

PPAD: a deep learning architecture to predict progression of Alzheimer's disease

Mohammad Al Olaimat¹, Jared Martinez¹, Fahad Saeed², Serdar Bozdag^{1,3,4,*},
Alzheimer's Disease Neuroimaging Initiative[‡]

¹Department of Computer Science and Engineering, University of North Texas, Denton, TX, United States

²School of Computing and Information Sciences, Florida International University, Miami, FL, United States

³Department of Mathematics, University of North Texas, Denton, TX, United States

⁴BioDiscovery Institute, University of North Texas, Denton, TX, United States

*Corresponding author. Department of Computer Science and Engineering, University of North Texas, 3940 N. Elm St., Denton, TX, 76207, United States.

E-mail: serdar.bozdag@unt.edu (S.B.)

[‡]Data used in preparation of this article were 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.

Abstract

Motivation: Alzheimer's disease (AD) is a neurodegenerative disease that affects millions of people worldwide. Mild cognitive impairment (MCI) is an intermediary stage between cognitively normal state and AD. Not all people who have MCI convert to AD. The diagnosis of AD is made after significant symptoms of dementia such as short-term memory loss are already present. Since AD is currently an irreversible disease, diagnosis at the onset of the disease brings a huge burden on patients, their caregivers, and the healthcare sector. Thus, there is a crucial need to develop methods for the early prediction AD for patients who have MCI. Recurrent neural networks (RNN) have been successfully used to handle electronic health records (EHR) for predicting conversion from MCI to AD. However, RNN ignores irregular time intervals between successive events which occurs common in electronic health record data. In this study, we propose two deep learning architectures based on RNN, namely Predicting Progression of Alzheimer's Disease (PPAD) and PPAD-Autoencoder. PPAD and PPAD-Autoencoder are designed for early predicting conversion from MCI to AD at the next visit and multiple visits ahead for patients, respectively. To minimize the effect of the irregular time intervals between visits, we propose using age in each visit as an indicator of time change between successive visits.

Results: Our experimental results conducted on Alzheimer's Disease Neuroimaging Initiative and National Alzheimer's Coordinating Center datasets showed that our proposed models outperformed all baseline models for most prediction scenarios in terms of F2 and sensitivity. We also observed that the age feature was one of top features and was able to address irregular time interval problem.

Availability and implementation: <https://github.com/bozdaglab/PPAD>.

1 Introduction

Alzheimer's disease (AD) is an irreversible neurodegenerative disease that leads to problems in cognitive functioning (e.g. memory loss and impaired reasoning) and behavioral changes (e.g. aggression, wandering, and anxiety) (Hampel et al. 2011; Rodríguez and Verkhatsky 2011; Huang et al. 2016; Cui et al. 2019). Unusual accumulation of amyloid plaques and neurofibrillary tangles in the brain are considered as the main causes of AD (Perrin et al. 2009; Tong et al. 2015; Patterson 2018; Lee et al. 2019). According to the World Health Organization, there are about 40 million AD cases worldwide. In the United States, there are about 6 million AD cases, and this number is expected to reach 14 million by 2050 (Alzheimer's Association 2015; Venugopalan et al. 2021).

Mild cognitive impairment (MCI) is an intermediary stage between cognitively normal (CN) state and AD (Petersen et al. 2014; Petersen, 2016; Patterson 2018). MCI is determined through an impairment of memory on standard tests with the absence of significant impairment in daily living activities and dementia (Winblad et al. 2004). Using a standardized test, impairment on cognitive is defined as performance below 1.5 SD of the age-, sex-, and education-adjusted

normative mean; according to the test, MCI can be classified based on the severity into early and late MCI. Early MCI (EMCI) refers to a case when the performance is between 1.0 SD and 1.5 SD below the normative mean on the test, whereas late MCI (LMCI) refers to a case when the performance is 1.5 SD below the normative mean on the test (Aisen et al. 2010; Jessen et al. 2014). About 15% of MCI patients convert to AD per year while 80% of MCI patients convert to AD within about 6 years (Tábuas-Pereira et al. 2016). MCI cases who progress to AD eventually are called MCI-converter and MCI cases who stay as MCI or revert to CN are called MCI-nonconverter.

The diagnosis of AD can be made after significant symptoms of dementia such as short-term memory loss are already present (Frost et al. 2010). The diagnosis after the onset of the disease creates emotional burden for patients and their family members and economic burden to the healthcare. The estimated healthcare cost for AD was over \$300 billion in 2020 (Wong 2020). As a result, developing a robust method that can early predict conversion from MCI to AD is crucial for patients to have better treatments, interventions to delay or prevent AD progression.

Electronic health record (EHR) is a sequential data represented as temporal sequences of clinical features and has been used to train machine learning models to classify and cluster patient records for improving clinical decision-making (Menachemi and Collum 2011). Traditional machine learning methods, however, aggregate clinical features; thus, ignores temporal relations between these sequences (Martí-Juan *et al.* 2020; Tanveer *et al.* 2020). Recurrent neural network (RNN) is a deep learning model used to process sequential data and maintains temporal relations between sequences (Goodfellow *et al.* 2016). Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) and gated recurrent unit (GRU) (Gers *et al.* 2000; Cho *et al.* 2014) are RNN variants that have the capability to handle long-term dependencies, which is considered the drawback of the vanilla RNN. The main difference between the GRU and LSTM is their complexity (i.e. number of learnable parameters). The GRU is a less complex architecture than LSTM, thus are more preferable especially when training data are not abundant (Greff *et al.* 2017).

To identify biomarkers of conversion from MCI to AD, various machine learning methods have been utilized. Support vector machine (SVM), a popular technique for classification and regression (Bisgin *et al.* 2018) has been widely used for the diagnosis of AD using a single data modality such as magnetic resonance imaging (MRI) data (Retico *et al.* 2015; Plochanski *et al.* 2016; Alam *et al.* 2017; Tangaro *et al.* 2017; Sheng *et al.* 2019). In Zeng *et al.* (2018), an optimized SVM based on a new switching delayed particle swarm optimization (PSO) was proposed for the classification of AD and MCI using MRI. The proposed model achieved 72.94% accuracy for static MCI versus AD classification, and 57.14% accuracy for progressive MCI versus AD classification. In Sun *et al.* (2018), spatial-anatomical information was integrated to MRI data to train and evaluate a new SVM model for the classification of AD. The proposed model achieved 65.7% accuracy to differentiate between MCI and AD. In Zhang *et al.* (2012), SVM was applied with a multitask learning to identify AD biomarkers and predict the 2 years conversion from MCI to AD using baseline measurements from MRI and cerebrospinal fluid (CSF). The proposed model achieved 73.9% accuracy, 68.6% sensitivity, and 73.6% specificity. In Cheng *et al.* (2015), a domain transfer learning model was proposed for using not only MCI samples but also AD and CN as auxiliary samples to identify biomarkers that can be used for a classification task to distinguish between MCI-converter and MCI-nonconverter samples. The proposed model achieved 79.4% accuracy, 84.5% sensitivity, and 72.7% specificity.

