RESEARCH

Early detection of Alzheimer's disease using neuropsychological tests: a predict–diagnose approach using neural networks

Devarshi Mukherji[1], Manibrata Mukherji[2], Nivedita Mukherji[3*] and Alzheimer's Disease Neuroimaging Initiative

Abstract

Alzheimer's disease (AD) is a slowly progressing disease for which there is no known therapeutic cure at present. Ongoing research around the world is actively engaged in the quest for identifying markers that can help predict the future cognitive state of individuals so that measures can be taken to prevent the onset or arrest the progression of the disease. Researchers are interested in both biological and neuropsychological markers that can serve as good predictors of the future cognitive state of individuals. The goal of this study is to identify non-invasive, inexpensive markers and develop neural network models that learn the relationship between those markers and the future cognitive state. To that end, we use the renowned Alzheimer's Disease Neuroimaging Initiative (ADNI) data for a handful of neuropsychological tests to train Recurrent Neural Network (RNN) models to predict future neuropsychological test results and Multi-Level Perceptron (MLP) models to diagnose the future cognitive states of trial participants based on those predicted results. The results demonstrate that the predicted cognitive states match the actual cognitive states of ADNI test subjects with a high level of accuracy. Therefore, this novel two-step technique can serve as an effective tool for the prediction of Alzheimer's disease progression. The reliance of the results on inexpensive, non-invasive tests implies that this technique can be used in countries around the world including those with limited financial resources.

Keywords: Alzheimer's disease, Recurrent neural networks, Multi-layer perceptron neural networks, Sequence prediction

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that affects around 50 million people globally [1]. Projections show that 1 in 85 people will develop AD by 2050. In spite of the scale of the problem, there is no cure for the disease and countless clinical trials have been unsuccessful in finding effective treatment [2, 3]. In 2018, the Alzheimer's Association estimated that the average cost of caring for an Alzheimer's disease patient is $350,174—making it the most expensive disease in the United States.

Research shows that AD is an incredibly slowly progressing disease [4–6] and takes many years from initial cognitive decline to full-blown disease development. Due to this slow progression, early detection of AD can be crucial in both its treatment and prevention. Currently, the detection of AD relies on invasive and expensive tests—spinal taps for CSF Tau protein, brain scans, and blood biomarker detections [7–9]. These tests contribute to the overall costs associated with AD assessment and treatment mentioned above. It is therefore critical to develop an effective, less expensive, and non-intrusive screening method to identify people who are at risk of developing AD so that the expensive and intrusive tests can be used only for those patients. Moreover, multiple ongoing global efforts aim to identify the

*Correspondence: mukherji@oakland.edu

[3] Oakland University, Rochester, MI, USA
Full list of author information is available at the end of the article

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 2 of 26

optimal candidates for their clinical trials (cohorts) to test new therapies and understand disease progression. The cohorts selected for these trials often include patients who do not develop cognitive impairment during their participation and result in an unfortunate waste of resources [8, 10].

It has been shown that neuropsychological tests are effective in the diagnosis of AD and in the identification of patients that are likely to experience AD progression [11–17]. These tests are much cheaper to administer than CSF spinal taps and brain scans and can be used to identify patients at risk of developing cognitive decline. The goal of this study is to use multiple Long Short-Term Memory Recurrent Neural Networks (LSTM RNNs, for their ability to handle the vanishing gradients problem and capture long-term dependencies efficiently) on a set of well-established neuropsychological tests such as the ones used by [18] to assess the current cognitive state of a subject and use current and past test score data to predict if a subject is likely to develop cognitive impairment within the next 2 to 4 years. This segmentation of subjects into those who are expected to remain cognitively normal and those likely to develop problems allows doctors and researchers to select appropriate subjects for clinical trials. Access to inexpensive assessments of cognitive state can be of great benefit to people around the world, especially those in underdeveloped countries where low-cost diagnosis methods can be used for screening and more expensive tests can be preserved for only the most high-risk patients.

All data for this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. ADNI was launched in 2003 in cooperation with the National Institutes of Health (NIH) and the National Institute on Aging to better understand AD development and discover new therapeutic techniques. This database includes data on neuropsychological tests, along with many other metrics used to monitor AD progression. Based on the research on neuropsychological tests in AD detection and prediction, five different neuropsychological tests (MMSE, ADAS Q4, ADAS Cog11, ADAS Cog13, and FAQ) were chosen for this study—[12] offers a brief description of these tests. The ADNI database includes data on all of these tests. Another study [19], assessed different neuropsychological tests to determine which have good prediction power.

Neural networks have been used in the medical field for several years, and recently data from the ADNI database have been used by many researchers to predict neuropsychological test scores or cognitive status primarily based on brain scans and volumes of different parts of the brain that have been identified as correlated to AD progression [20–30]. Even though our work differs from this body of research in its goal to utilize inexpensive markers like neuropsychological test scores, it does strengthen our hypothesis that ADNI's data on neuropsychological tests capture the subtle influences of important biomarkers like brain volume and can be used on their own to predict cognitive status.

An interesting research work that reinforces the relative contribution of the ADNI neuropsychological tests on the cognitive status of the trial participants is [31]. The authors use the Fuzzy-Rough Feature Selection technique to determine the most effective set of features that makes the largest contribution towards the cognitive status outcomes arrived at by ADNI. They use two search strategies and three similarity functions to determine the set of features that contribute the most in determining the ADNI outcome. They use a forward search in which features are inserted one at a time and a backward search in which features are removed iteratively. The best forward search feature set includes seven neuropsychological tests of which three are used in our work. The best backward search feature set contains no neuropsychological tests. Moreover, they use eight neural network models (no RNNs) to find the accuracy of these features in modeling the ADNI outcomes. The best forward-pass feature set yields the highest average accuracy of around 85% (Table 4 in the paper) and the best backward-pass feature set yields the lowest average accuracy of around 39% (Table 5 in the paper). The 85 percent accuracy number indicates that the neuropsychological tests contribute disproportionately to the accuracy of the ADNI cognitive status determinations. The 39 percent accuracy number indicates that the absence of neuropsychological tests has a disproportionately adverse effect on the determination of the ADNI cognitive status. Both outcomes corroborate the suitability of using neuropsychological tests on their own in any AD modeling work and reinforce the motivation behind our work.

Other research that uses past test data to predict future patient outcomes includes [32] where a five-step procedure is developed to record the impact of an auditory task on the EEG recordings of schizophrenia patients. The paper attempts to automate the decision-making process of whether to continue therapy for these patients in the future or to stop it. Two databases are used to store therapy outputs over multiple time steps and correlations between decisions made for those time steps. In contrast, our work captures the relationship between the input markers and the normal/abnormal status in the trained neural network models and we do not need any explicit equation to model the relationship between inputs and the predicted future state. Also, there is no storage of any data in a database for future lookup.

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 3 of 26

As mentioned before, this study uses LSTM RNNs (LSTMs, henceforth) for prediction and Multi-Layer Perceptrons (MLPs) for diagnosis. MLPs are a class of Artificial Neural Networks (ANNs) that use multiple layers of neurons and can be used to classify input data. LSTMs are machine learning models that use sequences of data to predict future values [33–36]. LSTMs have been used effectively in many domains ranging from language recognition to marketing to healthcare. In the context of AD, [37] uses sequence prediction models in which irregularity in time series data are handled by a neural network layer that models a continuous-time autoregressive (CAR) model followed by either an RNN, LSTM, or GRU layer to predict future test values for several markers from the ADNI database. Since our data have no time irregularities, we used just an LSTM to predict the sequence values, and, unlike [37], used the predicted values to perform diagnosis using MLPs. [18] is one of the recent studies that used LSTM in AD prediction. Using National Alzheimer's Coordinating Center (NACC) data, [18] used 78 features and a global CDR (Clinical Dementia Rating) score to identify patients that are likely to experience AD progression based on changes in CDR scores. Unlike ADNI, the NACC data are not collected at equal intervals for all subjects necessitating [18] to introduce time between visits as a separate feature. The paper utilized a large number of features (78) in a multi-feature model to identify patients that are likely to experience worsening of the disease in the future. In contrast, the LSTMs used in the current study are developed individually for each of the chosen tests, and their output is analyzed using MLPs to diagnose a patient as likely to remain cognitively normal or not in the next 2 to 4 years. Consequently, this paper can be viewed as a less expensive step that should be conducted before a full-blown LSTM that uses a comprehensive and expensive set of features is used to determine how rapidly a patient's cognitive decline is expected to progress in the future. The objectives of this study are:

1. To develop LSTMs, one for each of the chosen neuropsychological tests in the ADNI dataset, that can predict future test scores based on past values.
2. To develop MLPs that use all predicted test scores from step 1 above to diagnose whether the patient has cognitive impairment.
3. To combine the above two sets of neural network models to predict test scores for each test over the next 2–4 years and diagnose upon those predictions.

All of these objectives were met in this study. In the process of meeting the objectives, we have demonstrated that neuropsychological tests can be used on their own to predict the cognitive state of individuals. The goal of this work was not to supersede the cognitive assessments made by ADNI. Instead, the goal was to determine whether the cognitive assessments made by ADNI could be used as outputs to train MLPs which relied only on a subset of the set of features that ADNI gathers for each participant (namely, neuropsychological tests). Data for these features can be easily gathered and the two-step neural network algorithm can be used to forecast cognitive status for a 2- to 4-year window. In addition, our innovative approach of using predicted test results for diagnosing can be utilized for any set of markers from other established trials going on around the world. There are other studies that have predicted multiple feature values using neural networks, but we think that this work is the first one that used such predicted values as collective inputs to MLPs to diagnose cognitive status.

2 Methods
2.1 Data
2.1.1 Data, software, and packages
All data used in this project were obtained from the ADNI database. ADNI is a longitudinal study launched in 2003 in cooperation with the National Institutes of Health (NIH) and the National Institute on Aging. ADNI uses adult volunteers between the ages of 55 and 90. The initial cohort included 200 cognitively normal (CN) subjects, 400 subjects with mild cognitive impairment (MCI), and 200 subjects diagnosed with Alzheimer's disease (AD). Each individual is assigned a unique identification number or RID. CN subjects have a Clinical Dementia Rating (CDR) of 0 and MMSE score between 24 and 30; MCI subjects show signs of memory loss, have a CDR score of 0.5, and are at least one standard deviation below the mean score on the delayed recall portion of the Wechsler Memory Scale's Logical Memory II. Those who are diagnosed with AD are diagnosed in accordance with the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [38]. The ADNI study has completed three phases to date and is currently in its ADNI3 phase. The first phase, ADNI1, started in October 2004 and spanned 5 years. ADNI-GO was the second phase and lasted 2 years after its launch in September 2009, and ADNI2 spanned 5 years since its launch in September 2011. In each phase, new participants were recruited, and participants from the previous phases continued. The ADNI dataset contains demographic data of the subjects as well as results of various neuropsychological tests, results of brain scans, and other physiological tests conducted at various intervals. The relevant data for this study are obtained from a subset of the ADNI dataset called

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 4 of 26

Table 1 Descriptive statistics of the 66 test dataset subjects

Age	#	Gender	#	Education	#	Race	#	Marital status	#
58–68	8	Female	29	8–11	3	Asian	2	Divorced	4
68–78	41	Male	37	12–15	22	Black	4	Married	49
78–88	17			16–20	41	White	60	Never married	1
								Widowed	12

Table 2 Test score cutoffs and normalizations

Test	Normal cutoff	Min	Max	Normalization
MMSE	≥ 28	0	30	Value/30
ADAS4	≤ 5	0	10	Value/10
ADAS11	≤ 10	0	70	Value/70
ADAS13	≤ 13	0	85	Value/85
FAQ	≤ 2	0	30	Value/30

ADNIMERGE which was downloaded on May 14, 2019. Data extraction and processing were conducted using the R programming language running in RStudio version 1.2. The recurrent neural network analysis was conducted using the Keras package version 2.2.5.0 running in RStudio. Descriptive statistical analysis was conducted using the statistical software package Stata 14.

Since the objectives of this paper are based on LSTMs that use past neuropsychological test scores to predict future test scores, it is important to select subjects who have sufficient time series data for the chosen tests. ADNI subjects undergo an initial assessment that determines their baseline cognitive states and then have follow-up tests at months 6, 12, 18, 24, 36, 48, 60, 72, 84, and so on. The data for this study are primarily derived from the baseline, months 6, 12, 24, 36, 48, 60, and 72. (Months missing in this interval, such as 18, do not have any neuropsychological test values in the ADNI database.) A sequence of length four uses the baseline, and months 6, 12, and 24 data; a sequence of length six also includes months 36 and 48, and a sequence of length eight includes months 60 and 72.

