Early Prediction of Alzheimer's Disease with a Multimodal Multitask Deep Learning Model

Nitin Seshadri[1], Rishi Shah[2], and Dr. Serena McCalla[2]

[1]Somers High School, Lincolndale, NY, USA
[2] iResearch Institute, Glen Cove, NY, USA

ABSTRACT

Alzheimer's disease (AD) is the sixth leading cause of death in the United States and the most common neurodegenerative disease in adults over 65. Early-stage AD is often misinterpreted as normal cognitive aging because it may not cause adverse symptoms or visible behavioral changes for up to 20 years. Machine learning has been used to avoid misinterpretation of data and more accurately predict the onset of AD. This study aims to use the data typically available in a clinical setting to predict the onset of AD while maintaining a high level of accuracy. This study proposes a deep learning model that uses multimodal input data and performs multitask classification to predict AD diagnosis and scores of two commonly used cognitive assessments: Alzheimer's Disease Assessment Scale (ADAS) and Mini-Mental State Examination (MMSE). The model was validated using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset of 1737 patients. The current model achieved a greater accuracy in predicting AD diagnosis and a lower error in predicting ADAS and MMSE scores than existing state-of-the-art models. This model can be applied to the clinical setting so that accurate diagnosis can be achieved, and appropriate action can be taken. Future investigations could include using a convolutional neural network (CNN) to process data from clinical images directly or training and validating the model with other clinical datasets to further improve its accuracy.

Introduction

Alzheimer's disease (AD) is the sixth leading cause of death in the United States, and the most common neurodegenerative disease in the elderly population [1]. AD symptomology, which includes neuronal decay and brain atrophy, causes a significant decline in cognitive functions over time such as memory, recall, behavior, and language [2].

When a patient presents with symptoms of AD, image and clinical data are collected to diagnose the patient and monitor the progression of the disease. Image data is collected by performing MRI and PET scans. Clinical data is collected by interviewing the patient, and consists of age, gender, education, apolipoprotein E4 (APOE4; the presence of this gene increases AD risk) genotype, and cognitive assessment scores, including the Mini-Mental State Examination (MMSE) or the Alzheimer's Disease Assessment Scale (ADAS) [3] [4].

Even though a significant amount of data is collected, two main issues occur: (1) Early-stage AD is often misinterpreted as normal cognitive aging because it may not cause adverse symptoms or visible behavioral changes for up to 20 years [5] and (2) there is a 99.6% failure rate of clinical trials for AD treatments [6]. To avoid misinterpretation and allow for early action to mitigate AD symptoms, a need exists for a more accurate method of detection at an earlier stage of the disease until better treatments for AD can be discovered. [7].

Review of Literature

In the medical domain, machine learning methods can identify patterns using a vast library of existing patient tests and diagnosis data and use that to make predictions for another patient [8]. Machine learning methods to predict

Alzheimer's disease have improved greatly over the past decade [8]. With improved models, misinterpretation of data can be avoided, and more accurate predictions of AD onset can be achieved with the same data a doctor collects from the patient supplemented by a large library of previous patient data.

In AD machine learning research, input data is categorized into various modalities, including clinical, MRI, PET, cognitive, and neuropathological. Previous models have incorporated one or more modalities and have attempted to predict either diagnosis, cognitive assessment scores, or both.

There are four types of machine learning models applied to AD prediction (Table 1). Unimodal single-task models [9] [10] only use one type of data (unimodal) and only predict one metric (single-task). Unimodal multitask models [11], while more complex than single-task models because they predict multiple metrics, only use one type of data, potentially limiting the accuracy of their predictions. Multimodal single-task models [12] [13], while using more varied input data than a unimodal model, only predict one metric, potentially limiting the utility of their predictions because doctors need multiple statistics to confirm a diagnosis. Multimodal multitask models [14] use multiple types of data and predict multiple metrics and are not known to suffer from any of the limitations of the prior models. With multimodal data, subtle changes in a patient can be detected, leading to a more reliable diagnosis [14].

Table 1. Comparison of previous state-of-the-art machine learning models.