Moreover, artificial neural networks (ANN) have been used for the classification of AD using biomarkers or MRI data (Huang *et al.* 2008; Joshi *et al.* 2010; Quintana *et al.* 2012; Al-Naami *et al.* 2013; Aljović *et al.* 2016). In Wang *et al.* (2015), three variants of feed-forward neural network (FNN) were proposed based on PSO and artificial bee colony (ABC) to differentiate between normal and abnormal brain using MRI data. The proposed recombination of PSO and ABC FNN (FNN-HPA) model achieved 99.45% accuracy, 99.37% sensitivity, and 100% specificity. In Kar and Majumder (2019), diffusion tensor visualization-based neuro-fuzzy classification ANN model was proposed to differentiate between CN and AD using MRI. The proposed model was able to achieve 100% accuracy. However, these results were based on a small sample set (i.e. 9 AD and 11 CN patients).

Integration of multimodality data has been performed to improve the performance of predicting conversion from MCI to AD by extracting AD biomarkers from each modality. A graph-based semisupervised learning method that integrates brain image data from MRI and positron emission tomography (PET) was proposed to distinguish between EMCI from CN cases (Kim *et al.* 2013). The proposed method achieved 68.5% accuracy, 53.4% sensitivity, and 77% specificity. In Lee *et al.* (2019), a multi-modal GRU model was trained using longitudinal cognitive performance and CSF biomarkers data, and cross-sectional neuroimaging and demographic data to predict MCI to AD conversion. The results showed that the proposed model achieved 81% accuracy and an area under the receiver operation characteristics curve (AUC) of 86%. In Venugopalan *et al.* (2021), an integrative classification method was proposed to classify patients into AD, MCI, and CN. The model was trained on clinical and genetic features extracted using stacked denoising autoencoders and brain image features extracted using convolutional neural network (CNN). The results have shown that integrating multimodality data outperforms single modality models. In Nguyen *et al.* (2020), an RNN model was applied to the longitudinal cognitive performance, MRI, and CFS data of 1677 samples in Alzheimer's Disease Neuroimaging Initiative (ADNI) database to predict the diagnosis of patients in the future up to 6 years and achieved better performance than baseline models. In Li and Fan (2019), a deep learning model based on an LSTM autoencoder was proposed to extract the hidden temporal pattern in longitudinal data for five cognitive measures for 1-year follow-up. The new extracted features were combined with baseline hippocampal measures extracted from MRI scans to train and evaluate a prognostic model using Cox regression to predict AD progression for MCI individuals.

In analyzing longitudinal biomedical data, irregular time intervals between clinical visits poses a technical challenge. Deep learning methods that can handle sequential data (e.g. RNN) assume equal intervals between inputs in the sequence. To address this issue, time-aware LSTM (T-LSTM) was proposed to modify the memory state of the current cell state based on the time gap between the current and previous cell states (Baytas *et al.* 2017). The results on progression of Parkinson's disease data showed improved performance than baseline methods. In another study, T-LSTM was evaluated on synthetic and real data for chronic kidney disease (Luong and Chandola 2018). The results showed that T-LSTM autoencoder can be used to deal with sequential data to generate the latent space of the longitudinal profile, but the latent space of the longitudinal of real data was not able to subtype chronic kidney disease.

Studies have shown an association between AD and several genes such as CTNNA3, GAB2, PVRL2, and TOMM40. Among these, epsilon4 allele of apolipoprotein E gene (APOE ϵ 4) is the most important genetic risk factor for AD (Chalmers *et al.* 2003; Scheltens *et al.* 2021). In addition, demographic such as age, gender, alcohol consumption, smoking, depression, head injury, education, race, ethnicity, and nutrition have also been reported as nongenetic risk factors (Hall *et al.* 1998; Ikeda and Yamada 2010; Patterson 2018). As a result, these genetic and nongenetic risk factors can play role if they are utilized by a predictive model for AD progression.

In general, irregular number of visits for patients, irregular intervals between visits, and missing values are common drawbacks of EHR (Yadav et al. 2018). Since datasets from AD cohorts suffer from the same drawbacks, there is a crucial need to develop methods for early predicting conversion from MCI to AD while addressing such data irregularities. Most of the existing methods do not consider the irregular time intervals between consecutive visits and give equal weight to sensitivity and specificity of the model. However, increasing sensitivity (i.e. correctly predicting individuals who would convert to AD) is more important. Moreover, several existing studies do not integrate longitudinal data with cross-sectional demographic data such as gender, race, ethnicity, patients' education, and APOE ϵ 4. Most tools are also not made publicly available, which limits their application to new datasets.

To address these limitations, in this study, we propose two open-source deep learning models, PPAD and PAD-AE, for early predicting conversion from MCI to AD at the next visit and multiple visits ahead for patients, respectively. The novelty of our proposed models is three-fold. PPAD and PPAD-AE integrate multimodal longitudinal features with cross-sectional demographic data. To minimize the effect of the irregular time intervals between visits, PPAD and PPAD-AE use patient age in each visit as an indicator of time change between consecutive visits. In addition, PPAD and PPAD-AE utilize a customized loss function to give more weight on predicting conversion to AD cases, thereby increasing the model's sensitivity. To show robustness of our proposed models, we used two evaluation setups, by which (i) ADNI dataset was used to train and test the proposed models; (ii) ADNI dataset was used to train the proposed models and the National Alzheimer's Coordinating Center (NACC) dataset was used to test the proposed models. Our experimental results showed that our proposed models outperformed all baseline models for most of the prediction scenarios in terms of F2 and sensitivity. We also demonstrated that using age feature improved the model performance by helping address the irregular time interval between consecutive visits. We made PPAD and PPAD-AE publicly available at <https://github.com/bozdaglab/PPAD> under Creative Commons Attribution Non-Commercial 4.0 International Public License.