2.1.2 Neuropsychological test selections

Five neuropsychological test scores were used to assess the cognitive state of the subjects. The tests are MMSE, ADAS Q4, ADAS Cog-11, ADAS Cog-13, and FAQ. These tests have been used to demonstrate their effectiveness in the diagnosis and prediction of cognitive impairment [12]. Additionally, neuropsychological tests were used in a regression model [39] to determine the impact on change in CDR sum of scores, and "MMSE, FAQ, and ADAS-cog were identified as prognostic factors to detect cognitive decline in CDR-SB". Other test scores, such as the Wechsler Logical Memory Delay (LDELTOTAL), are also available. However, the number of missing values of the latter test was greater than for the other tests, and the test was given less frequently (every 2 years as opposed to half-yearly or yearly). As a result, LDELTOTAL could not be used as a marker in this study.

2.2 Training, validation and test data set creation

As noted earlier, the battery of neuropsychological tests is administered during scheduled visits that occur at the initial visit (baseline) and then at months 6, 12, 24, and beyond. All participants do not attend all of the scheduled visits, and even when they do, results for all scheduled tests are not available. Consequently, there are many missing values for all of the tests that are part of the ADNI study. In order to predict the cognitive status 2–4 years ahead, a minimum of four past test scores is needed for each test subject. As a result, sequences of lengths four, six, and eight were extracted for each of the five neuropsychological tests separately from the ADNI1,

Table 3 Number of observations for training and validation of sequence prediction RNNs
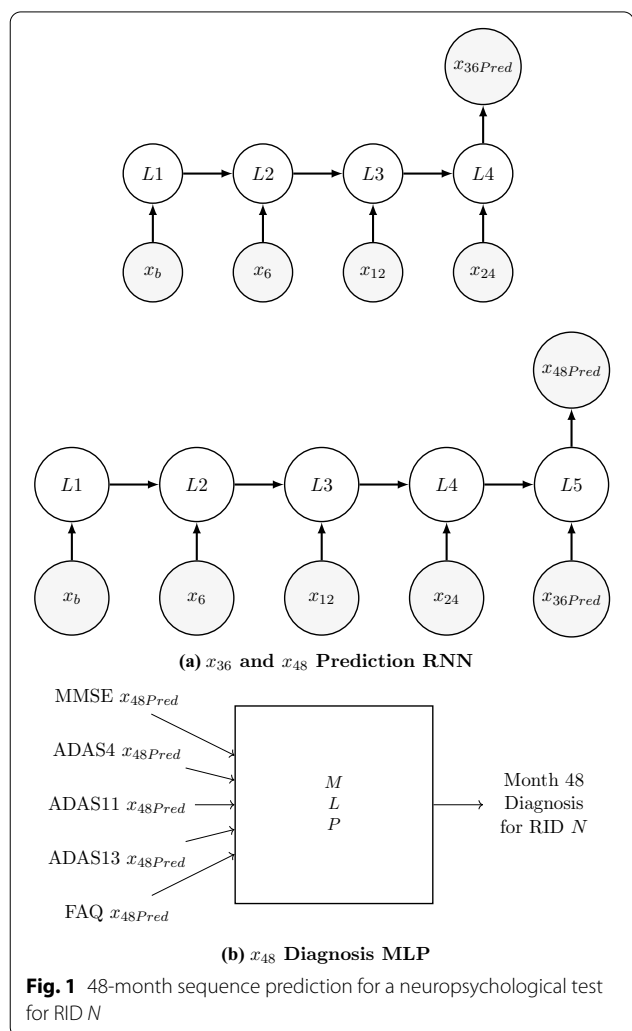
Test	Training 36th and 48th months	Training 60th and 72nd months	Validation 36th and 48th months	Validation 60th and 72nd months
MMSE	458	230	115	57
ADAS4	458	220	115	55
ADAS11	441	202	110	50
ADAS13	437	209	109	52
FAQ	466	250	116	63

Mukherji *et al. Brain Informatics*　　(2022) 9:23

Page 5 of 26

Table 4 Number of observations for training and validation of diagnosis MLPs

Diagnosis time	Training	Validation
Month 24	632	156
Month 48	428	156
Month 72	264	104

ADNI-GO, and ADNI2 data sets and combined to form the training and validation data for the test-score prediction LSTM models. No special care was taken to collect sequences for the same patient across the neuropsychological tests since the generic pattern of disease progression is what the LSTMs need to learn, not the progression pattern of the same patient across the tests.

Special care had to be taken to re-create or impute missing values. To use mostly real data values and reduce

(a) x_{36} and x_{48} **Prediction RNN**

(b) x_{48} **Diagnosis MLP**

MMSE x_{48Pred}
ADAS4 x_{48Pred}
ADAS11 x_{48Pred}
ADAS13 x_{48Pred}
FAQ x_{48Pred}

M
L
P

Month 48 Diagnosis for RID N

Fig. 1 48-month sequence prediction for a neuropsychological test for RID N

(a) x_{60} and x_{72} **Prediction RNN**

(b) x_{72} **Diagnosis MLP**

MMSE x_{72Pred}
ADAS4 x_{72Pred}
ADAS11 x_{72Pred}
ADAS13 x_{72Pred}
FAQ x_{72Pred}

M
L
P

Month 72 Diagnosis for RID N
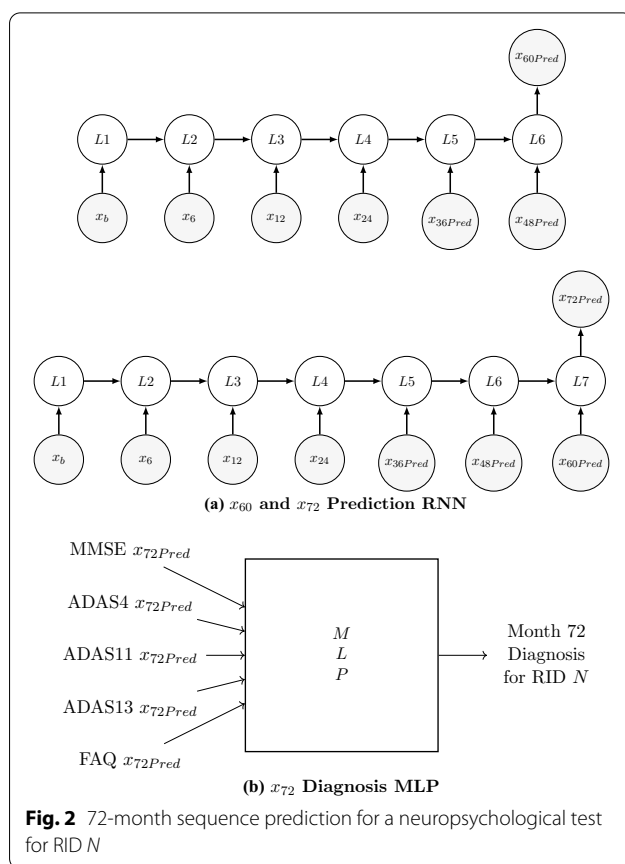
Fig. 2 72-month sequence prediction for a neuropsychological test for RID N

the number of imputations, it was ensured that subjects had no more than two missing test scores. Moreover, subjects who had a missing value on the last visit of a sequence of lengths four, six, or eight were also dropped. This constraint ensured that no model would learn the generic trend of disease progression using imputed values as the most recent data point. A simple average method was used to impute missing values. As an example, for a model of length four, a subject with missing data for month 12 was assigned the average of the baseline, month 6, and month 24 values. It is one of the many ways in which missing data can be imputed. This imputation method is advantageous since this method preserves the average of the sequence.

Imputed sequences of length six from the combined data set were used to generate the training and validation data for the LSTM models used to predict the fifth and sixth values. For example, for the LSTM model for predicting the fifth value, the model was trained with sequence values one through four as input data and sequence value five as the output data value. The model learned how the fifth value was associated with the previous four values using the training data and tested what it learned on the validation data to determine its accuracy.

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 6 of 26

Similarly, for the sixth value prediction model, sequence values one through five were used as the input data, and value six was used as the output data value. Eighty percent of the rows were used to create the training set and the remaining 20% of rows made up the validation set. For the seventh and eighth element prediction models, a similar set of steps were followed using the combined data set containing sequences of length eight.

(a) Diagnosis MLP for Month 48

```
Model: "sequential_143"
_____
Layer (type)                    Output Shape                       Param #
=============================================================================
dense_456 (Dense)               multiple                           30
_____
dense_455 (Dense)               multiple                           60
_____
dense_454 (Dense)               multiple                           220
_____
dense_453 (Dense)               multiple                           210
_____
dense_452 (Dense)               multiple                           55
_____
dense_451 (Dense)               multiple                           6
=============================================================================
Total params: 581
Trainable params: 581
Non-trainable params: 0
_____
```

(b) Diagnosis MLP for Month 72

```
Model: "sequential_410"
_____
Layer (type)                    Output Shape                       Param #
=============================================================================
dense_2177 (Dense)              multiple                           30
_____
dense_2176 (Dense)              multiple                           300
_____
dense_2175 (Dense)              multiple                           1275
_____
dense_2174 (Dense)              multiple                           130
_____
dense_2173 (Dense)              multiple                           30
_____
dense_2172 (Dense)              multiple                           6
=============================================================================
Total params: 1,771
Trainable params: 1,771
Non-trainable params: 0
_____
```

(c) Sequence Prediction RNN

```
Model: "sequential_17"
_____
Layer (type)                    Output Shape                       Param #
=============================================================================
lstm_34 (LSTM)                  (None, 5, 100)                     40800
_____
lstm_35 (LSTM)                  (None, 25)                         12600
_____
dense_17 (Dense)                (None, 1)                          26
=============================================================================
Total params: 53,426
Trainable params: 53,426
Non-trainable params: 0
_____
```

Fig. 3 Summary of diagnosis MLPs and sequence prediction RNN models

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 7 of 26

The MLPs need to learn to diagnose the cognitive state of a single patient given the predicted test score values of the five neuropsychological tests of that patient. Sequences of lengths four, six, and eight were extracted for a given RID for each of the five neuropsychological tests separately from the ADNI1, ADNI-GO, and ADNI2 datasets, yielding 15 different collections. The ADNI diagnosis values (CN, MCI, or AD) were also saved for each RID and for each sequence at the fourth, sixth, and eighth visits from the three data sets (ADNI1, ADNI-GO, and ADNI2). If the test score of any of the five tests at visits four, six, or eight and the corresponding ADNI diagnosis was not available, the RID was removed from the data set. Among the RIDs that remained, if any sequence for the five tests contained more than two unavailable scores, the RID was removed from the data set. The remaining test scores formed the core for generating the training and validation data for the diagnosis MLP models. A classification output value of 0, cognitively normal (CN), was generated for an ADNI diagnosis value of CN and a classification output value of 1, non-CN, was generated for ADNI diagnosis values of MCI and AD. For each sequence length, the data were then sorted according to the classification value (0 or 1), and 80% of the smaller of the two classified sets was combined with a matching number of sequences from the other set of sequences to create the training set. This procedure was carefully incorporated due to the models' enhanced ability to learn if there are an equal number of 0s and 1s. Finally, the remaining sequences were pulled together to form the validation data set.

To determine if a combination of neuropsychological tests can be used to predict the future cognitive state of an individual, it is necessary to create a test data set of individuals for whom longitudinal data of length eight (baseline to 72 months) are available for all of the five tests chosen for this study. After applying the same criterion of no more than two missing values, no missing diagnosis value, and ensuring that these sequences were not used as training or validation data for any LSTM or MLP model, only 66 patients who had values for all the five neurological tests could be obtained from the oldest cohort—ADNI1. Descriptive statistics for this cohort are given in Table 1.

