2020 International Conference on Computational Science and Computational Intelligence (CSCI) | 978-1-7281-7624-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/CSCI51800.2020.00144

A Deep-Learning Approach for the Prediction of Mini-Mental State Examination Scores in a Multimodal Longitudinal Study

Ulyana Morar
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
uwill002@fiu.edu

Harold Martin
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
hmart027@fiu.edu

Walter Izquierdo
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
wizqu003@fiu.edu

Parisa Forouzannezhad
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
pforo003@fiu.edu

Elaheh Zarafshan
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
ezara007@fiu.edu

Rosie E. Curiel
Dept Psychiatry & Behavioral Sciences
University of Miami
Miami, FL, USA
rcuriel2@med.miami.edu

Monica Roselli
Dept of Psychology
Florida Atlantic Univ.
Boca Raton, FL, USA
mrossell@fau.edu

David Loewenstein
Dept Psychiatry & Behavioral Sciences
University of Miami
Miami, FL, USA
dloewenstein@med.miami.edu

Ranjan Duara
Wien Ctr. Alzheimers Dis & Memory Disorders
Mt Sinai Med Center
Miami, FL, USA
ranjan.duara@msmc.com

Elona Unger
College of Pharmacy
Florida A&M Univ.
Tallahassee, FL, USA
ilonaunger@gmail.com

Malek Adjouadi
Department of Electrical and Computer Engineering
Florida International Univ.
Miami, FL, USA
adjouadi@fiu.edu

***Abstract*— This study introduces a new multimodal deep regression method to predict cognitive test score in a 5-year longitudinal study on Alzheimer's disease (AD). The proposed model takes advantage of multimodal data that includes cerebrospinal fluid (CSF) levels of tau and beta-amyloid, structural measures from magnetic resonance imaging (MRI), functional and metabolic measures from positron emission tomography (PET), and cognitive scores from neuropsychological tests (Cog), all with the aim of achieving highly accurate predictions of future Mini-Mental State Examination (MMSE) test scores up to five years after baseline biomarker collection. A novel data augmentation technique is leveraged to increase the numbers of training samples without relying on synthetic data. With the proposed method, the best and most encompassing regressor is shown to achieve better than state-of-the-art correlations of 85.07%(SD=1.59) for 6 months in the future, 87.39% (SD =1.48) for 12 months, 84.78% (SD=2.66) for 18 months, 85.13% (SD=2.19) for 24 months, 81.15% (SD=5.48) for 30 months, 81.17% (SD=4.44) for 36 months, 79.25% (SD=5.85) for 42 months, 78.98% (SD=5.79) for 48 months, 78.93%(SD=5.76) for 54 months, and 74.96% (SD=7.54) for 60 months.**

Keywords*— Longitudinal analysis, Alzheimer's disease, Cerebrospinal fluid, Alzheimer's Disease Neuroimaging Initiative (ADNI), magnetic resonance imaging (MRI)

I. INTRODUCTION

Alzheimer's is a neurodegenerative disease characterized by rapid decline in cognitive performance affecting approximately 5.5 million people in the United States [1]. Not only is the disorder a leading cause of death, but it typically involves long-term morbidity that largely affects individual's quality of life. Manifestations of AD are often expressed in brain atrophy with concurrent amyloid plaques through the toxic buildup of Amyloid-β (αβ) peptides and neural injury, resulting in progressive memory decline [2-11].

This area of study has gathered the most attention by the scientific community, and especially in exploring aplications of machine learning, where numerous studies have shown great prospects for augmenting our understanding of this complex disease [12-17]. In [12], the authors use Support Vector Machines (SVMs) to predict the Mini-Mental State Examination (MMSE) score of patients 24 months after the initial visit, and to further predict which patient will have an MMSE decline of more than three (3) points, which they deemed to be medically significant. They generate said predictions by only using demographic data, genetic biomarkers, neuro psychological tests, and baseline MMSE scores.