2 Materials and methods

2.1 Datasets

In this study, longitudinal and cross-sectional data from two datasets were used. The main dataset was from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild MCI and early AD. Since it has been launched, the public-private cooperation has contributed to significant achievements in AD research by sharing data to researchers from all around the world (Jack et al. 2010; Jagust et al. 2010; Risacher et al. 2010; Saykin et al. 2010; Trojanowski et al. 2010; Weiner et al. 2010, 2013).

The second dataset was from the NACC (Besser et al. 2018) database, which is a large, centralized resource for AD research. It contains data from multiple study sites across the United States, including demographic, cognitive, genetic, and

MRI data. The database is designed to facilitate research on the causes, diagnosis, and treatment of AD.

In this study, we have used two different experimental setups to train and evaluate our proposed models. In the first setup, ADNI dataset was split into training and test datasets to train and test the proposed models. In the second setup, the whole ADNI dataset was used to train the proposed models while the NACC dataset was used as an external dataset to test the proposed models.

2.1.1 ADNI dataset for training and testing the proposed models

To obtain longitudinal and cross-sectional data from all ADNI studies (ADNI1, ADNI2, and ADNI-GO), ADNImerge R package was used (<https://adni.bitbucket.io/>) (Jiang et al. 2020; McCombe et al. 2020). The raw data consisted of 15 087 records from 2288 unique patients. Each record represents a patient visit that consists of feature values from four data modalities namely cognitive performance measurement, MRI, CSF, and demographic, and the diagnosis label (i.e. CN, MCI, or AD). The dataset had several irregularities: the patients had varying number of visits ranging from 1 to 21; the time between consecutive visits for a patient varied from 3 to 60 months; and several visits had missing feature values. To address these issues, we preprocessed the dataset by performing several steps (Fig. 1). K nearest neighbor algorithm was employed to impute the missing values. Each missing feature was imputed using the average of values from the nearest neighbors that had a value for that feature. When finding the nearest neighbors, only the records with the same diagnosis (i.e. MCI or AD) were considered. The Euclidean distance metric was used and the number of neighbors, k , was set to five.

In the preprocessing stage, there was a trade-off between the number of patients and the quality of data. As a result, we were keen to reduce data imputation as much as possible through removing visits and features that had high missing rate based on carefully determined thresholds (40% and 60% for visits and features, respectively). Reducing data imputation—high data quality—was at the expense of the number of patients.

After the imputation, the final dataset had 20 longitudinal and 5 demographic features for 1169 patients and 5759 visits. To train the proposed models, the dataset was randomly split into 70% train and 30% test data in a stratified manner. After the splitting, each feature was normalized using min-max normalization. For generalization, the whole procedure was repeated for three random splits.

2.1.2 ADNI dataset for training and NACC dataset for testing the proposed models

In this setup, we performed a data harmonization to collect the overlapping features from ADNI and NACC datasets. To do so, first, we obtained three different data modalities from NACC dataset namely cognitive, genetic, and MRI data. Demographic information including gender, race, ethnicity, and education were in the cognitive data. We obtained APOE ϵ 4 status from genetic data for all the patients in the cognitive data. We could not integrate MRI data features with the cognitive data due to the low number of overlapping patients, mismatching number and date of visits for the overlapping patients, and high rate of missing values (>90%) of the overlapping features between ADNI and NACC datasets. At the end, we harmonized nine features between ADNI and NACC datasets (Supplementary Table S2). ADNI had 1205 patients

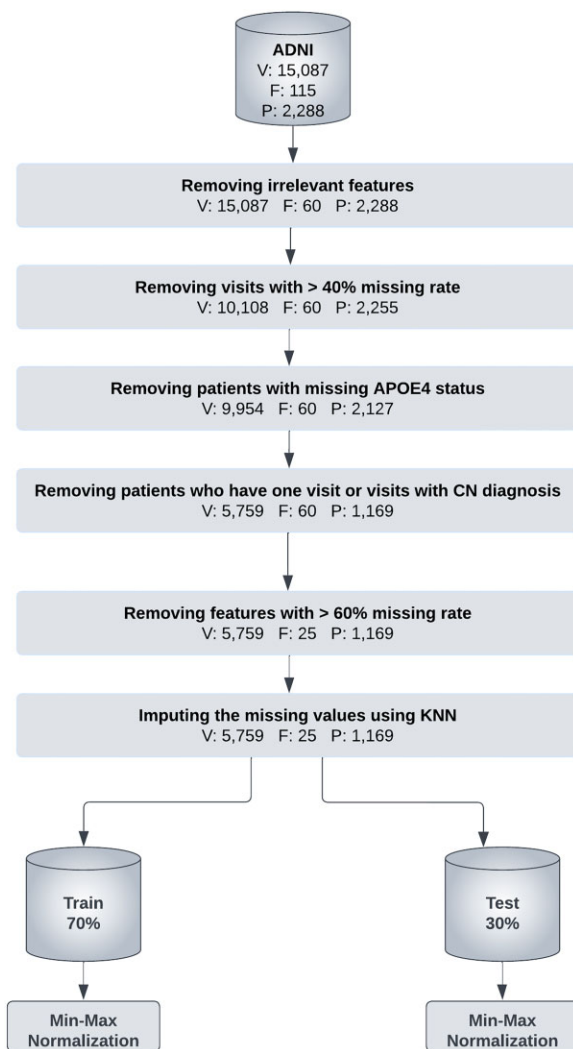


Figure 1. ADNI data preprocessing steps. V, F, and P represents the number of visits, features, and patients, respectively.

and 6066 visits while NACC had 8121 patients and 35 423 visits. The same preprocessing steps discussed in the previous section were applied to prepare ADNI and NACC datasets.

2.1.3 Dataset notations

After preprocessing, we prepared the datasets as a multivariate longitudinal data. Let M denote a dataset with N samples (patients), $M = (X_1, \dots, X_N)$ where each sample X represents measurements of F features collected over T time points (visits): $X = \{x_1, x_2, \dots, x_T\} \in \mathbb{R}^{T \times F}$. For each visit $t = 1, 2, \dots, T$, $x_t = \{x_t^1, x_t^2, \dots, x_t^F\} \in \mathbb{R}^F$ represents a vector of features of sample X at visit t . For each feature $f = 1, 2, \dots, F$, $x^f = \{x_1^f, x_2^f, \dots, x_T^f\} \in \mathbb{R}^T$ represents the f th feature value of sample X over T visits, and x_t^f represents the f th feature value of sample X at visit t . In M , each sample has a corresponding diagnosis (Y_1, \dots, Y_N) for each time point: $Y = \{y_1, y_2, \dots, y_T\} \in \mathbb{R}^{T \times 1}$. For each visit $t \in \{1, 2, \dots, T\}$, $y_t \in \{0, 1\}$ where 0 denotes MCI and 1 denotes AD.