Table 2 provides the range of test score values for each test, along with their cutoff scores for normal values which were obtained from the different references discussed in Sects. 1 and 2. The last column of the table provides the normalization calculations used for each test to convert the values to a scale of 0 to 1. Even though the cutoff values were not used to determine the classification output values of the MLP models, these cutoff values were used to determine the efficacy of the sequence

prediction LSTM models and to determine the contribution of the normal and abnormal patients towards the overall diagnosis as discussed in a later section.
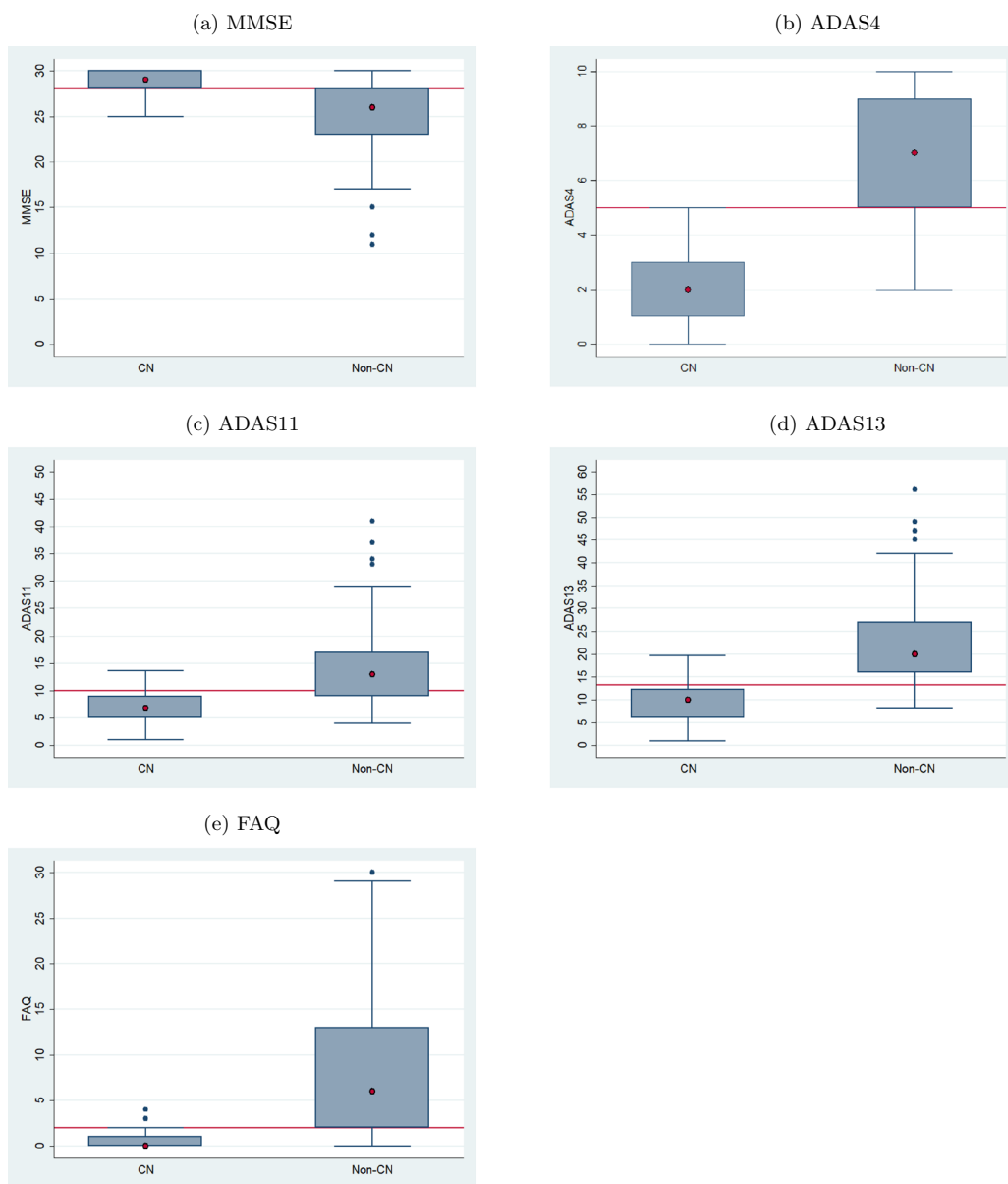
2.2.1 Model creation and application

Each diagnosis and sequence prediction model was trained using its training data and validated using its validation data. The loss curve was plotted for each model, and the model and the loss curve were saved. The models with the highest prediction accuracy were saved for later use in the test data prediction phase.

Table 3 gives the number of unique RIDs used to train and validate the sequence prediction models. Table 4 gives the number of unique RIDs used to train and validate the fourth step, sixth step, and eighth step diagnosis models. The numbers show that for each test, the number of observations for the data sets decreases when going from sequences of length six to sequences of length eight since the number of RIDs that have missing values over a 72-month period is much higher than it is over a 48-month period.

Figures 1 and 2 show the schematic of how the different neural network models were used. The main goal of this study lies in the combination of the predictive and diagnostic models. The combination of the models is used on the final test data created for each neuropsychological test. The final test data set is the set of RIDs that is used to test the accuracy and practicality of the study, and it consists of 66 patients who all have test scores for each of the neuropsychological tests. As shown in Fig. 1a, the first four values of a sequence for a specific test and for a specific RID, N, say, were fed into the best fifth-element sequence prediction model to predict the month 36 value for the test (x_{36Pred}). Then, the same four values along with x_{36Pred} were fed into the best sixth-element sequence prediction model to predict the month 48 value for the test (x_{48Pred}). This process was repeated for every test for RID N to obtain the x_{48Pred} values for all tests. Then, as shown in Fig. 1b, all of the x_{48Pred} values were fed to the best month-48 diagnosis model to diagnose the CN–non-CN status of RID N. Figure 2 shows the same process repeated for the generation of (x_{72Pred}) for all the tests and the corresponding diagnosis of the CN–non-CN status for RID N. Then these diagnosis predictions,

Table 5 Diagnosis models' accuracies

Diagnosis time	Accuracy (%)
Month 24	85.25
Month 48	78.85
Month 72	77.88

Fig. 4 Ranges of validation data split by the predicted diagnosis (CN or non-CN) assigned by the diagnosis model at month 48 (with normal cutoffs)

(f) Boxplot and Descriptive Statistics

CN or Non-CN	Test	Low Adj Value	p25	Median	p75	Up Adj Value	Mean	Min	Max
CN	MMSE	25	28	29	30	30	29	25	30
Non-CN	MMSE	17	23	26	28	30	25	11	30
CN	ADAS4	0	1	2	3	5	2	0	5
Non-CN	ADAS4	2	5	7	9	10	7	2	10
CN	ADAS11	1	5	7	9	14	7	1	14
Non-CN	ADAS11	4	9	13	17	29	15	4	41
CN	ADAS13	1	6	10	12	20	10	1	20
Non-CN	ADAS13	8	16	20	27	42	23	8	56
CN	FAQ	0	0	0	1	2	1	0	4
Non-CN	FAQ	0	2	6	13	29	9	0	30

for months 48 and 72, were compared to the real diagnosis given by ADNI for RID N (using DX values) at months 48 and 72, respectively, to determine the accuracy of the combined model.

Figure 3 summarizes the parametric structure of the three types of neural network models that were used in this study. Figure 3a shows the structure of the diagnosis MLP that was used to diagnose the five neurological test scores at month 48. The neural network has an input layer, a "sigmoid" output later, and four hidden layers. Figure 3b shows the structure of the diagnosis MLP that was used to diagnose the five neurological test scores at month 72. It also has an input layer, a "sigmoid" output later, and four hidden layers but uses a different number of neurons in each hidden layer. The 'relu' activation was used for every non-output layer and the "mean-squared-error" loss function and the 'adam' optimizer were used to compile the models. Figure 3c shows the common structure of the sequence prediction LSTMs that were used to predict the five neurological test scores. Two layers of LSTMs were used followed by a dense output layer. The 'relu' activation was used for every layer and the "mean-squared-error" loss function and the 'adam' optimizer were used to compile the models.

3 Results

There are three sets of results for this work, and they relate to (1) accuracies of the diagnosis of cognitive states, (2) how closely the predicted values of the test scores follow the trend of the actual values in the validation data set 2 years (sixth value) and 4 years (eighth value) after month 24, and (3) accuracies of the diagnosis when the sequence prediction LSTMs and the diagnosis MLPs are combined to predict on the test data set.

3.1 Diagnosis models' accuracies

The diagnosis models learn from the training data to determine which combination of the five test scores should be assigned a CN (0) value and which combination should be assigned a non-CN (1) value. When the trained models are applied to the validation data, the models' assignments of 0 and 1 are compared with the labeling of 0 (ADNI diagnosis CN, DX = 1) and 1 (ADNI diagnosis MCI/AD, DX = 2/3) of the validation data sequences to arrive at the accuracy results.

Table 5 gives the accuracies of the diagnosis models. The diagnosis at month 48 is 78.85% accurate and the diagnosis at month 72 is 77.88% accurate. The accuracy

values were determined by taking the proportion of correct diagnoses to the total number of diagnoses.

The boxplots of Figs. 4 and 5 serve as a corroboration of the accuracies noted above. The plots also corroborate the accuracy of the normal–abnormal cutoff values shown in Table 2. The boxplots for each graph are split into two groups, CN and non-CN. CN depicts the range of scores diagnosed as 0 by the diagnosis models, and non-CN depicts those diagnosed as 1. The stark contrast in the ranges of the CN and non-CN boxplots shows that the diagnosis models were able to effectively learn for all of the five tests. The tables in Figs. 4f and 5f contain the descriptive statistics of the boxplots along with the minimum value, maximum value, and the mean of the boxplot ranges for each of the five tests.

An aspect of the boxplots of Fig. 4 that deserves special mention is that even though the prediction accuracies are in the high seventies, there are still some RIDs that are incorrectly classified. For example, for MMSE (Fig. 4a), the CN group contains some RIDs that are in the abnormal range (< 28) and the non-CN group contains some RIDs that are in the normal range (≥ 28). Figure 6a shows a bar graph for all the five tests in which the CN and non-CN groups are further split into normal and abnormal percentage values based on the cutoff values for the test from Table 2. For example, for MMSE, the CN group contains 88% RIDs that are in the normal range and 12% RIDs that are in the abnormal range. It is these 12% abnormal RIDs that run the risk of being misdiagnosed as normal and hence dropped from further monitoring and treatment. Similarly, for MMSE, the non-CN group contains 31% RIDs that are in the normal range and 69% RIDs that are in the abnormal range. Once again, it is the 31% normal RIDs that run the risk of being monitored and treated causing a waste of valuable resources. Such misdiagnosis is inevitable in the context of machine learning approaches, especially when training data are scarce. But one important observation that should be made in regard to the diagnosis reached by ADNI, as shown in Fig. 6b, is that the mix of normal−abnormal misdiagnosis percentages are not any better when we consider the validation data and group them by the ADNI DX value (CN and non-CN, that is, MCI/AD). As seen in Fig. 6b, for MMSE, say, the non-CN group contains 43% RIDs that are in the normal range (compared to 27% using the predicted diagnosis of our models) and run the risk of being monitored and treated causing a waste of valuable resources. All the tests in Fig. 6a, when

(See figure on next page.)
Fig. 5 Ranges of validation data split by the predicted diagnosis (CN or non-CN) assigned by the diagnosis model at month 72 (with normal cutoffs)
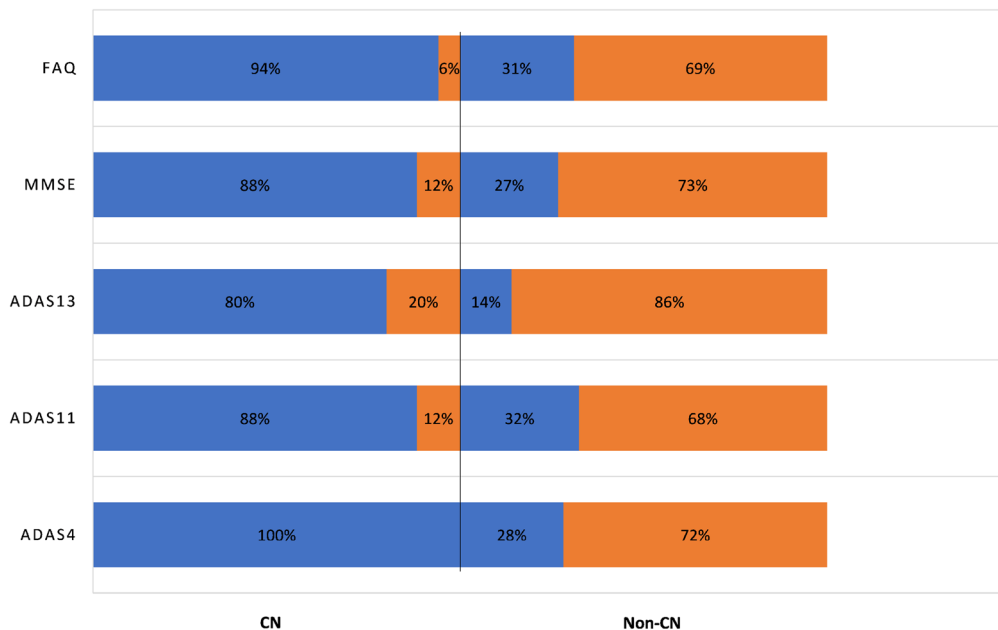
(a) MMSE