Researchers	Model Type	Limitations
Ito et al. [9] Yang et al. [10]	Unimodal, Single-Task	Only uses one type of data; only predicts one metric
Zhou et al. [11]	Unimodal, Multitask	Only uses one type of data
Liu et al. [12] Qiu et al. [13]	Multimodal, Single-Task	Only predicts one metric
El-Sappagh et al. [14]	Multimodal, Multitask	Uses more data than available in clinical setting

El-Sappagh et al. [14] is the most recent implementation of multimodal multitask machine learning applied to AD, and uses over 600 features from clinical, MRI, and PET data to achieve over 90% accuracy for the diagnosis task, greater than previous models. However, in a clinical setting, not as much data is collected, which results in fewer features that can be extracted by a machine learning model [15].

Objective

The purpose of this research is to identify a multitask machine learning model that can predict AD diagnosis (DX), 13-Question Alzheimer's Disease Assessment Scale (ADAS), and Mini-Mental State Examination (MMSE) score based on multimodal data available in a clinical setting.

Hypothesis

It is hypothesized that
H1: The identified model will achieve a higher accuracy on the diagnosis task and lower RMSE values on the ADAS and MMSE tasks than current state-of-the-art models.
H0: The identified model will not achieve a higher accuracy on the diagnosis task and lower RMSE values on the ADAS and MMSE tasks than current state-of-the-art models.

Methodology

Dataset

Data used in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

The dataset contains data for 1737 patients between 50 and 95 years of age in varying stages of cognitive condition, from normal cognitive aging to mild cognitive impairment (MCI) to AD. There are 957 males and 780 females. For each patient, there are data from the initial visit, categorized as baseline, and subsequent visits every 6 months for up to 10 years. At their baseline visit, 342 patients were diagnosed with AD.

Data Processing

The dataset was processed using the NumPy, pandas, and scikit-learn Python libraries based on criteria established by The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge [7]. The features selected for the machine learning model include 13 clinical features known to correlate [6] with AD progression (Table 2), and a variable number of features obtained through Principal Component Analysis (PCA) of 328 MRI and PET features that were preprocessed by ADNI with the FreeSurfer image analysis program. To obtain an 80%-20% training-test set split, approximately 1390 patients were assigned to the training set and the remaining patients to the test set at random.

Table 2. Clinical features used from the ADNI dataset.

Feature Name	Meaning
DX_bl	Baseline diagnosis
ADAS13_bl	Baseline ADAS score
MMSE_bl	Baseline MMSE score
VISCODE	Visit code
AGE	Age
PTGENDER	Gender
PTEDUCAT	Years of education
APOE4	Number of APOE4 alleles
RAVLT_immediate	Total number of words memorized over 5 trials
RAVLT_learning	Number of words learned between trial 1 and trial 5
RAVLT_forgetting	Number of words forgotten between trial 5 and trial 6
RAVLT_perc_forgetting	Percentage of words forgotten between trial 5 and trial 6
FAQ	Functional Activities Questionnaire score

The data were grouped by patient ID and then ordered by visit code (baseline to 120 months) to create a time series. Missing values for each feature were imputed with the median value of that feature from all patients in the training and test sets.

Model Structure and Feature Engineering

A deep learning model was constructed with the Keras Python library on the TensorFlow backend (Figure 1). As in [14], a bidirectional long short-term memory (BiLSTM) layer is coupled to two hidden Dense layers which split into three branches, one for each task.
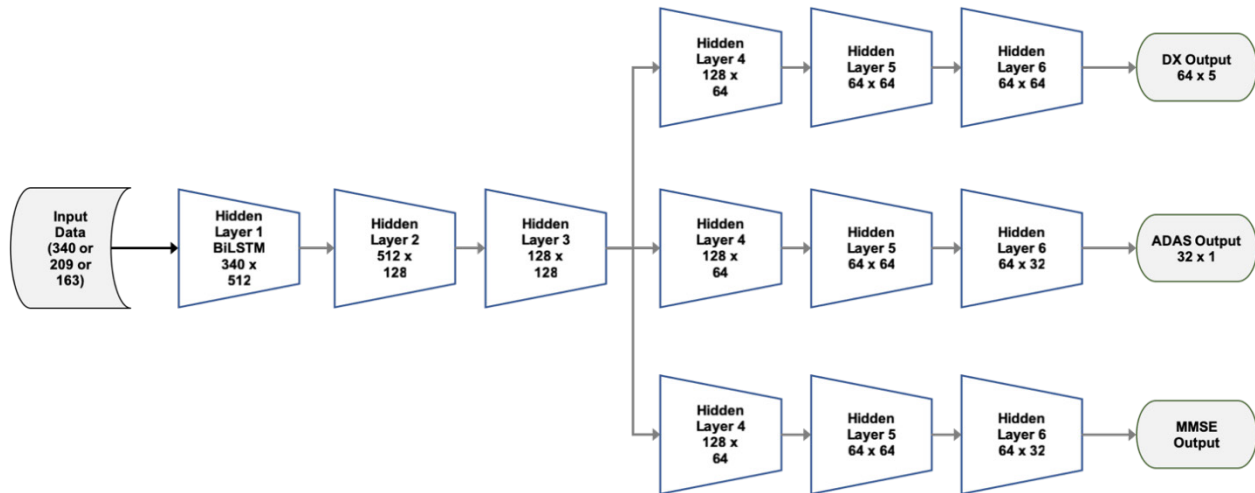