2.2 Method

In this study, we built two RNN-based deep learning models to predict conversion from MCI to AD at the next visit and the multiple visits ahead. In both models, we utilized the age feature to alleviate the limitation of RNN models with

irregular time intervals between consecutive inputs in the sequence. Our proposed models also utilized a customized binary cross entropy loss function to give a higher weight to its sensitivity since early prediction of conversion cases correctly is more important than making a false positive prediction. In the proposed models, the type RNN cell was a hyperparameter with possible choices of LSTM, GRU, bidirectional LSTM (Bi-LSTM), and bidirectional GRU (Bi-GRU).

2.2.1 A primer on GRU and Bi-GRU

In GRU, each cell consists of reset (Equation (1)) and update gate (Equation (2)) through which the cell determines what portion of the previous cell state and the current input will be utilized. To compute the hidden state at time point t , h_t , first a candidate hidden state, \tilde{h}_t is computed (Equation (3)) by utilizing the current input vector x_t , the reset gate r_t , and the hidden state at the previous time point, h_{t-1} . Then, utilizing the update gate z_t , the current hidden state is computed as a weighted average of h_{t-1} and \tilde{h}_t (Equation (4)). In Equations (1)–(4), W_r , U_r , W_z , U_z , W_b , and U_b are the trainable linear transformation matrices; b_r , b_z , and b_b are the bias vectors; σ is the sigmoid function, and \odot is the Hadamard product.

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (1)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$\tilde{h}_t = \tanh(W_b x_t + U_b(r_t \odot h_{t-1}) + b_b) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

In Bi-GRU, two unidirectional GRUs are used to learn information from the previous and the later inputs in the sequence while processing the current input (Liu et al. 2021). The first GRU is a forward GRU (GRU_f), which is explained in the previous paragraph. The second GRU is a backward GRU (GRU_b), which is exactly same as (GRU_f) except that the hidden state of the cell is computed based on the current and the later inputs. In other words, the hidden state of a backward GRU cell is calculated based on Equations (1)–(4) except that all h_{t-1} terms are replaced with h_{t+1} . To compute the hidden state of Bi-GRU at time point t , the hidden states of GRU_f and GRU_b are computed and concatenated (Equation (5)). In Equation (5), \oplus denotes the concatenation operation.

$$\text{Bi-GRU}(x_t) = GRU_f(x_t) \oplus GRU_b(x_t) \quad (5)$$

2.2.2 Prediction model for conversion to AD at the next visit

To predict AD conversion at the next visit, we developed a framework named Predicting Progression of Alzheimer's Disease (PPAD) that consists of a RNN and multilayer perceptron (MLP) (Fig. 2). In this architecture, the RNN component learns x_t , a latent representation of the longitudinal clinical data up to t visits (Equation (6)). Then, the MLP model is trained with concatenation of the cross-sectional demographic data (D) and x_t to predict conversion to AD at next visit (Equation (7)).

$$\hat{x}_t = \text{RNN}(X) \quad (6)$$

$$y' = \sigma(W_1(ReLU(W_2(x_t \oplus D) + b_2)) + b_1) \quad (7)$$

In Equation (7), y' represents the predicted diagnosis, W_1 and W_2 are the trainable linear transformation matrices, and b_1 and b_2 are the bias vectors.

2.2.3 Prediction model for conversion to AD at multiple visits ahead

For early predicting conversion of AD at multiple visits ahead, we propose another architecture PPAD-AE that composes of a RNN autoencoder and an MLP (Fig. 3). In this architecture, the RNN component learns a latent representation (x_t) of the longitudinal clinical data up to t visits (Equation (6)). Then, the latent representation is used by the decoder component to generate representations of multiple visits ahead up to n visits. Finally, the MLP model is trained with concatenation of the cross-sectional demographic data (D) and the representation of the last generated visit by the decoder to predict conversion to AD at the $(t + n)$ th visit (Equation (8)).

$$y' = \sigma(W_1(ReLU(W_2(x_{t+(n-1)} \oplus D) + b_2)) + b_1) \quad (8)$$

2.2.4 Parameter learning and evaluation metrics

To increase the prediction's sensitivity for both architectures, all trainable parameters for the RNN, RNN autoencoder, and the MLP were learned in an integral way using a customized binary cross-entropy loss function (Equation (9)) to give more weight on predicting conversion to AD cases. By using this customized loss function, we seek to minimize the false negative cases while predicting diagnosis of the future visit which leads to increased sensitivity of the predictive model.

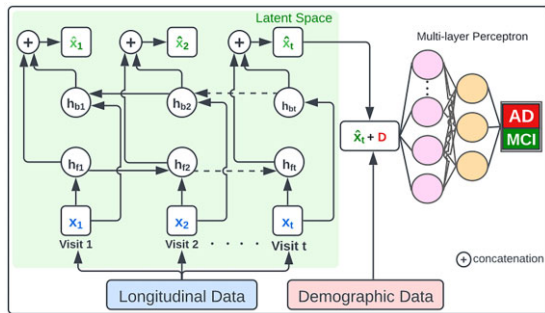


Figure 2. The architecture of PPAD to predict the conversion to AD at the next visit.