MRIs are used in [13] and [14]. The authors of [13], attempt to predict the clinical scores of patients within three months of their MRIs by using Relevance Vector Regression (RVC) through a Bayesian inference framework . They report that the predicted MMSE, CDR, and ADAS test scores are highly

This research is supported by the National Science Foundation (NSF) under grants CNS-1920182, CNS-1532061, CNS-1338922, CNS- 2018611 and CNS-1551221; and by the National Institutes of Health (NIH) under grants NIA/NIH 1R01 AG061106-02, NIA/NIH 1RO1AG04764905, and the NIA/NIH P30AG066506 with the Florida Alzheimer's Disease Research Center (Florida ADRC).

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00144

correlated within one another, which is to be expected as they all measure different aspects of cognitive performance. Whereas [14] uses two key MRI biomarkers to predict the MMSE at baseline and ADAS clinical scores at baseline and at the year's mark (12 months after). They instead employ an anatomically partitioned artificial neural network (APANN) composed of multiple hidden layers to encode the latent features of the input data.

Other studies such as [15-17] use multiple biomarker modalities to estimate the changes or future MMSE scores at different timepoints. The method suggested in [15] utilizes MRI, CSF, and FDG-PET to attain a classification of the subject under consideration along with their two-year (24 months) prognosis for MMSE and ADAS-Cog. They present a multimodal-multitask regressor-classifier in an attempt to simultaneously learn multiple tasks from multimodal data while taking advantage of the existing collinearities of the different modalities. While [16] expands the number of features used to include certain cognitive test scores to generate a running MMSE prediction from baseline to 48 months at 6-month intervals. In their paper, a Distributed Multitask Multimodal (DMM) approach is suggested to tackle the hard problem of learning useful features from distinct but correlated data. However, they ignore valuable cognitive data (MMSE, CDR, ADAS13) at baseline that would be relevant when trying to predict the progression of such scores at future time point. Finally, [17] uses gradient boosting of decision trees to generate MMSE predictions for the 24th month after baseline by using multimodal data from the previous 18th months.

Unlike previous approaches, we propose to make use of all available data at any given time point, through the use of a novel data augmentation technique, to predict the MMSE progression for an entire 5-year span, starting six months after the relevant features are measured. Our approach uses a simple and tested multilayer perceptron neural network with a newly-designed architecture. Therein generating a multitask single-shot regressor that can effectively tackle this difficult problem in a simple but yet most effective manner.

II. Methods

A. Study participants and Data description

In this work, all individuals were participating in the Alzheimer's Diseases Neuroimaging Initiative (ADNI) study which aims to obtain and maintain MRI, PET, CSF, biochemical biomarkers, and neuropsychological tests for the early detection and study of the progression of Alzheimer's disease (http://adni.loni.usc.edu). We include 1845 ADNI participants between the ages of 55 and 90, from whom 487 are cognitively normal (CN), 71 converted from a cognitively normal state to mild impairment (CNc), 497 with mild cognitive impairment (MCI), 337 converted from mild cognitive impairment state to AD (MCIc), 331 subjects diagnosed with Alzheimer's disease (AD), and 26 individuals who converted from CN to AD. The differentiation of clinical diagnosis depends on MMSE and Clinical Dementia Rating (CDR) cutoffs from ADNI protocols, among others. Due to the inherent variability of some subjects' performance on the MMSE test at different examinations, it is possible to have backwards reversal in clinical groups (i.e. from MCI to CN). We also include those subjects as CN unstable (78) and MCI unstable (18) in our study to explore the complexity of the ADNI dataset without artificially manipulating its distribution, as cleaning the data of "outlier" could possibly result in better performance metric for the proposed model than those attained from realistic subsets of the population at large that would inherently contain these types of samples.