Figure 1. Model structure.
Plot of model nodes, layers, inputs, and outputs. This figure shows the 6-hidden layer configuration, with 3 common layers and 3 output-specific hidden layers.

Several combinations of deep learning hyperparameters were evaluated to determine the highest-performing model, including the number of features based on PCA, the number of common hidden layers at the head of the model, the number of hidden layers for each of the output branches, and the number of nodes in each layer.

Model Evaluation

The model was trained 5 times each with 163, 209, or 340 features (for 85%, 90%, and 100% explained variance for the PCA features) and a total of 4, 5, or 6 hidden layers. The metrics used were accuracy for the diagnosis multiclass classification task and root-mean-square error (RMSE) for the ADAS and MMSE regression tasks.

Results and Discussion

Figure 2a shows that, with 163 features, the model achieves an accuracy of 87.1% for the diagnosis task, with accuracy increasing with the addition of more hidden layers. Figure 2b shows that the model achieves a root-mean-square error of 9.95 and 8.6 for ADAS and MMSE, respectively, with the errors decreasing with the addition of more hidden layers.
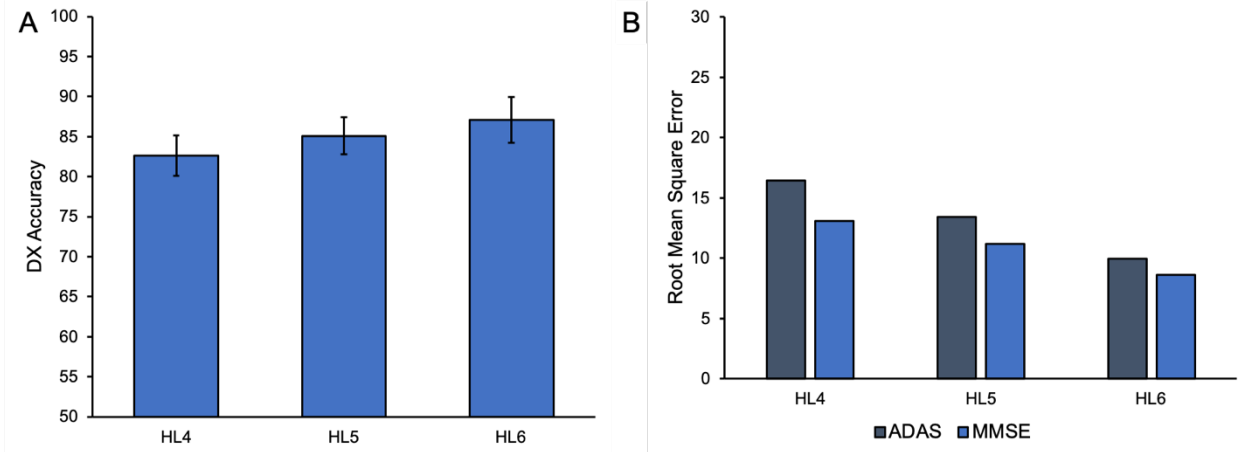
Figure 2. (A) Diagnosis classification accuracy. (B) ADAS and MMSE root-mean-square error.
163 features. Mean of 5 runs. Abbreviations: HL4, 4 hidden layers; HL5, 5 hidden layers; HL6, 6 hidden layers.