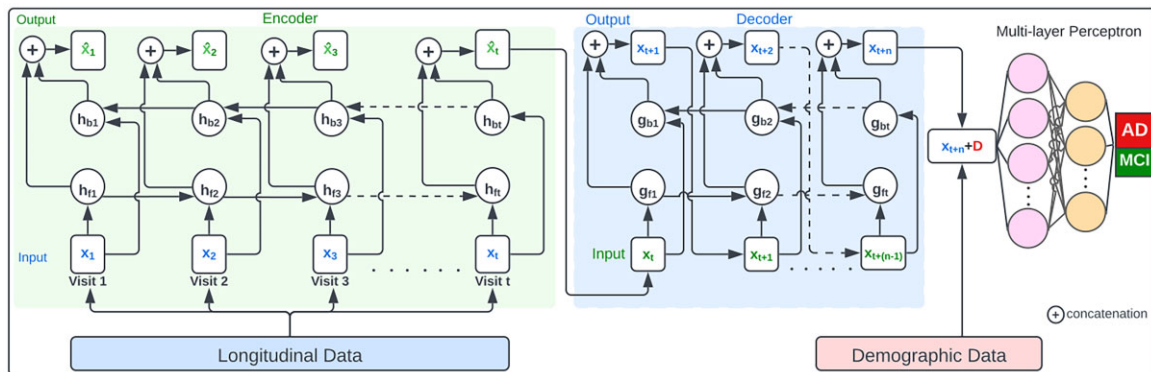


Figure 3. The architecture of PPAD-AE to predict the conversion to AD at the future visits.

$$\text{Loss} = -\frac{1}{N} \sum (\alpha \cdot (y \cdot \log y') + ((1 - \alpha) \cdot (1 - y) \cdot \log(1 - y'))) \quad (9)$$

In Equation (9), α is a real number between 0 and 1 to define the relative weight of positive prediction, y is the true diagnosis, and y' is the predicted diagnosis. In this study, we set α to 0.7. Based on the proposed customized loss function, all trainable parameters for the RNN, RNN autoencoder, and MLP (Equations (7) and (8)) were updated while training models in the backpropagation. For optimization, all models were trained using adaptive moment estimation (Adam) optimizer and the learning rate was set to 0.001. RNN cell, number of epochs, batch size, dropout rate, and L2 regularization are the hyperparameters that have been tuned. For model evaluation, F2 score (Equation (10)) and sensitivity were used.

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (10)$$

In Equation (10), recall is considered β times more important than precision. In this study, β was set to 2. We decided to use F2 score instead of other popular evaluation metrics such as AUC because in situations where there are wide disparities in the cost of false negatives versus false positives, it may be crucial to maximize one type of prediction error. In our case, increasing sensitivity (i.e. predicting patients who will convert to AD correctly) is more important than making false positive predictions (i.e. predicting someone will convert to AD incorrectly).

3 Results and discussion

In this study, we propose two RNN-based architectures, namely PPAD and PPAD-AE for the prediction of conversion to AD at the next visit and the multiple visits ahead, respectively. We evaluated the proposed architectures in two experimental setups. In the first setup, ADNI dataset was utilized to train and test the proposed architectures using the longitudinal multimodal and the cross-sectional demographic data. The longitudinal data consisted of 20 features from cognitive performance and neuroimaging biomarker data modalities (Supplementary Table S1). The cross-sectional demographic data consisted of gender, race, ethnicity, education, and APOEε4. We split the data into 70% training and 30% test three times and reported the average performance across these splits. In the second setup, the models were trained on the entire ADNI longitudinal and cross-sectional data and tested on

NACC data. The longitudinal data consisted of five features and the cross-sectional demographic data consisted of gender, race, ethnicity, education, and APOE ϵ 4 (Supplementary Table S2). For both experimental setups, we used age as a longitudinal feature to represent the time difference between consecutive visits. To select the optimal values for the hyperparameters (i.e. RNN cell, number of epochs, batch size, dropout rate, and L2 regularization), we performed grid search with 5-fold cross-validation for each investigated scenario in both experimental setups. Supplementary Tables S3 and S4 show the best hyperparameter values for PPAD and PPAD-AE for all splits in the first experimental setup, respectively. Supplementary Table S5 and S6 show the best hyperparameter values for PPAD and PPAD-AE in the second experimental setup, respectively. Supplementary Tables S7 and S8 show the number of converters and nonconverters in each scenario in the first and second experimental setup, respectively.

3.1 Predicting the conversion to AD at the next visit

To evaluate PPAD, which predicts conversion to AD at the next visit, we trained it using different scenarios for both experimental setups. Specifically, we trained four models using two, three, five, and six visits to predict the conversion to AD at the next visit (i.e. at the third, fourth, sixth, and seventh visits, respectively). For comparison, we trained a T-LSTM-based architecture using the same training data. Since T-LSTM can handle irregular intervals between visits internally, to check the effectiveness of utilizing the age feature, we did not use the age feature for T-LSTM. In addition, Random Forest (RF)- and SVM-based models were trained as baseline. Since RF and SVM cannot handle longitudinal data, they were trained using the same training data used for RNN models after aggregating each feature value by computing its mean. For a fair comparison, demographic features were not utilized in PPAD, RF, and SVM since T-LSTM does not have the ability to integrate these features. For generalization, the whole procedure was repeated 15 times. The results on both experimental setups showed that PPAD outperformed all baseline models in terms of F2 (Fig. 4A and B) and sensitivity (Supplementary Fig. S1A and B). The results also showed that, as expected, training using more visits improved the performance of next visit diagnosis prediction for all RNN models. In addition, PPAD outperformed T-LSTM in seven out of eight cases, which shows the ability of our model to address the irregular time intervals problem better than T-LSTM.

3.2 Predicting the conversion to AD at multiple visits ahead

To evaluate PPAD-AE, which predicts the conversion to AD at multiple visits ahead, we trained it using different scenarios. Specifically, we trained four models using two, three, five, and six visits to predict the diagnosis at the following second, third, and fourth visits ahead. For example, the model that was trained using two visits was evaluated on predicting the diagnosis at the fourth, fifth, and sixth visits. For generalization, the whole procedure was repeated 15 times. We compared PPAD-AE to RF and SVM. T-LSMT was unable to predict multiple visits ahead, thus was not used in this evaluation. We observed that PPAD-AE outperformed all baseline models in terms of F2 (Fig. 5) and sensitivity (Supplementary Fig. S2) in both experimental setups except for one scenario (Fig. 5F and Supplementary Fig. S2F). As observed in the PPAD results (Fig. 4), training the model with more visits improved the prediction performance in general, whereas the performance of most models dropped when predicting diagnosis at the farther time points. We also observed that the performance of PPAD-AE on NACC dataset was higher than held-out dataset of ADNI. This could be due to using a larger training set (i.e. the entire ADNI data) when testing the model on NACC cohort.

3.3 Feature importance analysis

We investigated the performance of the proposed models (Figs 2 and 3) to determine feature importance through SHapley Additive exPlanations (SHAP). Figure 6 and Supplementary Fig. S3 show the mean absolute SHAP value for the longitudinal features for the proposed models with first and second experimental setups, respectively. The results indicate that functional activities questionnaire and the logical memory delayed recall total (LDELTTOTAL) are the most important features to predict conversion to AD in all scenarios. In addition, age is among important features in predicting conversion to AD. For example, age was the seventh important feature among 20 features (Fig. 6A).