The proposed longitudinal multimodal deep learning model uses four main modalities from the ADNI dataset:

- Neuroimaging measurements from MRIs: ventricular volume, hippocampus volume, whole brain volume, entorhinal cortical thickness, fusiform gyrus, middle temporal gyrus. Where the volumetric measurements are normalized by the intracranial volume (ICV).
- CSF measurements: amyloid-β 1–42 peptide (Aβ1-42), total tau (tTau), and tau phosphorylated at the threonine 181 (pTau).
- PET measurements: FDG (18-Fluoro-DeoxyGlucose), Pittsburgh Compound-B (PIB), and AV45.
- Cognitive scores (Cog): Rey Auditory Verbal Learning Tests, Functional Activities Questionnaires, Everyday Cognition (Ecog) scales, ADAS13, CDR, and MMSE.

One of the main challenges in longitudinal studies is panel attrition which requires additional preprocessing steps for the handling of missing data. In our study, we include subjects who have at least two separate visits, as we need at least one more datum after initial admission in order to validate our predictions. Fig. 1 depicts the percentage of available samples for neuroimaging, cognitive, CSF, and PET modalities for the different visits at Δt=6 months intervals. Understandably, the total available CSF samples is the smallest of all modalities for month 30 (0.1%), month 42 (0.2%), and month 54 (1%). While MRI samples are the most available data points over the duration of the study.

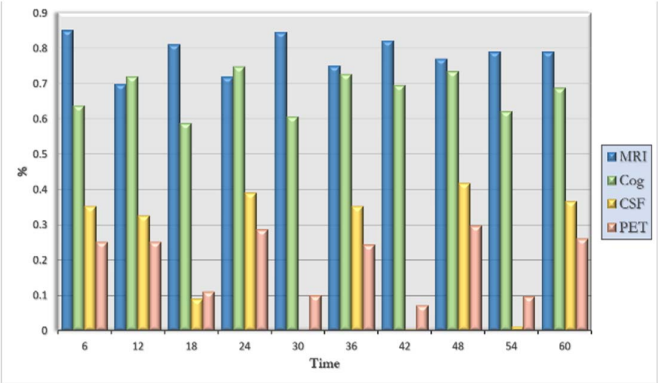

Figure 1. Percent of samples with available data for MRI, CSF, PET, and cognitive tests biomarkers for the different time points.

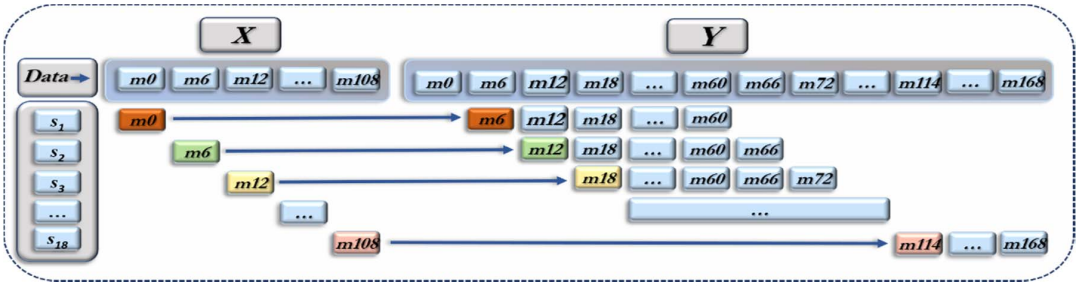

Figure 2. Dataset Augmentation Example

B. Data Preparation

Although the ADNI cohort is quite extensive, deep-learning models like the one presented in this study require large amounts of data, but as this study demonstrates it is still possible to achieve better performance even with a somewhat limited number of samples. ADNI usually follows patients for an extended period of time lasting for at most 168 months at the time of this study. However, the current literary work aimed at longitudinal prognosis tends to only use the biomarkers gathered at the baseline visit to predict trends in MMSE [12-16]. This practice effectively ignores valuable data collected from further visits that could be used to enhance training and provide greater variability for generalization. Therefore, to address this issue, we introduce a dataset augmentation technique that will make use of the data collected from all available follow-up visits. Herein producing a richer dataset to be used for testing and training. This augmentation technique is described in general terms by (1) and an example of its usage is displayed in Fig. 2.

$$\cup_{i=1}^{18}[X_{i\Delta t}]; [Y_{(i+1)\Delta t}, Y_{(i+2)\Delta t}, \dots, Y_{(i+n)\Delta t}] \quad (1)$$

$$\Delta t = 6, n = 10$$

Where X_i stands for the input features and Y_i stands for the target or predicted values at the i^{th} visit. In our particular case, we link a set of features X at any given time with a corresponding set of targets Y at times ranging from six (6) to sixty (60) months after X is collected.