Figure 3a shows that, with 209 features, the model achieves an accuracy of 89.8% for the diagnosis task, with accuracy increasing with the addition of more hidden layers. Figure 3b shows that the model achieves a root-mean-square error of 5.14 and 3.88 for ADAS and MMSE, respectively, with the errors decreasing with the addition of more hidden layers.
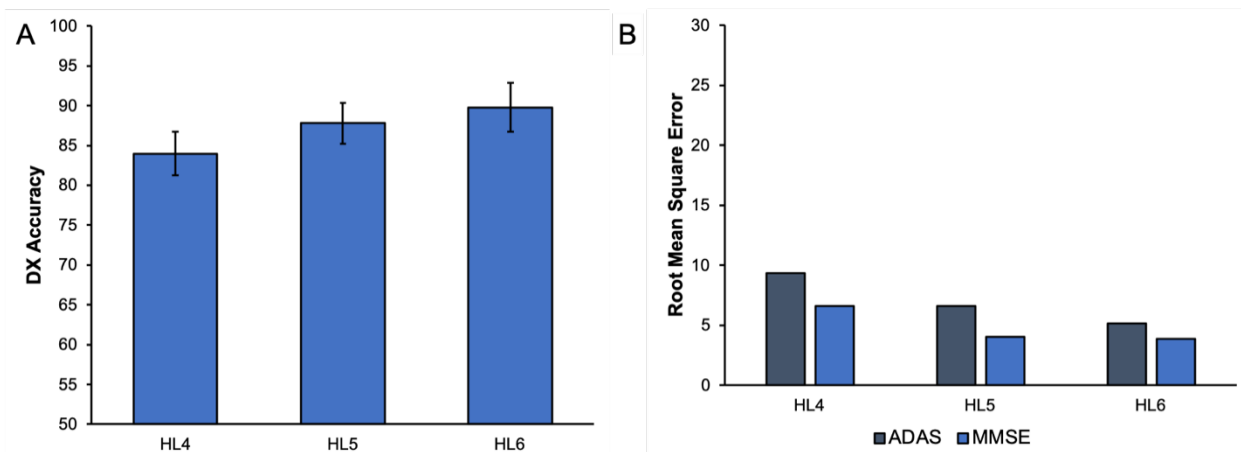

Figure 3. (A) Diagnosis classification accuracy. (B) ADAS and MMSE root-mean-square error.
209 features. Mean of 5 runs. Abbreviations: HL4, 4 hidden layers; HL5, 5 hidden layers; HL6, 6 hidden layers.

Figure 4a shows that, with 340 features, the model achieves an accuracy of 90.6% for the diagnosis task, with accuracy increasing with the addition of more hidden layers. Figure 4b shows that the model achieves a root-mean-square error of 3.59 and 3.82 for ADAS and MMSE, respectively, with the errors decreasing with the addition of more hidden layers.
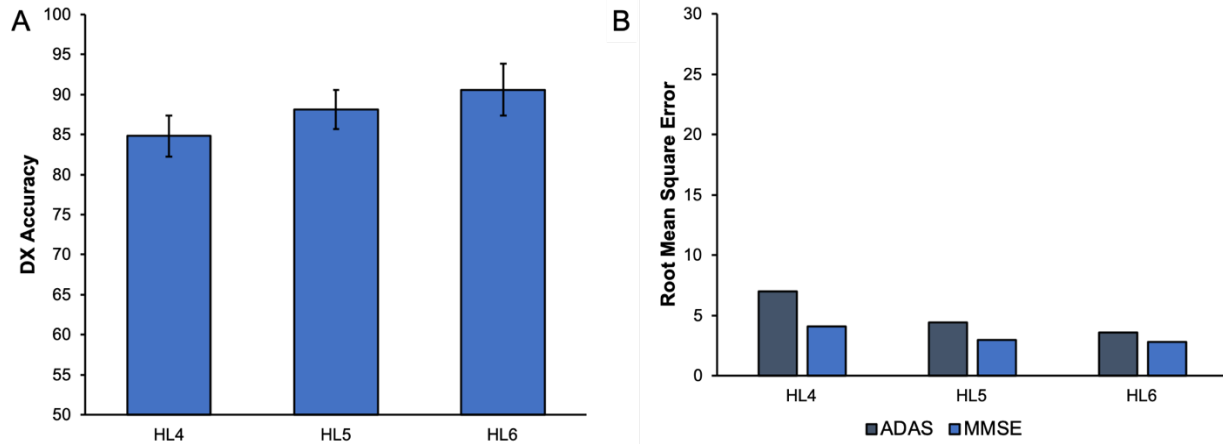
Figure 4. (A) Diagnosis classification accuracy. (B) ADAS and MMSE root-mean-square error.
340 features. Mean of 5 runs. Abbreviations: HL4, 4 hidden layers; HL5, 5 hidden layers; HL6, 6 hidden layers.