(b) ADAS4

(c) ADAS11

(d) ADAS13

(e) FAQ
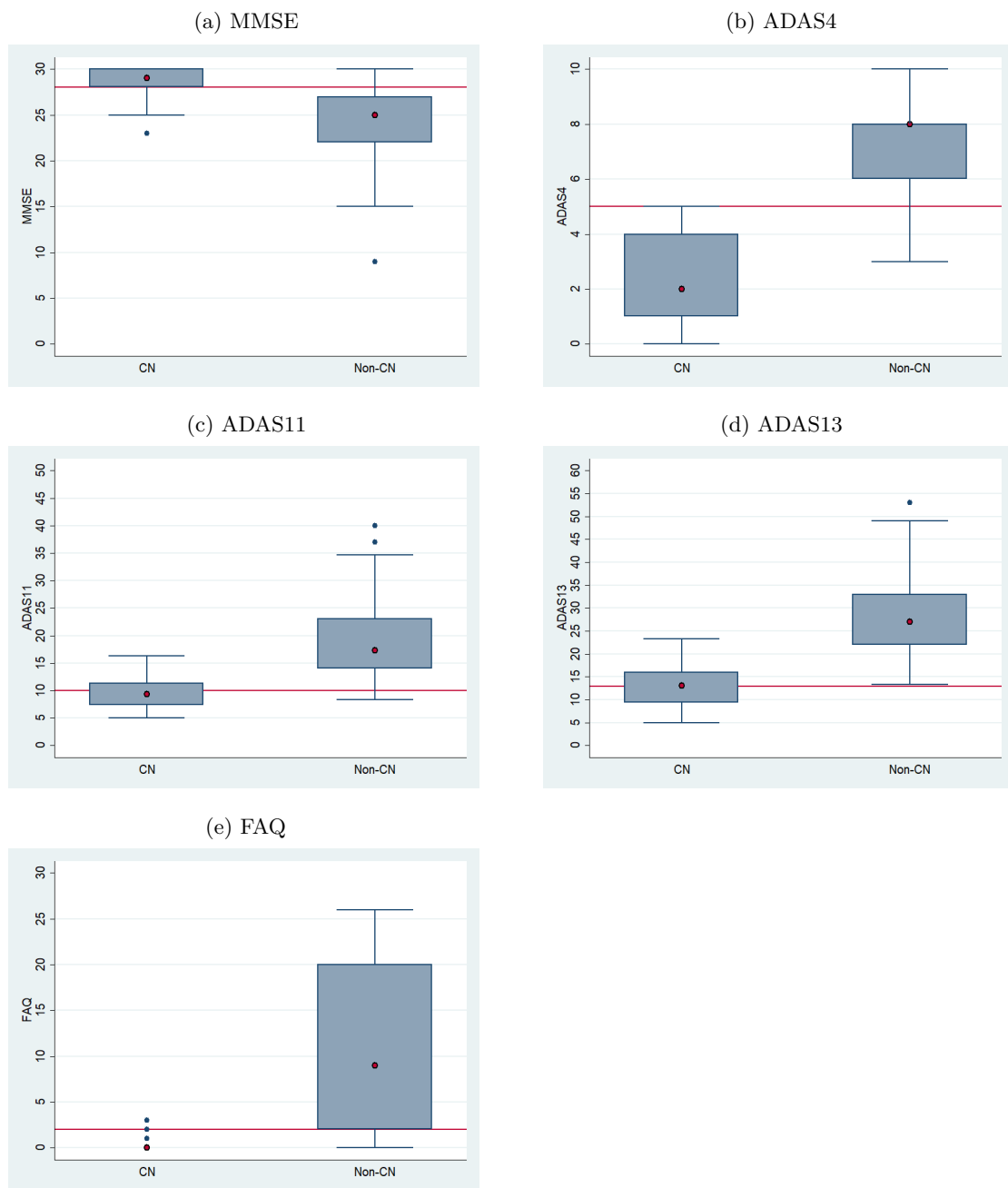
(f) Boxplot and Descriptive Statistics

CN or Non-CN	Test	Low Adj Value	p25	Median	p75	Up Adj Value	Mean	Min	Max
CN	MMSE	25	28	29	30	30	29	23	30
Non-CN	MMSE	15	22	25	27	30	24	9	30
CN	ADAS4	0	1	2	4	5	2	0	5
Non-CN	ADAS4	3	6	8	8	10	7	3	10
CN	ADAS11	5	7	9	11	16	10	5	16
Non-CN	ADAS11	8	14	17	23	35	19	8	40
CN	ADAS13	5	9	13	16	23	13	5	23
Non-CN	ADAS13	13	22	27	33	49	29	13	53
CN	FAQ	0	0	0	0	0	0	0	3
Non-CN	FAQ	0	2	9	20	26	11	0	26

Fig. 5 (See legend on previous page.)

MONTH 48 VALIDATION DATA AND PREDICTED DIAGNOSIS

	CN	Non-CN
FAQ	94% · 6%	31% · 69%
MMSE	88% · 12%	27% · 73%
ADAS13	80% · 20%	14% · 86%
ADAS11	88% · 12%	32% · 68%
ADAS4	100%	28% · 72%

(a)

MONTH 48 VALIDATION DATA AND ADNI DIAGNOSIS

	CN	Non-CN
FAQ	94% · 6%	50% · 50%
MMSE	94% · 6%	43% · 57%
ADAS13	72% · 28%	39% · 61%
ADAS11	81% · 19%	53% · 47%
ADAS4	98% · 2%	51% · 49%
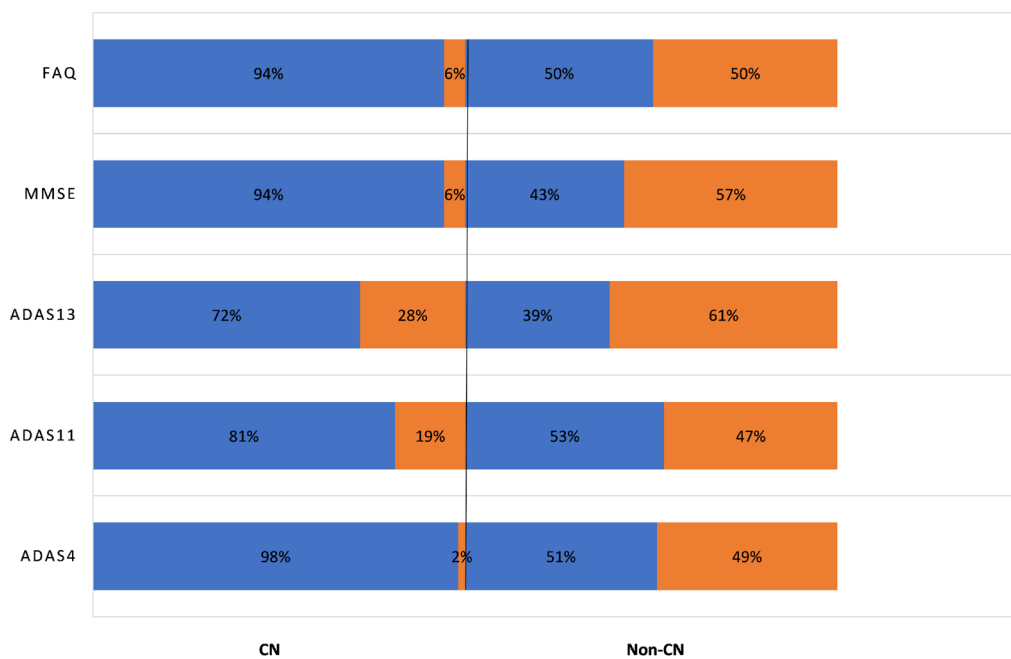
(b)

Fig. 6 Proportion of correct and incorrect diagnosis in CN and non-CN categories for validation data at month 48

72 MONTH VALIDATION DATA AND PREDICTED DIAGNOSIS

	CN	Non-CN
FAQ	98% · 2%	26% · 74%
MMSE	89% · 11%	21% · 79%
ADAS13	51% · 49%	100%
ADAS11	57% · 43%	5% · 95%
ADAS4	100%	15% · 85%
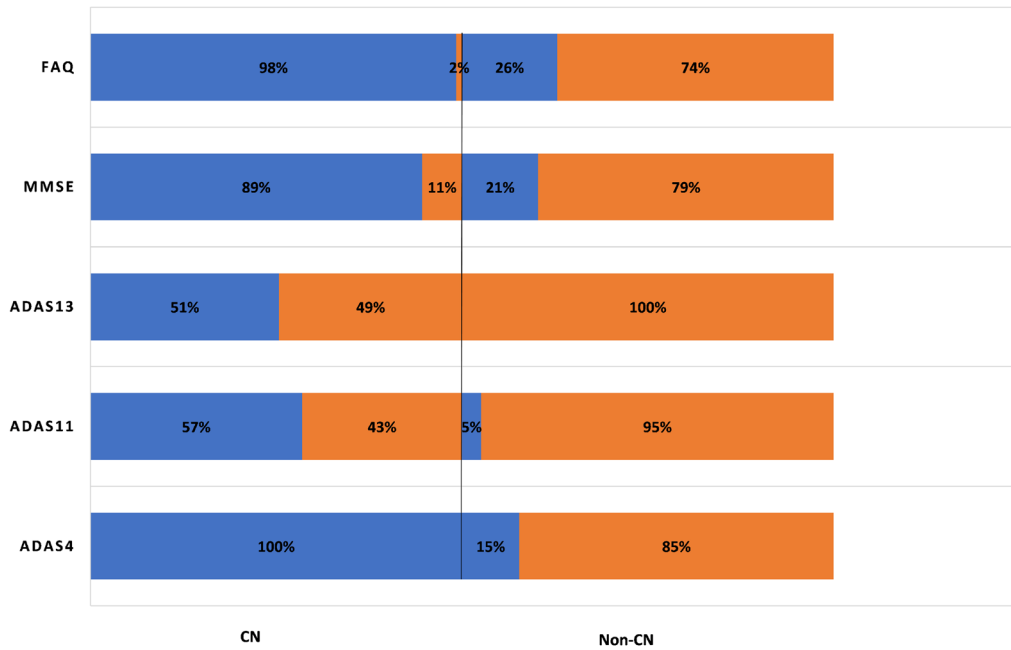
(a)

MONTH 72 VALIDATION DATA AND ADNI DIAGNOSIS

	CN	Non-CN
FAQ	100%	58% · 42%
MMSE	94% · 6%	49% · 51%
ADAS13	67% · 33%	15% · 85%
ADAS11	70% · 30%	23% · 77%
ADAS4	100%	54% · 46%
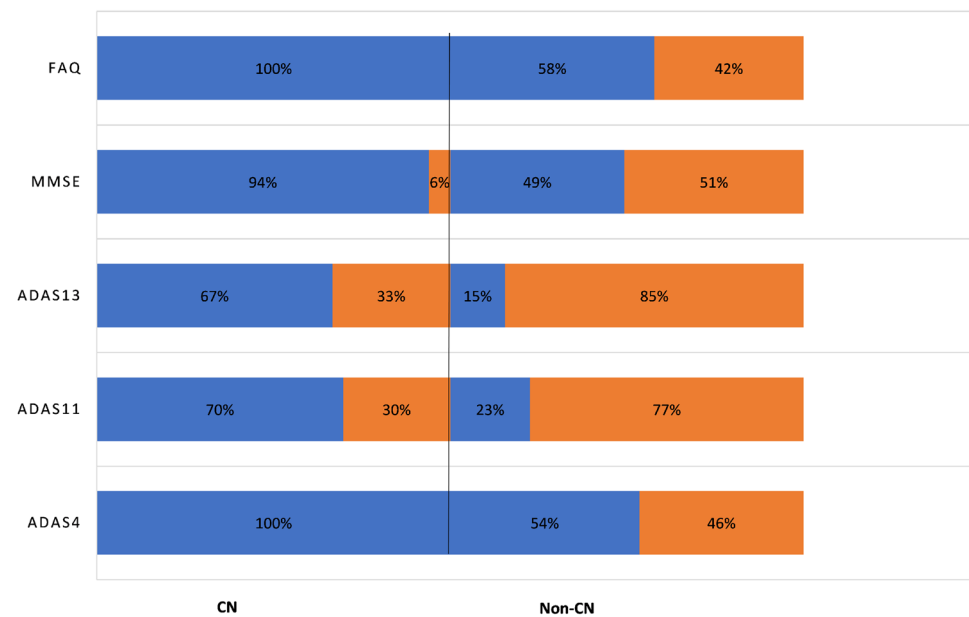
(b)

Fig. 7 Proportion of correct and incorrect diagnosis in CN and non-CN categories for validation data at month 72

Fig. 8 Comparison of actual and predicted values by sequence prediction models at month 48 for validation data (with normal cutoffs)