By following (1), we generate the set of all possible $[X_i] \rightarrow [Y_{i+\Delta t}, Y_{i+2\Delta t}, \dots, Y_{i+10\Delta t}]$ for every single available follow-up visit from month 0 (baseline) to month 108 of every single available patient. Thereby generating a sample rich dataset from which to train and test the proposed network. Fig. 2 shows in practical details how data from a single subject can generate 18 different samples when using the method proposed herein.

Once the dataset has been augmented using (1), we are left with 8066 distinct samples from the original 1845 patients. Thereafter, we proceed to split these samples into training and testing sets by following a 90-10 split (90% train and 10% test). We repeat this process 10 times in accordance with the 10-fold cross-validation technique to generate 10 distinct and non-overlapping testing sets along with their associated training sets. All results presented through this paper display the mean and standard deviation of the reported metrics on the different testing sets. It is especially relevant to point out that, although the data split is random, special care has been taken to ensure that no data from subjects seen during the training of the model is seen in testing phase. That is to say, that if any given sample of a particular patient is used for training, no other samples from that same individual can be used for testing.

C. Network Architecture

A deep fully connected neural network is trained as a regressor to predict MMSE progression at multiple future time points. The prediction window of interest spans from six (6) months through the full duration of sixty (60) months. The model itself uses four biomarker modalities (MRI, PET, CSF, and Cognitive test scores) with a total of thirty-three (33) biomarkers to produce MMSE scores for ten distinct future time points at six-month intervals.

The network's architecture is shown in Fig. 3. A five (5) layer network is used with four (4) hidden layers. The first layer is composed of 50 neurons, the second and third consist of 100 layers, the fourth has 50, and the last has 10 (one for each of the predicted timepoints). Layers one through five have Rectified Linear Units (ReLU) activations and employ L2 regularization [18] to help the network weights remain small and avoid weight saturation. Weights are randomly initialized using the approach described by Golorot in [19].

Stochastic gradient descent is used for training the model with an RMSprop optimizer [20], learning rate of 0.01, and a 0.8 weight decay. Early stopping is employed to prevent the network from excessively overfitting the training data and prolong the training for an unnecessarily long time after no new features are being learned. Therefore, we stop learning and revert back to the best set of weight after 100 epochs (run through the training set) of no accuracy improvement when evaluating the testing dataset.

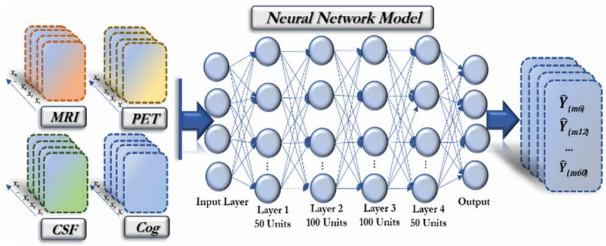

Figure 3. Proposed Neural Network Architecture

Furthermore, as not all timesteps are available for every single sample (Y matrix has missing values), we employ a masking technique to prevent missing values in the Y matrix from influencing the learning process. This masking algorithm effectively nullifies any gradient contribution from only the

output neurons that have missing data during any particular sample and only for that instance.

This missing data problem is also present for the X input matrix, as previously shown in Fig. 1. In this case we simply replace the missing values with the mean of that feature for the training set. This allows our model to generate predictions even for cases where data is missing from the input features. And by doing so we are able to learn valuable information from samples that would be otherwise discarded.