3.4 Ablation study for the demographic features

Through an ablation study, we investigated the performance of the proposed models (Figs 2 and 3) with and without integrating the demographic features. We observed that for both models, the performance did not change much due to the demographic features. In the first experimental setup, we

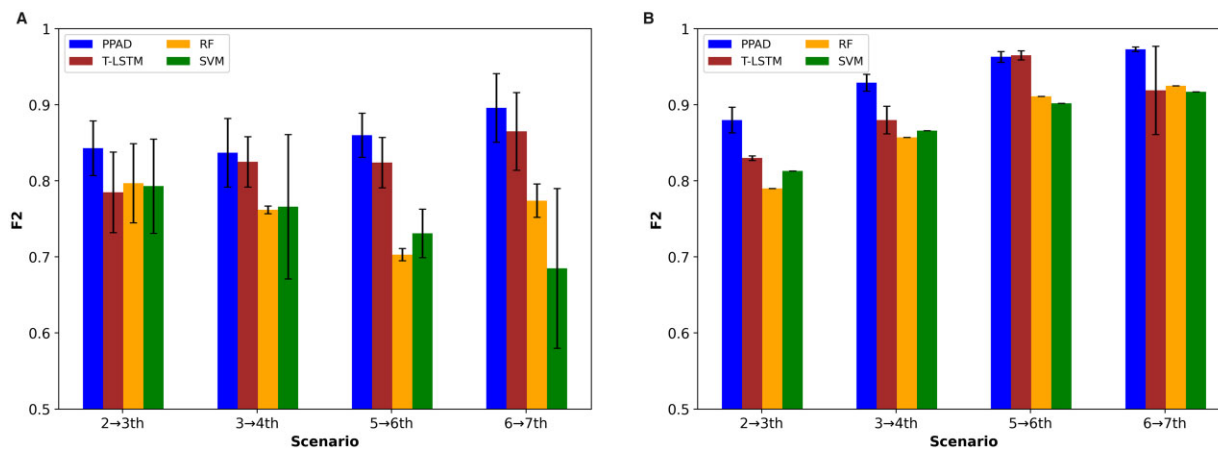


Figure 4. F2 scores for PPAD models to predict conversion to AD at the next visit. (A) Models tested on held-out samples in ADNI after training using two, three, five, and six visits in ADNI, respectively. (B) Models tested on NACC after training using two, three, five, and six visits in ADNI, respectively.

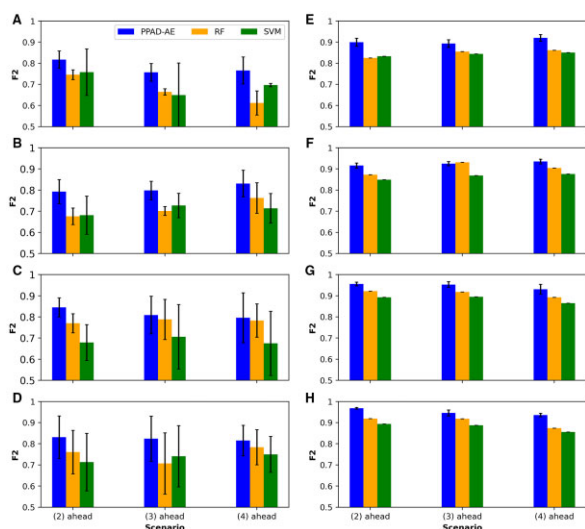


Figure 5. F2 scores for PPAD-AE models to predict conversion to AD at the second, third, and fourth visits ahead. (A–D) Models tested on held-out samples in ADNI after training using two, three, five, and six visits in ADNI, respectively. (E–H) Models tested on NACC after training using two, three, five, and six visits in ADNI, respectively.

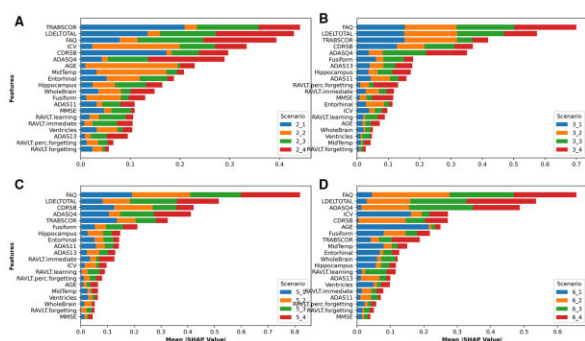


Figure 6. SHAP values for all features used in the first experimental setup using (A) two, (B) three, (C) five, and (D) six visits to train the models. 2_1 means trained using two visits to predict conversion to AD at the next visit, 3_2 means trained using three visits to predict conversion at two visits ahead, and so on.

observed a slight increase when excluding demographic features in PPAD (Fig. 7), whereas there was a slight increase when including them in PPAD-AE for 9 out of 12 cases (Fig. 8). In the second experimental setup, we observed a slight increase when excluding demographic features in PPAD for three out of four cases (Supplementary Fig. S4), whereas there was a slight increase when including them in PPAD-AE for 6 out of 12 cases (Supplementary Fig. S5).

4 Conclusion

In this study, we present two deep learning architectures PPAD and PPAD-AE to predict the conversion to AD in the future. We utilized longitudinal cognitive and neuroimaging features and cross-sectional demographic data from two large AD databases to evaluate our models. We conducted two experimental setups where ADNI data were used partially or completely to train the model and held-out data and NACC data were used to test the models. In both experimental setups, our tools outperformed other existing tools and baseline models. We also investigated other tools, but could not test

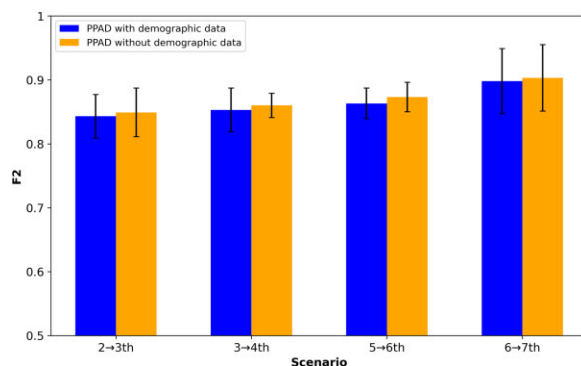


Figure 7. PPAD ablation results for the demographic features. PPAD tested on held-out samples in ADNI after training using two, three, five, and six visits in ADNI.

