521.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *ArXiv Preprint ArXiv:1505.00853*.

Yeatts, S. D., Palesch, Y. Y., & Temkin, N. (2018). Biostatistical issues in TBI clinical trials. In *Handbook of Neuroemergency Clinical Trials* (pp. 167–185). Elsevier.