III. Results

We use ten-fold cross validation to avoid reporting on any particularly beneficial or detrimental dataset split. We train the network on a 64bit Windows 10 machine with an AMD FX-8350 Eight-Core Processor, 16 GB of DDR3 RAM, and an NVIDIA GeForce RTX2070 graphics card. The network is deployed in a Python 3.7 environment, using Keras 2.3.1, and TensorFlow GPU 1.14.0.

Performance was measured using three standard metrics widely used for regression problems, Correlation, Root-Mean-Squared Error (RMSE), and the Coefficient of Determination (R^2).

In Table 1 we report these metrics for each individually predicted timepoint along with their respective standard deviations for the 10 folds cross validation. Fig. 4 displays the scatter plots of predicted versus actual Mini-Mental State Examination scores for the ten different periods (from 6 to 60 months after the feature measurement) and across subgroups.

It is worth noting that there is a significant correlation spike at the twelfth month mark that would not be expected, as the correlation tends to go down over time due to the growing uncertainty of initial measurements. However, this fluctuation can be partially explained by an increased number of subjects present at the twelfth month (M12 in Fig. 4), especially in the low end of the MMSE scores. This added data helps the algorithm better fit the trend line and yields higher correlation, RMSE, and R^2 values. Other such spikes, although less significant, can also be observed at M24, M48, and to a very lesser extent M36, where there is a marked increase in the number of available datapoints that stretches to the lower ends of the MMSE scale.

Table 2 has an in-depth comparison between the current state-of-the-art MMSE regression algorithms and the proposed method. We provide a detailed overview of the approaches, datasets, and predicted correlations, along with standard deviations, for different timepoints.

It is evident from this table that the proposed model outperforms the competition across time, except for the sixth (6) month, where [17] outperforms. This is most likely due to the fact that [17] uses input features that stretch across multiple timepoints to generate a prediction six months after the last measurement. Therefore, taking advantage of the timeseries correlation to better approximate the progression rate of the MMSE.