(a) MMSE Actual vs Predicted

(b) ADAS4 Actual vs Predicted

(c) ADAS11 Actual vs Predicted

(d) ADAS13 Actual vs Predicted

(e) FAQ Actual vs Predicted

(f) % of Scores on the Same Side of the Cutoff Line for Month 48

Test	% Same Side of Cutoff
MMSE	83.0%
ADAS4	97.0%
ADAS11	99.0%
ADAS13	88.0%
FAQ	93.0%

compared with the corresponding tests in Fig. 6b show that our prediction and diagnosis models perform better in reducing the misdiagnosis percentage values thereby promising to be a cost-effective solution for the initial screening of Alzheimer's patients.
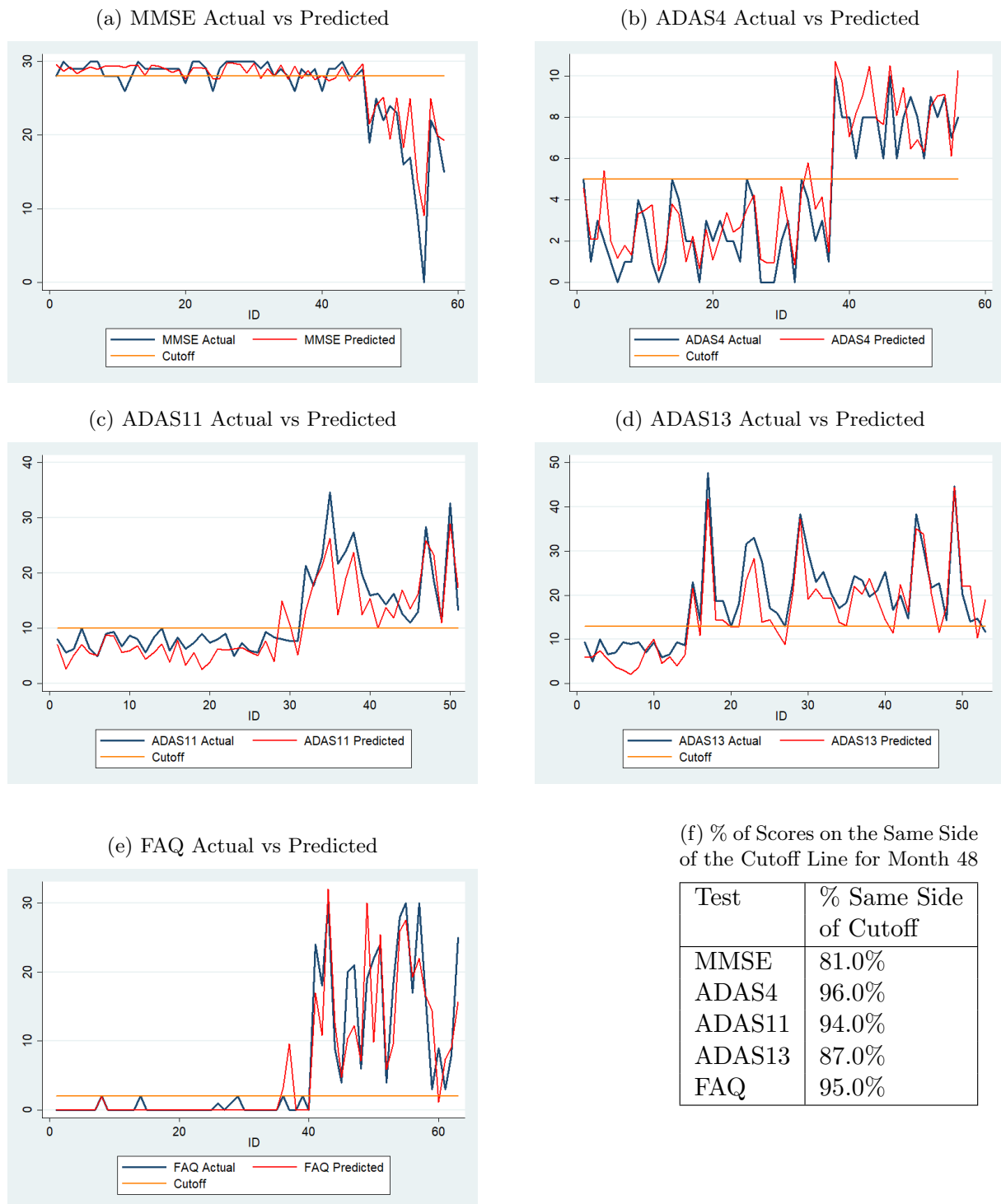
Fig. 9 Comparison of actual and predicted values by sequence prediction models at month 72 for validation data (with normal cutoffs)
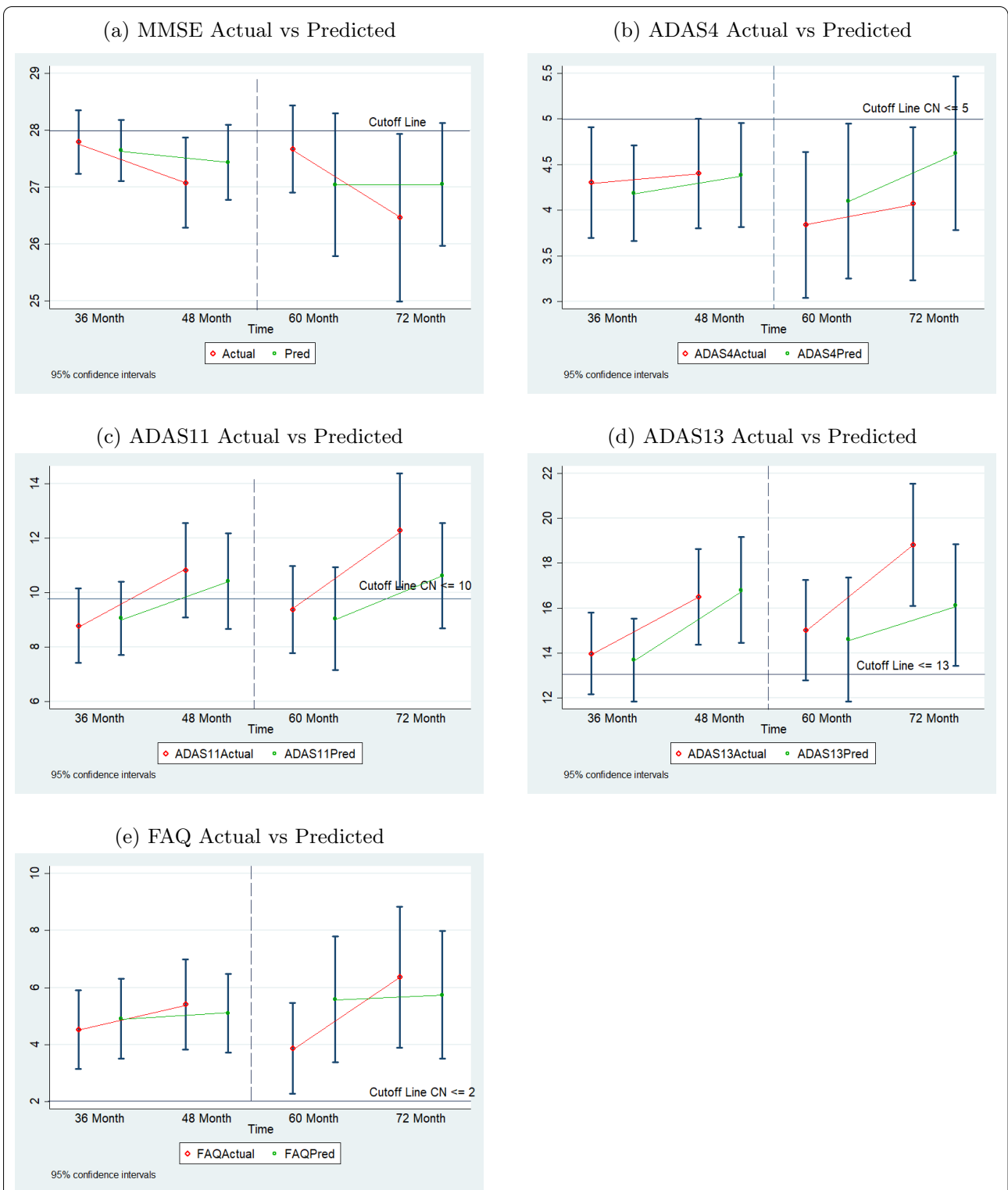
Fig. 10 95% confidence intervals of actual and predicted values for sequence prediction models for validation data (with normal cutoffs)

Fig. 11 Comparison of actual and predicted values by sequence prediction models at month 48 for 66 test dataset subjects (with normal cutoffs)

Fig. 12 Comparison of actual and predicted values by sequence prediction models at month 72 for 66 test dataset subjects (with normal cutoffs)
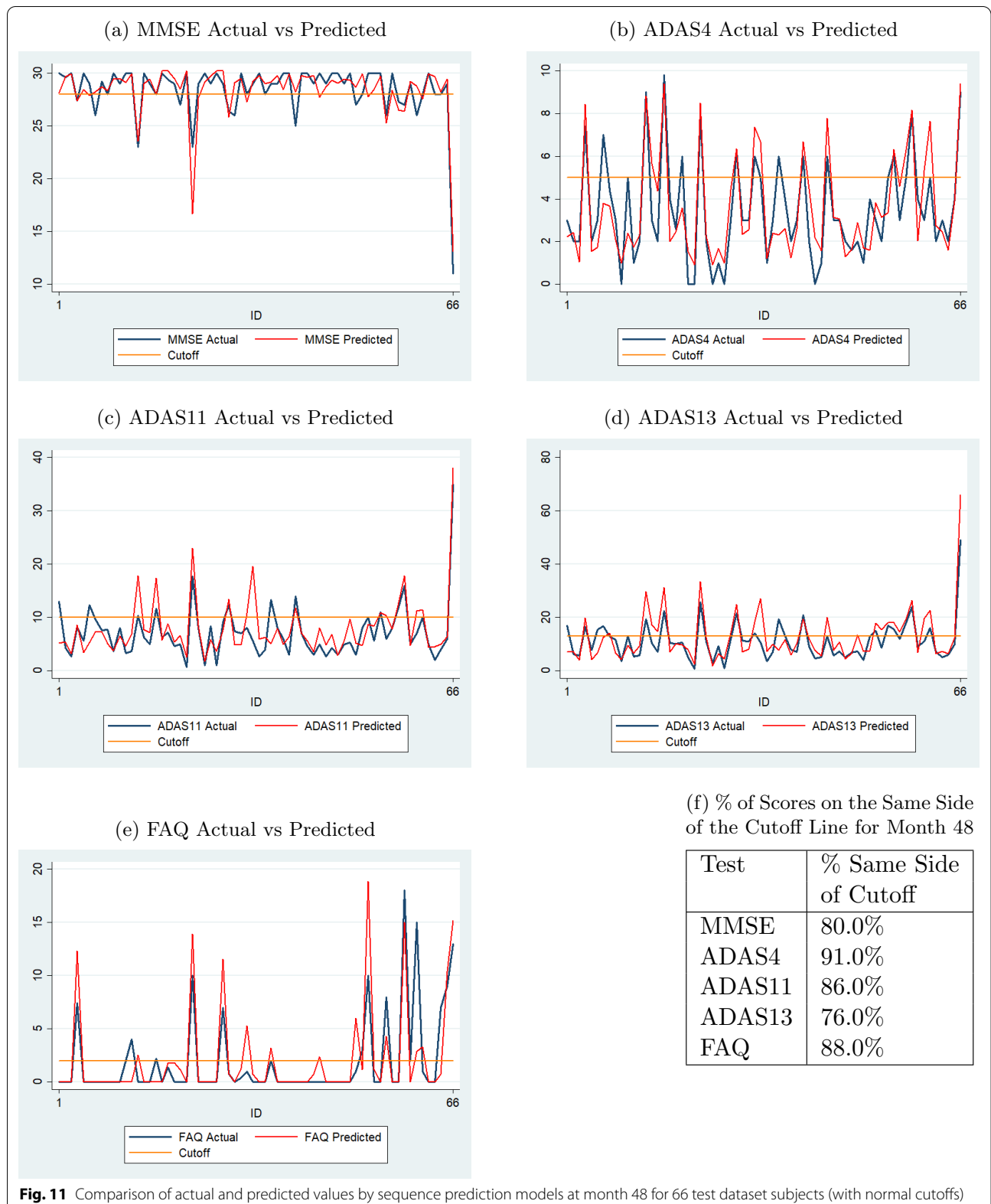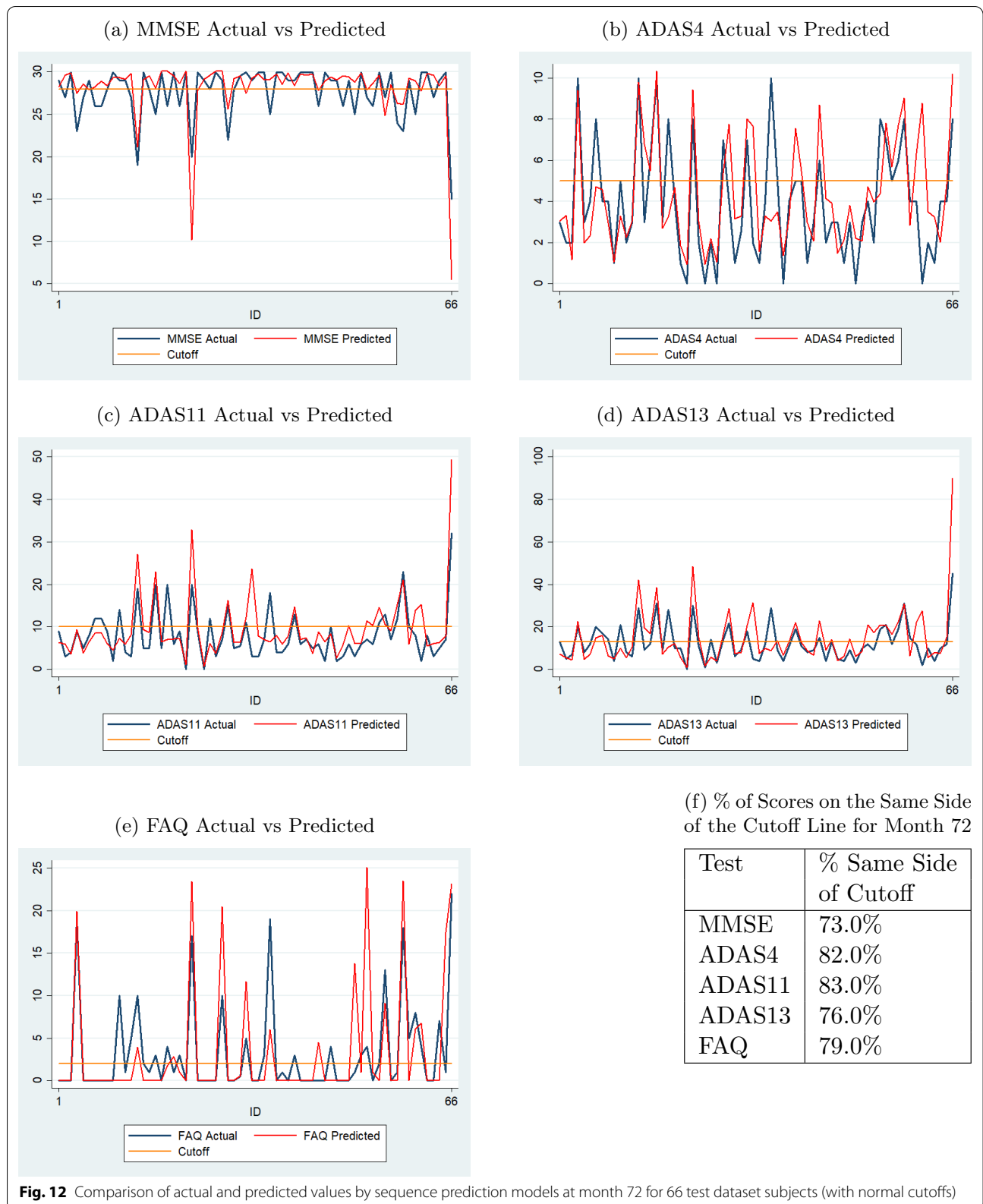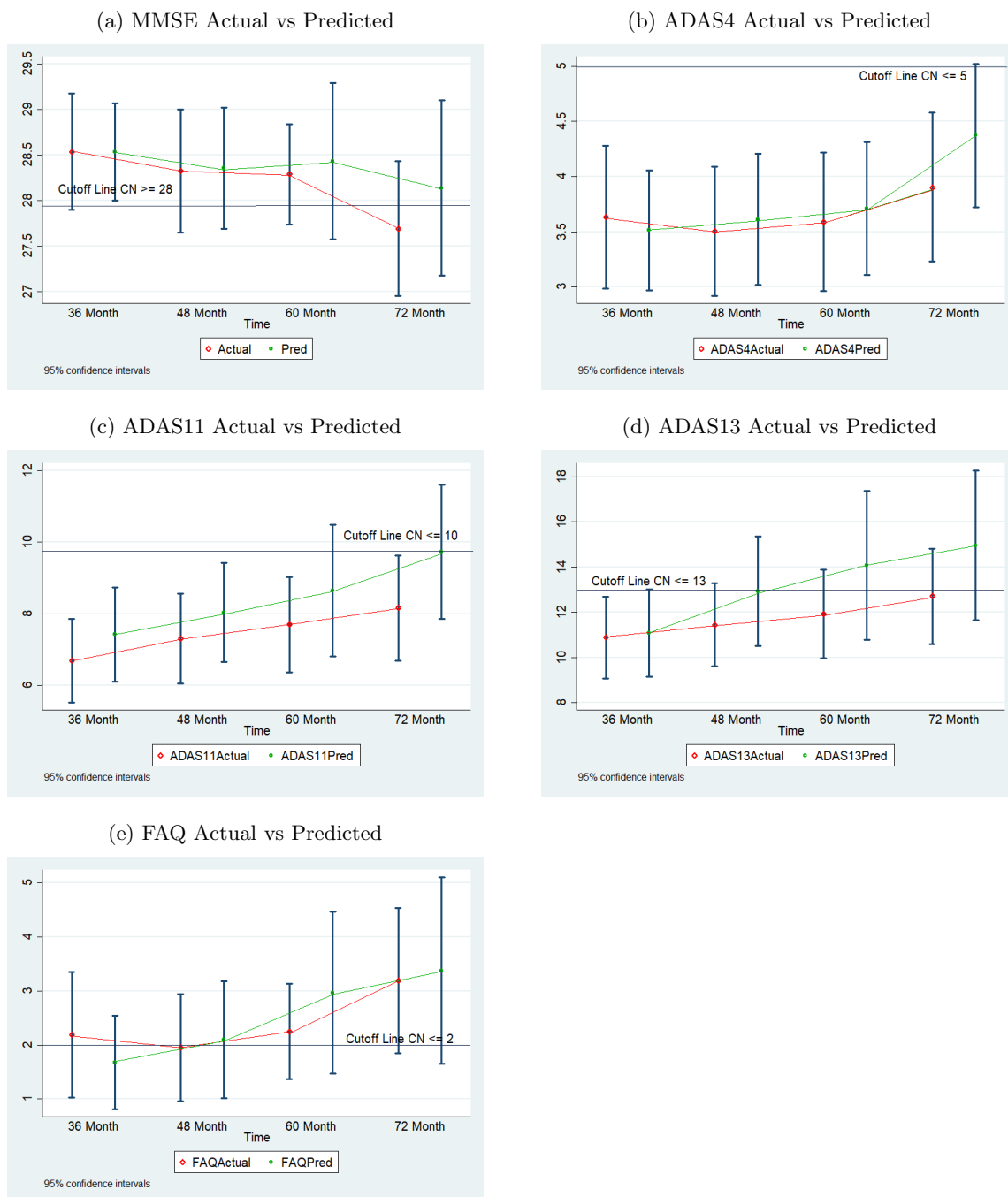
Fig. 13 95% confidence intervals of actual and predicted values for sequence prediction models for test data (with normal cutoffs)