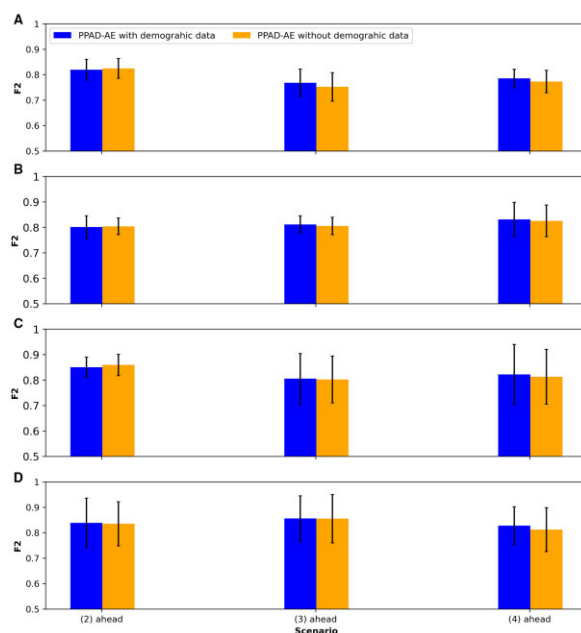


Figure 8. PPAD-AE ablation results for the demographic features. (A–D) PPAD-AE tested on held-out samples in ADNI after training using two, three, five, and six visits in ADNI, respectively.

them as the code or the tool was not made available. By utilizing a customized loss function, we gave higher emphasis on the sensitivity of the models. Because for AD prediction, a false positive prediction (predicting someone to convert to AD falsely) is less severe than a false negative prediction (not being able to predict a conversion case). PPAD and PPAD-AE are also flexible to incorporate additional features including omics features from gene expression, DNA methylation datasets, and blood-based biomarker measurements. To increase its usability, we make PPAD and PPAD-AE publicly available at <https://github.com/bozdaglab/PPAD/>.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI was funded by the National

Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The NACC database is funded by NIA/NIH [Grant U24 AG072122]. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank aFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health [Award Number R35GM133657] and the startup funds from the University of North Texas.

References

- Aisen PS, Petersen RC, Donohue MC *et al.*; Alzheimer's Disease Neuroimaging Initiative. Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans. *Alzheimers Dement* 2010;**6**:239–46.
- Alam S, Kwon GR, Kim JI *et al.* Twin SVM-based classification of Alzheimer's disease using complex dual-tree wavelet principal coefficients and LDA. *J Healthc Eng* 2017;**2017**:1.
- Aljović A, Badnjević A, Gurbeta L. Artificial neural networks in the discrimination of Alzheimer's disease using biomarkers data. In *2016 5th Mediterranean Conference on Embedded Computing (MECO)*, 286–9. IEEE, 2016.
- Al-Naami B, Gharaibeh N, Khesman AA. Automated detection of Alzheimer disease using region growing technique and artificial neural network. *World Acad Sci Eng Technol Int J Biomed Biol Eng* 2013;**7**(5):204–8.
- Alzheimer's Association. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 2015;**11**:332–84.
- Baytas IM, Xiao C, Zhang X *et al.* Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 65–74, 2017.
- Besser L, Kukull W, Knopman DS *et al.* Version 3 of the National Alzheimer's Coordinating Center's uniform data set. *Alzheimer Dis Assoc Disord* 2018;**32**:351–8.
- Bisgin H, Bera T, Ding H *et al.* Comparing SVM and ANN based machine learning methods for species identification of food contaminating beetles. *Sci Rep* 2018;**8**:1–12.
- Chalmers K, Wilcock GK, Love S. APOE ϵ 4 influences the pathological phenotype of Alzheimer's disease by favouring cerebrovascular over parenchymal accumulation of A β protein. *Neuropathol Appl Neurobiol* 2003;**29**:231–8.
- Cheng B, Liu M, Zhang D *et al.* Domain transfer learning for MCI conversion prediction. *IEEE Trans Biomed Eng* 2015;**62**:1805–17.
- Cho K, Van Merriënboer B, Bahdanau D *et al.* On the properties of neural machine translation: Encoder-decoder approaches. In: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation 2014 Oct* (pp. 103–111). 2014.
- Cui R, Liu M; Alzheimer's Disease Neuroimaging Initiative. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput Med Imaging Graph* 2019;**73**:1–10.
- Frost S, Martins RN, Kanagasangam Y. Ocular biomarkers for early detection of Alzheimer's disease. *J Alzheimers Dis* 2010;**22**:1–16.
- Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Comput* 2000;**12**:2451–71.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press, 2016: 305–307.
- Greff K, Srivastava RK, Koutník J *et al.* LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 2017;**28**:2222–32.
- Hall K, Gureje O, Gao S *et al.* Risk factors and Alzheimer's disease: a comparative study of two communities. *Aust N Z J Psychiatry* 1998;**32**:698–706.
- Hampel H, Prvulovic D, Teipel S, German Task Force on Alzheimer's Disease (GTF-AD) *et al.* The future of Alzheimer's disease: the next 10 years. *Prog Neurobiol* 2011;**95**:718–28.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;**9**:1735–80.
- Huang C, Yan B, Jiang H *et al.* Combining voxel-based morphometry with artificial neural network theory in the application research of diagnosing Alzheimer's disease. In *2008 International Conference on BioMedical Engineering and Informatics*, vol. 1, 250–254. IEEE, 2008.