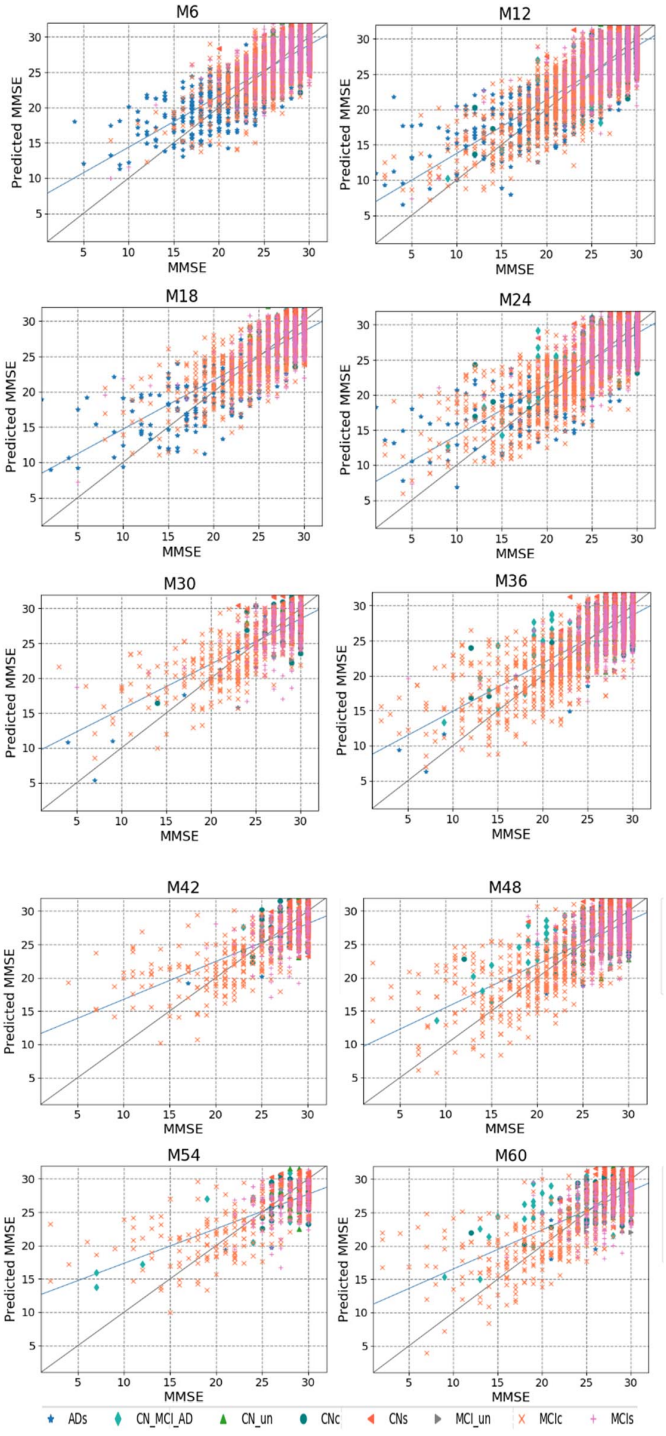

Figure 4. Scatter Plots of Predicted vs Actual MMSE for the different subgroups

TABLE II. EXPERIMENTAL RESULTS

Time(months):	**6**	**12**	**18**	**24**	**30**	**36**	**42**	**48**	**54**	**60**
Correlation ,% (SD)	85.3 (1.5)	87.6 (1.2)	85.3 (2.6)	85.6 (1.8)	81.9 (4.7)	81.7 (4.8%)	79.4 (4.8)	79.9 (5.2)	78.4 (6.5)	76.3 (7.3)
RMSE (SD)	2.42 (0.30)	2.52 (0.28)	2.76 (0.36)	2.71 (0.32)	3.00 (0.43)	2.87 (0.38)	3.11 (0.47)	3.05 (0.31)	3.64 (0.40)	3.23 (0.39)
R^2 , % (SD)	56.01 (12.01)	65.06 (7.50)	60.66 (8.75)	61.71 (7.75)	54.12 (12.68)	54.85 (14.41)	47.98 (18.33)	48.25 (18.16)	43.05 (15.92)	43.43 (22.49)

TABLE I. PREDICTION CORRILATION COMPARISON ACROSS MULTIPLE METHODS , % (SD)

Method	**Subjects (Samples)**	**Approach**	**Modalities**	**Time (month)**									
				6	12	18	24	30	36	42	48	54	60
[15]	167 (167)	M3T	MRI, FDG-PET, CSF	n/a	n/a	n/a	51.1 (2.1)	n/a	n/a	n/a	n/a	n/a	n/a
[16]	1620 (1620)	DMM	MRI, PET, CSF, Cog, Demog, APOE4	85.8	79.8	n/a	81.2	n/a	79.0	n/a	75.9	n/a	n/a
[17]	1141 (1141)	GBDT	MRI, PET, CSF, Cog, Demog, APOE4	**90.4**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Proposed	1845 (8066)	ANN	MRI, PET, CSF, Cog	85.7 (1.6)	**87.4 (1.5)**	**84.8 (2.7)**	**85.1 (2.2)**	**81.2 (5.5)**	**81.1 (4.4)**	**79.3 (5.9)**	**78.9 (5.8)**	**78.9 (5.8)**	**75.0 (7.5)**

IV. CONCLUSION

The proposed model, along with its accompanying data augmentation technique achieve better than state-of-the-art testing metrics for predicting Mini-Mental State Examination scores up to five years (sixty months) after the examination date. Our results further highlight the power of combining multiple biomarker modalities, as had been demonstrated by previous studies.

However, more investigation and added scrutiny remain to be performed pertaining the design of such network architectures and the ability to combine them with augmentation techniques such as gradient boosting in an effort to attain even higher prediction accuracy than those reported in this new study. Although the prediction results are nearly 85% or higher in correlation for up to 24 months and exceed 80% correlation for time points as far as 30 months from baseline, more investigation is needed in terms of determining what more could be done to overcome the missing data challenge and what other means could rigorously address the co-linearity issue inherent to longitudinal studies.