Figure 7a and b similarly capture the misdiagnosis percentage values for predicted diagnosis and ADNI diagnosis outcomes at month 72 and it is clear once again that our prediction and diagnosis models perform better in reducing the misdiagnosis percentage values.

3.2 Performance of the sequence prediction models

Figure 8 consists of a series of graphs that demonstrate the performance of the sequence prediction models when predicting the sixth value based on the first four actual values and the predicted fifth value. The validation data sets were chosen for these figures. The blue line depicts the actual sixth value of each of the subjects, while the

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 19 of 26

red line depicts the predicted sixth value using the best sequence prediction model. The horizontal yellow line represents the threshold that differentiates test scores defined as CN or non-CN from Table 2. The comparisons show that even if the predicted values are not exactly identical to the actual values, they are close and more importantly, follow the general trend of the actual data. If the predicted values did not follow the general trend, the diagnoses based on these values would not be accurate. Moreover, the inaccuracy is of greater concern if the predicted and actual values were on two different sides of the cutoff line. The percent values in the table in Fig. 8f show that the percentages of predicted scores on the same side of the cutoff lines are high for all tests for month 48.

Figure 9 displays the performance of the sequence prediction models when predicting the eighth value, the 72nd month, based on the first four actual values and the predicted fifth, sixth, and seventh values. In this case, as well, the data are for the individuals in the validation data sets. The results are similar to those reported above for predicting the sixth value. The percent values in the table in Fig. 9f show that the percentages of predicted scores on the same side of the cutoff lines are high for all tests for month 72.

Figure 10 shows the 95 percent confidence intervals of actual and predicted values for validation data for each test. Since the validation data sets for predicting the month 36 and month 48 values are different from that for predicting the month 60 and month 72 values, the means of the 48-month and 60-month intervals could not be connected and the two sections of each chart are separated by a dashed line. The charts demonstrate that the predicted values follow the general trend of the actual data for each test and the spread of the predicted values are comparable to that of the actual data.

Figures 11 and 12 display results, similar to those in Figs. 8 and 9, for the 66 individuals in the test data set. The percent values in the table in Fig. 11f show that the percentages of predicted scores on the same side of the cutoff lines are high for all tests for month 48. The same is true for month 72 as shown in the table in Fig. 12f.

Figure 13 shows the 95 percent confidence intervals of actual and predicted values for test data for each test. The charts demonstrate that the predicted values follow the general trend of the actual data for each test and the spread of the predicted values are comparable to that of the actual data.

3.3 Combined model's accuracy

The key objective of this study is to determine if neuropsychological test scores can be used to predict if an individual will remain cognitively normal in the next 2–4 years. If the models, when combined together, predict that a patient will not remain CN over the next 2–4 years, then those individuals should be monitored more frequently and are likely to be recommended to undergo further testing.

Tables 6 and 7 collectively show that the combined model, when applied to the test data set, is successful in predicting that an individual will be cognitively normal at the end of the next 2 years with 84.62% accuracy. This accuracy drops to 83.33% for predictions that are 4 years in the future. Similarly, the combined model is successful in predicting that an individual will be cognitively abnormal at the end of the next 2 years with 78.9% accuracy. This accuracy rises to 82.8% for predictions that are 4 years in the future. Table 8 shows the overall accuracies of the combined model at months 48 (82.76%) and 72 (83.08%) which reiterates the efficacy of this this novel, two-steptechnique. The subjects predicted to be cognitively abnormal are good candidates for more invasive testing and are likely to serve as good candidates for clinical trials for AD treatments. Hence, even though the ADNI diagnosis is not solely based on these five neuropsychological tests, our technique shows that if we use just the outcomes of these five tests, we can determine with a high degree of accuracy how the disease will progress for patients for whom there are no other data available to aid the diagnosis.

Table 6 Combined model's prediction accuracy by category at month 48

Combined model's predictions		ADNI's assessment			Accuracy (%)
Diagnosis	No. of subjects	CN	Non-CN	NA	
CN	44	33	6	5	84.62
Non-CN	22	4	15	3	78.9

Table 7 Combined model's prediction accuracy by category at month 72

Combined model's predictions		ADNI's assessment			Accuracy (%)
Diagnosis	No. of subjects	CN	Non-CN	NA	
CN	37	30	6	1	83.33
Non-CN	29	5	24	0	82.8

Table 8 Combined model's overall diagnosis accuracies

Diagnosis time	Accuracy (%)
Month 48	82.76
Month 72	83.08
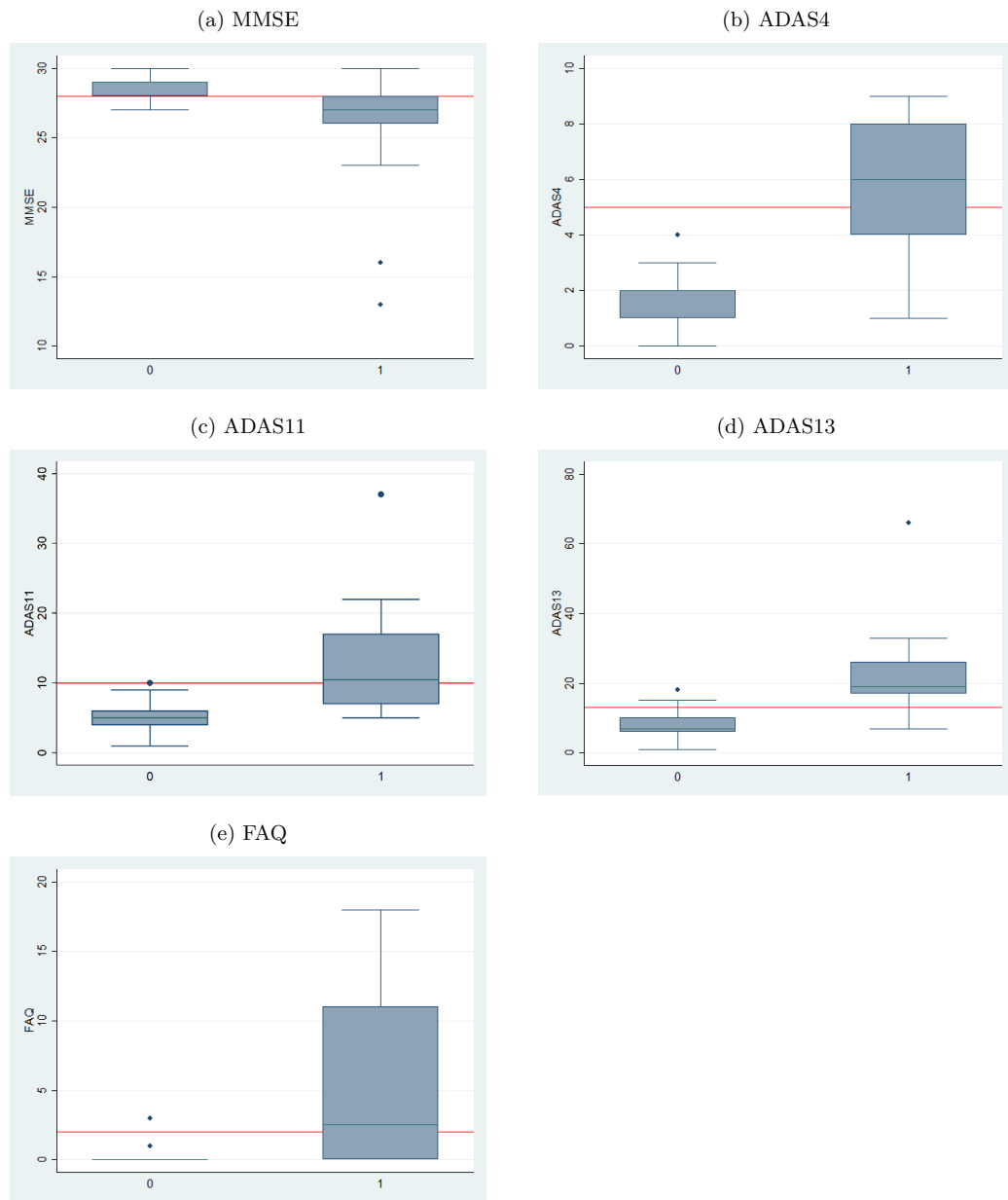
(a) MMSE