- Huang L, Jin Y, Gao Y *et al.*; Alzheimer's Disease Neuroimaging Initiative. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based Random Forest. *Neurobiol Aging* 2016;46:180–91.
- Ikeda T, Yamada M. Risk factors for Alzheimer's disease. *Brain Nerve* 2010;62:679–90.
- Jack CR Jr, Bernstein MA, Borowski BJ *et al.*; Alzheimer's Disease Neuroimaging Initiative. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement* 2010;6:212–20.
- Jagust WJ, Bandy D, Chen K *et al.*; Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative positron emission tomography core. *Alzheimers Dement* 2010;6:221–9.
- Jessen F, Wolfsgruber S, Wiese B *et al.*; German Study on Aging, Cognition and Dementia in Primary Care Patients. AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimers Dement* 2014;10:76–83.
- Jiang L, Lin H, Chen Y *et al.*; Alzheimer's Disease Neuroimaging Initiative. Sex difference in the association of APOE4 with cerebral glucose metabolism in older adults reporting significant memory concern. *Neurosci Lett* 2020;722:134824.
- Joshi S, Shenoy D, Simha GV *et al.* Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. In *2010 Second International Conference on Machine Learning and Computing*, 218–222. IEEE, 2010.
- Kar S, Majumder DD. A novel approach of diffusion tensor visualization based neuro fuzzy classification system for early detection of Alzheimer's disease. *J Alzheimers Dis Rep* 2019;3:1–18.
- Kim D, Kim S, Risacher SL *et al.* A graph-based integration of multimodal brain imaging data for the detection of early mild cognitive impairment (E-MCI). In *International Workshop on Multimodal Brain Image Analysis*. Springer, Cham, 2013, 159–69.
- Lee G, Nho K, Kang B *et al.*; for Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using multimodal deep learning approach. *Sci Rep* 2019;9:1–12.
- Li H, Fan Y. Early prediction of Alzheimer's disease dementia based on baseline hippocampal MRI and 1-year follow-up cognitive measures using deep recurrent neural networks. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 368–71. IEEE, 2019.
- Liu X, Wang Y, Wang X *et al.* Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system. *Opt Express* 2021;29:5923–33.
- Luong DTA, Chandola V. An empirical evaluation of time-aware LSTM autoencoder on chronic kidney disease. *CoRR*, 2018.
- Martí-Juan G, Sanroma-Guell G, Piella G. A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease. *Comput Methods Programs Biomed* 2020;189:105348.
- McCombe N, Ding X, Prasad G *et al.* Predicting feature imputability in the absence of ground truth. In: *37th International Conference on Machine Learning (ICML): The Art of Learning with Missing Values (ARTEMIS) Workshop 2020 Jul 2*. 2020.
- Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy* 2011;4:47–55.
- Nguyen M, He T, An L *et al.*; Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using deep recurrent neural networks. *Neuroimage* 2020;222:117203.
- Patterson C. *World Alzheimer Report 2018*. London: Alzheimer's Disease International. 2018.
- Perrin RJ, Fagan AM, Holtzman DM. Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* 2009;461:916–22.
- Petersen RC. Mild cognitive impairment. *Continuum* 2016;22:404–18.
- Petersen RC, Caracciolo B, Brayne C *et al.* Mild cognitive impairment: a concept in evolution. *J Intern Med* 2014;275:214–28.
- Plocharski M, Østergaard LR; Alzheimer's Disease Neuroimaging Initiative. Extraction of sulcal medial surface and classification of Alzheimer's disease using sulcal features. *Comput Methods Progr Biomed* 2016;133:35–44.
- Quintana M, Guàrdia J, Sánchez-Benavides G *et al.*; Neuronorma Study Team. Using artificial neural networks in clinical neuropsychology: high performance in mild cognitive impairment and Alzheimer's disease. *J Clin Exp Neuropsychol* 2012;34:195–208.
- Retico A, Bosco P, Cerello P *et al.* Predictive models based on support vector machines: whole-brain versus regional analysis of structural MRI in the Alzheimer's disease. *J Neuroimaging* 2015;25:552–63.
- Risacher SL, Shen L, West JD *et al.*; Alzheimer's Disease Neuroimaging Initiative (ADNI). Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiol Aging* 2010;31:1401–18.
- Rodríguez JJ, Verkhratsky A. Neurogenesis in Alzheimer's disease. *J Anat* 2011;219:78–89.
- Saykin AJ, Shen L, Foroud TM *et al.*; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers Dement* 2010;6:265–73.
- Scheltens P, De Strooper B, Kivipelto M *et al.* Alzheimer's disease. *Lancet* 2021;397:1577–90.
- Sheng J, Wang B, Zhang Q *et al.* A novel joint HCPMMP method for automatically classifying Alzheimer's and different stage MCI patients. *Behav Brain Res* 2019;365:210–21.
- Sun Z, Qiao Y, Lelieveldt BP *et al.*; Alzheimer's Disease Neuroimaging Initiative. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. *Neuroimage* 2018;178:445–60.
- Tábuas-Pereira M, Baldeiras I, Duro D *et al.* Prognosis of early-onset vs. late-onset mild cognitive impairment: comparison of conversion rates and its predictors. *Geriatrics* 2016;1:11.
- Tangaro S, Fanizzi A, Amoroso N *et al.*; Alzheimer's Disease Neuroimaging Initiative. A fuzzy-based system reveals Alzheimer's disease onset in subjects with mild cognitive impairment. *Phys Med* 2017;38:36–44.
- Tanveer M, Richhariya B, Khan RU *et al.* Machine learning techniques for the diagnosis of Alzheimer's disease: a review. *ACM Trans Multimedia Comput Commun Appl* 2020;16:1–35.
- Tong H, Lou K, Wang W. Near-infrared fluorescent probes for imaging of amyloid plaques in Alzheimer's disease. *Acta Pharm Sin B* 2015;5:25–33.
- Trojanowski JQ, Vandevertichele H, Korecka M *et al.*; Alzheimer's Disease Neuroimaging Initiative. Update on the biomarker core of the Alzheimer's disease neuroimaging initiative subjects. *Alzheimers Dement* 2010;6:230–8.
- Venugopalan J, Tong L, Hassanzadeh HR *et al.* Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep* 2021;11:1–13.
- Wang S, Zhang Y, Dong Z *et al.* Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int J Imaging Syst Technol* 2015;25:153–64.
- Weiner MW, Aisen PS, Jack CR, Jr *et al.*; Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement* 2010;6:202–11.e7.
- Weiner MW, Veitch DP, Aisen PS *et al.*; Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dementia* 2013;9:e111–94.
- Winblad B, Palmer K, Kivipelto M *et al.* Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* 2004;256:240–6.
- Wong W. Economic burden of Alzheimer disease and managed care considerations. *Am J Manage Care* 2020;26(8 Suppl):S177–83.
- Yadav P, Steinbach M, Kumar V *et al.* Mining electronic health records (EHRs) A. *ACM Comput Surv* 2018;50:1–40.
- Zeng N, Qiu H, Wang Z *et al.* A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing* 2018;320:195–202.
- Zhang D, Shen D; Alzheimer's Disease Neuroimaging Initiative. Multimodal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* 2012;59:895–907.