REFERENCES

[1] "Alzheimer's Disease Facts and Figures. SPECIAL REPORT On the Front Lines: Primary Care Physicians and Alzheimer's Care in America," Alzheimer's Association, [Online]. Available at website: https://www.alz.org/media/Documents/alzheimers-facts-and-figures_1.pdf.

[2] J. Becker and et al., "Amyloid-β associated cortical thinning in clinically normal elderly," Ann Neurol, vol. 69, no. 6, pp. 1032-1042, 2011.

[3] K. Blennow, B. Dubois, A. Fagan, P. Lewczuk, M. de Leon and H. Hampel, "Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease," Alzheimer's & dementia : the journal of the Alzheimer's Association, vol. 11, no. 1, pp. 58-69, 2015.

[4] C. Sutphen, L. McCue, E. Herries, C. Xiong and et al., "Longitudinal decreases in multiple cerebrospinal fluid biomarkers of neuronal injury in symptomatic late onset Alzheimer's disease," Alzheimer's & dementia : the journal of the Alzheimer's Association, vol. 14, no. 7, pp. 869-879, 2018.

[5] A. Fagan, C. Xiong and et al., "Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease," Science translational medicine, vol. 6, no. 226, p. 226ra30, 2014.

[6] L. Fan, K. Tzen and et al, "The Relation Between Brain Amyloid Deposition, Cortical Atrophy, and Plasma Biomarkers in Amnesic Mild Cognitive Impairment and Alzheimer's Disease," Front Aging Neurosci., vol. 10, no. 175, 2018.

[7] K. Bateman, R. Xiong, T. Benzinger and et al, "Clinical and biomarker changes in dominantly inherited Alzheimer's disease," The New England journal of medicine, vol. 367, no. 9, pp. 795-804, 2012.

[8] N. Sharma and A. Singh, "Exploring Biomarkers for Alzheimer's Disease," Journal of clinical and diagnostic research : JCDR, vol. 10, no. 7, pp. KE01-KE6, 2016.

[9] J. Pegueroles and et al, "Longitudinal brain structural changes in preclinical Alzheimer's disease," Alzheimer's & dementia : the journal of the Alzheimer's Association, vol. 13, no. 5, pp. 499-509, 2017.

[10] R. Duara and D. Loewenstein, "Effect of age, ethnicity, sex, cognitive status and APOE genotype on amyloid load and the threshold for amyloid positivity," Neuroimage Clin, vol. 22, no. 101800, 2019.

[11] D. Loewenstein and R. Curiel, "Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment," Neurology, vol. 91, no. 10, pp. e976-e984, 2018.

[12] F. Zhu and et al, "COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease," Scientific reports, vol. 6, no. 34567, 2016.

[13] C. Stonnington and et al, "Predicting clinical scores from magnetic resonance scans in Alzheimer's disease," NeuroImage, vol. 51, no. 4, pp. 1405-13, 2010.

[14] N. Bhagwat and et al, "An artificial neural network model for clinical score prediction in Alzheimer disease using structural neuroimaging measures," Journal of psychiatry & neuroscience, vol. 44, no. 4, pp. 246-260, 2019.

[15] D. Zhang, D. Shen and et al, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," Neuroimage, vol. 59, no. 2, pp. 895-907, 2012.

[16] S. Tabarestani and et al, "A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study," Neuroimage, vol. 206, no. 116317, 2019.

[17] W. Izquierdo, H. Martin and et al, "Robust prediction of cognitive test scores in Alzheimer's patients," in IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, 2017.

[18] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," TECHNOMETRICS, vol. 12, pp. 56-67, 1970.

[19] X. Golorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," Journal of Machine Learning Research, vol. 9, pp. 249-256, 2010.

[20] G. Hinton, N. Srivastava and K. Swersky, "Neural Networks for Machine Learning,"[Online].Available:http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. [Accessed 13 9 2020].