(b) ADAS4

(c) ADAS11

(d) ADAS13

(e) FAQ

(f) Boxplot and Descriptive Statistics

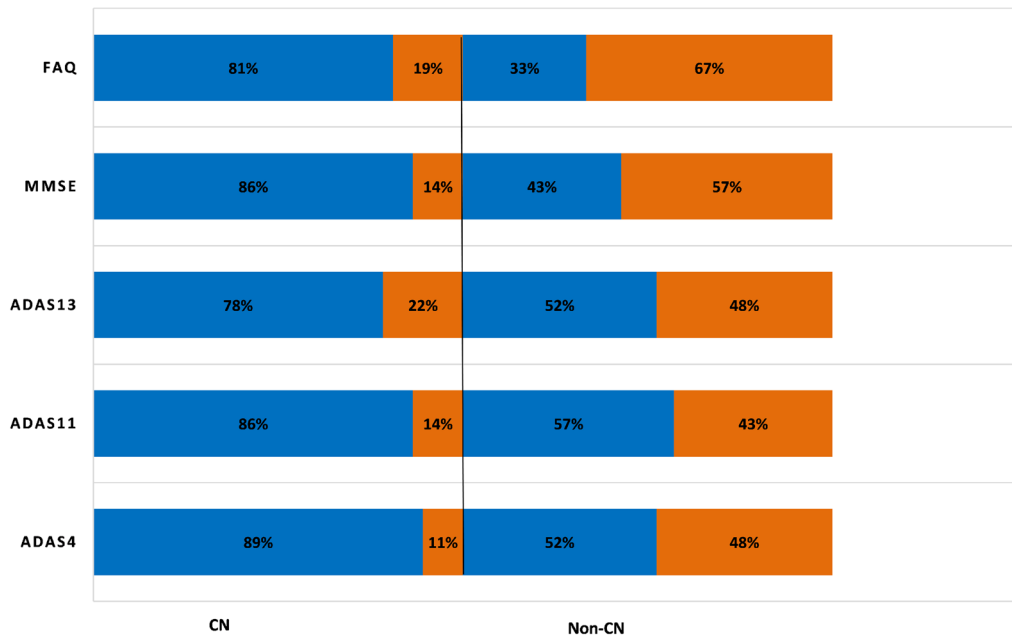| CN or Non-CN	Test	Low Adj Value	p25	Median	p75	Up Adj Value	Mean	Min	Max
CN	MMSE	28	29	29	30	30	29	28	30
Non-CN	MMSE	24	26	28	29	30	27	13	30
CN	ADAS4	1	2	2	3	4	2	1	4
Non-CN	ADAS4	2	5	7	8	9	6	2	9
CN	ADAS11	2	5	5	7	10	6	2	11
Non-CN	ADAS11	5	8	11	17	23	13	5	38
CN	ADAS13	2	6	7	10	15	8	2	18
Non-CN	ADAS13	7	17	20	26	33	22	7	66
CN	FAQ	0	0	0	0	0	0	0	3
Non-CN	FAQ	0	1	3	12	19	6	0	19

Fig. 14 Ranges of test data split by the predicted diagnosis (CN or non-CN) assigned by the diagnosis model at month 48 (with normal cutoffs)

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 21 of 26

MONTH 48 PREDICTED TEST DATA AND PREDICTED DIAGNOSIS

	CN	Non-CN
FAQ	98% / 2%	36% / 64%
MMSE	95% / 5%	41% / 59%
ADAS13	89% / 11%	14% / 86%
ADAS11	98% / 2%	41% / 59%
ADAS4	100%	27% / 73%
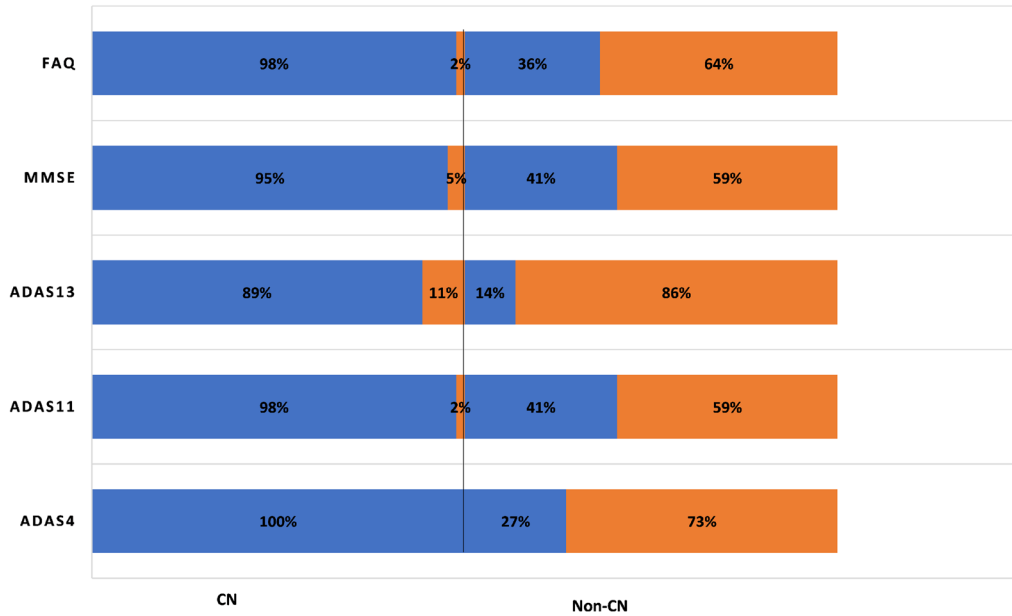
(a)

MONTH 48 ADNI TEST DATA AND ADNI DIAGNOSIS

	CN	Non-CN
FAQ	92% / 8%	57% / 43%
MMSE	95% / 5%	57% / 43%
ADAS13	89% / 11%	48% / 52%
ADAS11	92% / 8%	67% / 33%
ADAS4	95% / 5%	57% / 43%
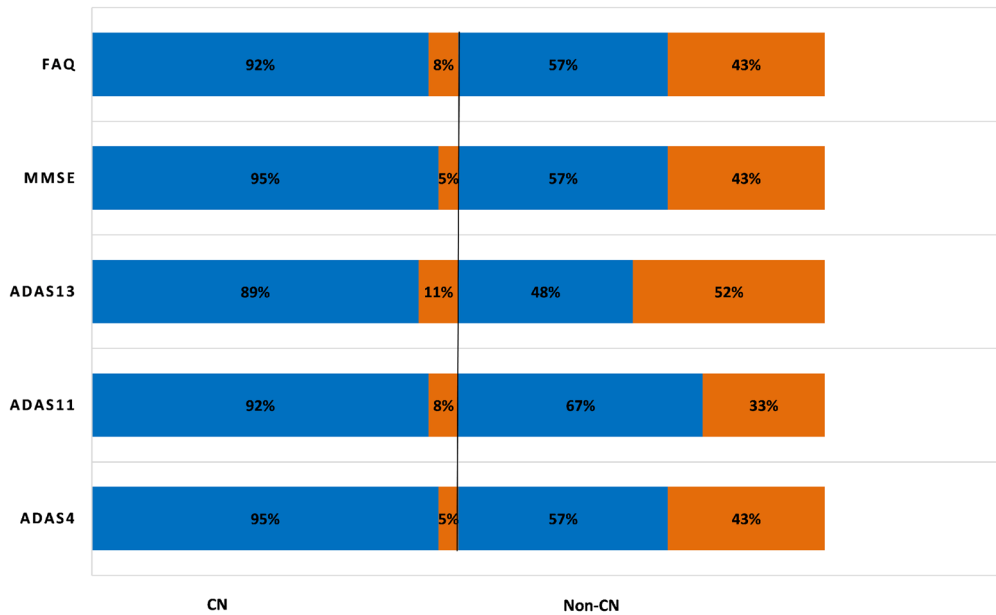
(b)

Fig. 15 Proportion of correct and incorrect diagnosis in CN and non-CN categories for test data and predicted test data at month 48
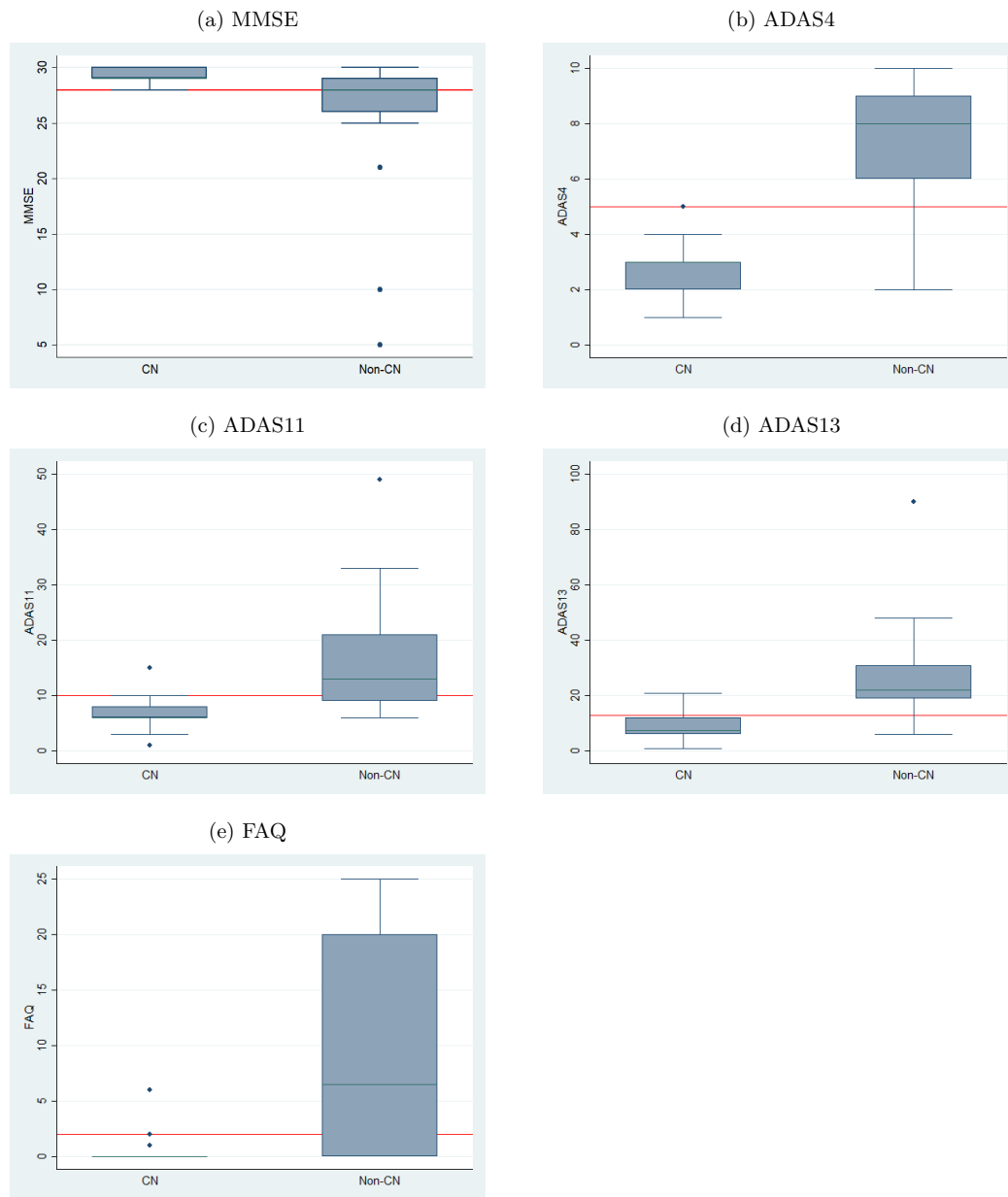
(a) MMSE

(b) ADAS4

(c) ADAS11

(d) ADAS13

(e) FAQ

(f) Boxplot and Descriptive Statistics

CN or Non-CN	Test	Low Adj Value	p25	Median	p75	Up Adj Value	Mean	Min	Max
CN	MMSE	28	29	29	30	30	29	28	30
Non-CN	MMSE	25	26	28	29	30	26	5	30
CN	ADAS4	1	2	3	3	4	3	1	5
Non-CN	ADAS4	2	6	8	9	10	7	2	10
CN	ADAS11	3	6	6	8	10	7	1	15
Non-CN	ADAS11	6	9	13	21	33	16	6	49
CN	ADAS13	1	6	7.5	12	21	9	1	21
Non-CN	ADAS13	6	19	22	31	48	27	6	90
CN	FAQ	0	0	0	0	0	0	0	6
Non-CN	FAQ	0	0	6.5	20	25	10	0	25

Fig. 16 Ranges of test data split by the predicted diagnosis (CN or non-CN) assigned by the diagnosis model at month 72 (with normal cutoffs)

Fig. 17 Proportion of correct and incorrect diagnosis in CN and non-CN categories for test data and predicted test data at month 72

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 24 of 26

The boxplots of Figs. 14 and 16 serve as a corroboration of the accuracies noted above. The plots also corroborate the accuracy of the normal–abnormal cutoff values shown in Table 2. The boxplots for each graph are split into two groups, CN and non-CN. CN depicts the range of scores diagnosed as 0 by the diagnosis models, and non-CN depicts those diagnosed as 1. The stark contrast in the ranges of the CN and non-CN boxplots shows that the diagnosis models were able to predict very effectively on data that it had not seen before during training and validation. The tables in Figs. 14f and 16f contain the descriptive statistics of the boxplots along with the minimum value, maximum value, and the mean of the boxplot ranges for each of the five tests.