Table 3 summarizes the results obtained.

Table 3. Summary of results.

Features	DX Accuracy			ADAS RMSE			MMSE RMSE		
	HL4	HL5	HL6	HL4	HL5	HL6	HL4	HL5	HL6
[11]	—			0.854(MAE)			0.824(MAE)		
[12]	90.56%			—			—		
[13]	90.9 ± 2.7%			—			—		
[14]	91.17 ± 2.18%			0.076(MAE)			0.085(MAE)		
[16]	51.8%			8.537			2.373		
163	82.60%	85.10%	87.10%	2.19	2.16	2.20	1.28	1.10	1.22
209	84.00%	87.80%	89.80%	2.20	2.15	2.11	1.27	1.11	1.24
340	84.82%	88.16%	**90.60%**	1.52	1.66	**1.41**	0.802	0.570	**0.803**

The model with 340 features and 6 hidden layers was shown to have the greatest accuracy on the diagnosis task and lowest root-mean-square error on the ADAS and MMSE tasks.

Conclusion

This work has demonstrated that it is possible for a multitask multimodal deep learning model to predict diagnosis, ADAS, and MMSE with greater accuracy than prior state-of-the-art models [11, 12, 13, 14]. The model with 340 features and 6 hidden layers was shown to have the greatest accuracy on the diagnosis task and lowest root-mean-square error on the ADAS and MMSE tasks. By using clinically available features [7], this work improves upon existing research.

Future investigations could include using a convolutional neural network (CNN) or ResNet to process data from clinical images directly or training and validating the model with other clinical datasets to further improve its accuracy and evaluate its transferability.

In the clinical sector, this work could be applied to more accurately diagnose early-stage AD before symptoms appear or supplement doctor diagnoses by leveraging insights from large libraries of previous patient data.

Acknowledgements

References

NIH, "Alzheimer's Disease Fact Sheet," *National Institute on Aging*, 22-May-2019. [Online]. Available: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.

Society for Neuroscience, *Brain Facts: A Primer on the Brain and Nervous System*, 2018th ed. Society for Neuroscience, 2018, p. 70.

S. E. O'Bryant *et al.*, "Detecting Dementia With the Mini-Mental State Examination in Highly Educated Individuals," *Archives of Neurology*, vol. 65, no. 7, Jul. 2008.

J. K. Kueper, M. Speechley, and M. Montero-Odasso, "The Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review," *Journal of Alzheimer's Disease*, vol. 63, no. 2, pp. 423–444, Apr. 2018.

R. J. Caselli *et al.*, "Neuropsychological decline up to 20 years before incident mild cognitive impairment," *Alzheimer's & Dementia*, vol. 16, no. 3, pp. 512–523, Jan. 2020.

J. Albright, "Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm," *Alzheimer's & Dementia (New York, N. Y.)*, vol. 5, pp. 483–491, 2019.

R. V. Marinescu *et al.*, "TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data," *arXiv:2001.09016 [eess, q-bio, stat]*, vol. 11843, pp. 1–10, 2019.

K. Hao, "What is machine learning?," *MIT Technology Review*, 17-Nov-2018. [Online]. Available: https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/.

K. Ito *et al.*, "Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database," *Alzheimer's & Dementia*, vol. 7, no. 2, pp. 151–160, Sep. 2010.
Z. Yang and Z. Liu, "The risk prediction of Alzheimer's disease based on the deep learning model of brain 18F-FDG positron emission tomography," *Saudi Journal of Biological Sciences*, vol. 27, no. 2, pp. 659–665, Feb. 2020.

J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via multi-task learning," *NeuroImage*, vol. 78, pp. 233–248, Sep. 2013.

F. Liu, L. Zhou, C. Shen, and J. Yin, "Multiple Kernel Learning in the Primal for Multimodal Alzheimer's Disease Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 984–990, May 2014.

S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, "Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, no. 1, pp. 737–749, Jan. 2018.

S. El-Sappagh, T. Abuhmed, S. M. Riazul Islam, and K. S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, Oct. 2020.

European Progression of Neurological Disease Initiative and Alzheimer's Disease Neuroimaging Initiative, "TADPOLE - Grand Challenge," *grand-challenge.org*. [Online]. Available: https://tadpole.grand-challenge.org/Data/.

M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1195–1206, May 2019.