Figure 15a shows a bar graph for all the five tests in which the CN and non-CN groups are further split into normal and abnormal percentage values based on the cutoff values for the test from Table 2. For example, for MMSE, the CN group contains 98% RIDs that are in the normal range and 2% RIDs that are in the abnormal range. Similarly, for MMSE, the non-CN group contains 41% RIDs that are in the normal range and 59% RIDs that are in the abnormal range. As before, in regard to the diagnosis reached by ADNI, as shown in Fig. 15b, the mix of normal–abnormal misdiagnosis percentages is not any better when we consider the test data and group them by the ADNI DX value (CN and non-CN, that is, MCI/AD). As seen in Fig. 15b, for MMSE, say, the non-CN group contains 57% RIDs that are in the normal range (compared to 41% using the predicted diagnosis of our models) and run the risk of being monitored and treated causing a waste of valuable resources. All the tests in Fig. 15a, when compared with the corresponding tests in Fig. 15b show that our prediction and diagnosis models perform better in reducing the misdiagnosis percentage values thereby promising to be a cost-effective solution for the initial screening of Alzheimer's patients.

Figures 16 and 17 similarly capture the misdiagnosis percentage values for predicted diagnosis and ADNI diagnosis outcomes at month 72 and it is clear once again that our prediction and diagnosis models perform better in reducing the misdiagnosis percentage values.

4 Conclusion

The contribution of this work is the demonstration that simple neuropsychological tests can be used on their own to predict the future cognitive state of individuals. Using data from hundreds of subjects who participated in the ADNI project, this study used LSTM models to predict future values of their tests and used MLPs to diagnose individuals as CN or non-CN at a future date. For a cohort of 66 test subjects, the test scores of all five tests were combined to generate a prediction of their future cognitive states. These predictions were then compared with the actual CN or MCI/AD status assigned by ADNI. The results show that individuals who are predicted by the model to continue to remain CN in all of the tests are highly likely to remain cognitively normal in real life over the next 2 to 4 years. Individuals who are predicted to have test scores outside of normal ranges are likely to experience cognitive impairment in the future. These individuals should be monitored and treated.

The goal of this work was not to supersede the cognitive assessments made by ADNI. Instead, the goal was to determine whether the cognitive assessments made by ADNI could be used as outputs to train MLPs which relied only on a subset of the set of features that ADNI gathers for each participant (namely, the five neuropsychological tests). Data for these five features are relatively easy to collect and the two-step neural network algorithm can be used to forecast cognitive status for a 2- to 4-year window.

The analysis of this paper focused on only five tests for which data over at least a 72-month period was available in the ADNI database for a large number of individuals and there were no more than two missing values for each participant. The combined model's structure allows one to introduce any test for which sufficient longitudinal data are available from ADNI or any other data source such as NACC (National Alzheimer's Coordinating Center).

The benefit of developing a tool based on neuropsychological test scores that can predict with high accuracy the likelihood of an individual experiencing cognitive difficulties in the next few years derives from the low cost of administering these tests. If these tests become routine for individuals above a certain age or who have some risk factors for developing cognitive impairment, longitudinal performance data can be used with a high level of accuracy to determine which patients will require close monitoring. As treatments for Alzheimer's disease continue to develop, this ability to determine who will require close monitoring can allow more invasive and expensive tests to be reserved for them. This can also allow for early intervention, which can be crucial in the treatment or prevention of the disease.

Mukherji *et al. Brain Informatics*   (2022) 9:23

Page 25 of 26

Author contributions

DM and MM contributed equally to this work. NM contributed to the formal analysis, visualization, and manuscript preparation. All authors read and approved the final manuscript.

Authors' information

This research combines the various interests and expertise of the authors. The main research interest that has motivated the work stems from DM's interest in AD as a student of Neuroscience. As such, DM contributed to the main research concept and methodology of this project. MM's expertise as a computer scientist was critical for the machine learning and neural network programming aspects of the project. NM's expertise in econometrics as a professor of economics was helpful for the statistical analysis and graphical representations of the paper. All authors contributed equally in writing the manuscript.

Funding

Availability of data

Data used in this project are available from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Users need access permission from ADNI.

Declarations

Ethics approval and consent to participate

ADNI provided access to all data used in this study and the authors used their guidelines for acknowledging them in the author section. The manuscript received approval from ADNI for submission of the research for publication. All authors of the study have provided consent for submitting this research for publication.

Consent for publication

All authors have reviewed the final copy of the manuscript and have provided consent for its publication.

Competing interests

The authors declare that there are no competing interests.

Author details

[1]University of Michigan, Ann Arbor, MI, USA. [2]Rochester, MI, USA. [3]Oakland University, Rochester, MI, USA.

References

1. Report (2018) 2018 Alzheimer's disease facts and figures. Alzheimer's Dement. 14(3):367–429
2. Brookmeyer R, Johnson E, Ziegler-Graham K, Michael Arrighi H (2007) Forecasting the global burden of Alzheimer's disease. Alzheimer's Dement 3(3):186–191
3. Cummings JL, Morstorf T, Zhong K (2014) Alzheimer's disease drug development pipeline: few candidates, frequent failures. Alzheimer's Res Ther 6(4):37
4. Berti V, Polito C, Lombardi G, Ferrari C, Sorbi S, Pupi A (2016) Rethinking on the concept of biomarkers in preclinical Alzheimer's disease. Neurol Sci 37(5):663–672
5. Dubois B, Hampel H et al (2016) Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimer's Dement 12(3):292–323
6. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Szoeke C, Lance MS, Martins R, Maruff P, Ames D, Rowe CC, Masters CL (2013) Amyloid deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 12(4):357–367
7. Goudey B, Fung BJ, Schieber C, Faux NG (2019) A blood-based signature of cerebrospinal fluid A1–42 status. Sci Rep 9(1):4163
8. Sevigny J, Suhy J, Chiao P, Chen T, Klein G, Purcell D, Oh J, Verma A, Sampat M, Barakos J (2016) Amyloid PET screening for enrichment of early-stage alzheimer disease clinical trials: experience in a phase 1b clinical trial. Alzheimer Dis Assoc Disord 30(1):1–7
9. Shi L, Baird AL, Westwood S, Hye A, Dobson R, Thambisetty M, Lovestone S (2018) A decade of blood biomarkers for Alzheimer's disease research: an evolving field, improving study designs, and the challenge of replication. J Alzheimer's Dis 62(3):1181–1198
10. Cummings J, Lee G, Mortsdorf T, Ritter A, Zhong K (2017) Alzheimer's disease drug development pipeline: 2017. Alzheimer's Dement Transl Res Clin Interv 3(3):367–384
11. Albert MS, Moss MB, Tanzi R, Jones K (2001) Preclinical prediction of AD using neuropsychological tests. J Int Neuropsychol Soc 7(5):631–639
12. Battista P, Salvatore C, Castiglioni I (2017) Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: a machine learning study. Behav Neurol 2017:1–19
13. Bondi MW, Jak AJ, Delano-Wood L, Jacobson MW, Delis DC, Salmon DP (2008) Neuropsychological contributions to the early identification of Alzheimer's disease. Neuropsychol Rev 18(1):73–90
14. Gainotti G, Quaranta D, Vita MG (2014) Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease'. J Alzheimer's Dis 38(3):481–495
15. Tierney MC, Yao C, Kiss A, McDowell I (2005) Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology 64(11):1853–1859
16. Weakley A, Williams JA, Schmitter-Edgecombe M, Cook DJ (2015) Neuropsychological test selection for cognitive impairment classification: a machine learning approach. J Clin Exp Neuropsychol 37(9):899–916
17. Pimentel EML (2009) Role of neuropsychological assessment in the differential diagnosis of Alzheimer's disease and vascular dementia. Dement Neuropsychol 3(3):214–221
18. Wang T, Qiu RG, Ming Y (2018) Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks. Sci Rep 8(1):9161
19. Tabert MH, Manly JJ, Liu X, Pelton G, Rosenblum S, Jacobs M, Zamora D, Goodkind M, Bell KL, Stern Y, Devanand DP (2017) Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. JAMA Psychiatry (Chicago, Ill.)
20. Seyed HH, Babajani-Feremi A, The Alzheimer's Disease Neuroimaging Initiative (2022) Prediction and modeling of neuropsychological scores in Alzheimer's disease using multimodal neuroimaging data and artificial neural networks. Front Comput Neurosci 15:769982
21. Tabarestani S, Aghili M, Eslami M, Cabrerizo M, Barreto A, Rishe N, Curiel RE, Loewenstein D, Duara R, Adjouadi M (2020) A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. NeuroImage (Orlando, Fla) 206:116317
22. Liu M, Zhang J, Lian C, Shen D (2020) Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores. IEEE Trans Cybern 50(7):3381–3392
23. Liu M, Cheng D, Yan W (2018) Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform 12:35
24. Kung T-H, Chao T-C, Xie Y-R, Pai M-C, Kuo Y-M, Lee GGC (2021) Neuroimage biomarker identification of the conversion of mild cognitive impairment to Alzheimer's disease. Front Neurosci 15:584641
25. Seyed HH, Ebrahimzadeh A, Babajani-Feremi A (2019) Identification of the early stage of Alzheimer's disease using structural MRI and resting-state fMRI. Front Neurol 10:904
26. Eskildsen SF, Coupe P, Garcıa-Lorenzo D, Fonov V, Pruessner JC, Louis CD (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage (Orlando, Fla) 65:511–521
27. Ferrarini L, Palm WM, Olofsen H, van der Landen R, Blauw GJ, Westendorp RGJ, Bollen LEM, Middelkoop HAM, Reiber JHC, van Buchem MA, Admiraal Behloul F (2008) MMSE scores correlate with local ventricular enlargement in the spectrum from cognitively normal to Alzheimer disease. NeuroImage (Orlando, Fla) 39(4):1832–1838

28. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Aparici CM, Behr SC, Flavell RR, Huang S-Y, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Pampaloni MH, Hadley D, Franc BL (2019) A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology 290(2):456–464

29. Crane PK, Carle A, Gibbons LE, Insel P, Scott Mackin R, Gross A, Jones RN, Mukherjee S, McKay Curtis S, Harvey D, Weiner M, Mungas D (2012) Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imaging Behav 6(4):502–516

30. Albright J (2019) Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. Alzheimer's Dement Transl Res Clin Interv 5(1):483–491

31. Chen T, Shang C, Su P, et al (2021) Assessing significance of cognitive assessments for diagnosing alzheimer's disease with fuzzy-rough feature selection. In: Advances in computational intelligence systems, Cham, Springer International Publishing, pp 450–462

32. Mugruza-Vassallo CA (2016) Database methodology for therapy evaluation in auditory schizophrenia disorder based on continuity evolution of symptoms. In: IEEE Oct 2016, pp 298–303

33. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J (2016) Doctor AI: predicting clinical events via recurrent neural networks. In: JMLR workshop and conference proceedings, Aug 2016, 56, pp 301–318

34. Esteban C, Staeck O, Yang Y, Tresp V (2016) Predicting clinical events by combining static and dynamic information using recurrent neural networks

35. Lipton ZC, Kale DC, Elkan C, Wetzel R (2015) Learning to diagnose with LSTM recurrent neural networks

36. Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning

37. Mostafa MG, Sørensen L, Ourselin S, Nielsen M (2021) CARRNN: a continuous autoregressive recurrent neural network for deep representation learning from sporadic temporal data

38. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34(7):939

39. Yagi T, Kanekiyo M, Ito J, Ihara R, Suzuki K, Iwata A, Iwatsubo T, Aoshima K (2019) Identification of prognostic factors to predict cognitive decline of patients with early Alzheimer's disease in the Japanese Alzheimer's Disease Neuroimaging Initiative study. Alzheimer's Dement Transl Res Clin Interv 5(1):364–373

Publisher's Note