Alzheimer Disease Prediction Model Based on Decision Fusion of CNN-BiLSTM Deep Neural Networks

Shaker El-Sappagh[1], Tamer Abuhmed[2], and Kyung Sup Kwak[3](✉)

[1] Centro Singular de Investigación en Tecnoloxías Intelixentes, Universidade de Santiago de Compostela, Santiago, Spain
sh.elsappagh@gmail.com
[2] College of computing, Sungkyunkwan University, Seoul, Korea
tamer@skku.edu
[3] Department of Information and Communication Engineering, Inha University, Incheon, Korea
kskwak@inha.ac.kr

Abstract. Alzheimer's disease (AD) is a chronic neurodegenerative disorder. Early prediction of Alzheimer's progression is a crucial process for the patients and their families. As a chronic disease, AD data are multimodal and time series in nature. Building a deep learning model to optimize multi-objective cost function produces a more stable and accurate model. In this paper, we propose a multimodal multitask deep learning model for AD progression detection based five time series modalities and a collection of static data. The model predicts AD progression as a multi-class classification task and four critical cognitive scores as regression tasks. The experimental results show that our model is medically intuitive, more accurate, and more stable than the state-of-the-art studies.

Keywords: Alzheimer's disease · Convolutional neural network · Long Short-term memory · Time-series data analysis

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder [1]. AD has no care, and all current treatments are only for decelerating its progression. The early prediction of AD is a crucial step for timely treatment and progression delay. The early stage of AD is called mild cognitive impairment (MCI), where around 10% to 20% of MCI patients progress to AD per year [2]. AD is a chronic disease, where its conditions are gradually changing. Identifying patients who will not progress to AD (i.e. stable MCI [sMCI]) from progressive MCI (pMCI) patients, who will later be AD, is a complex process.

AD progression detection has been extensively studied by the literature [3, 4]. However, none of these studies reached a satisfactory performance level [5]. In addition, these studies are based on a limited set of features, mostly from neuroimaging data. The deployment of ML outputs in routine use for AD diagnoses is still very rare because the results of current studies are inconclusive. This is mainly due to the complexity of

K. Arai et al. (Eds.): IntelliSys 2020, AISC 1252, pp. 482–492, 2021.
https://doi.org/10.1007/978-3-030-55190-2_36

the problem, and also due to the inefficient utilization of available data and inefficient design of the ML models [6].

AD data are multimodal and longitudinal [7]. These data include heterogeneous modalities such as magnetic resonance imaging (MRI), cognitive scores, genetics, and cerebrospinal fluid (CSF). There is an increasing interest in multivariate approaches [8]. The majority of existing studies mainly depends on the extracted features from MRI neuroimaging, and they failed to accurately predict MCI-to-AD progression [9, 10]. Donnelly-Kehoe et al. [11] concluded that the maximum accuracy achieved by MRI features did not reach the one using the mini-mental state examination (MMSE) alone.

Furthermore, the current literature studies the single-modality and single-task models, where the model only optimizes a single objective function based on one type of data [12]. These models could not provide the optimized results, and integrating multiple data modalities yields comprehensive insights and more accurate results [13]. Liu et al. [14] used regular machine learning techniques to study the multimodal single-task classification problem. Besides, MCI progression detection has been designed as multimodal single-task models, where cognitive scores, such as MMSE, are used as indicators for AD progression [15]. In the medical environment, multiple modalities are chronically analyzed, and multiple clinical variables must be predicted. This ML paradigm is called multimodal multitask models, where each task has features from many sources, and multiple tasks are connected in a chronological sequence [7].

Most MCI progression studies binary sMCI vs. pMCI classification problem based exclusively on baseline data [2]. However, baseline data are less discriminative for progression detection than considering a patient's longitudinal data. AD is a chronic disease, so patient's data are always time series in nature. The disease state at a certain point in time is not independent of the state at a previous point in time. As a result, AD data are not only multimodal but also time series. However, most research studies do not consider this temporal/sequential nature of AD data [1]. Modeling AD progression as a multimodal multitask problem that based on time series data is a challenge, but it promises a great improvement in a models' performance [16].

All previous challenges could be effectively managed by using deep learning (DL) techniques [12, 17]. Choi and Jin [18] used CNN to predict pMCI based on PET images and a single-task model. Spasov et al. [19] proposed a multimodal single-task classification model based on a CNN to detect AD progression based on the late fusion of MRI, demographics, neuropsychological, and apolipoprotein E4 (APOe4). However, these models are less accurate, less sufficient, and not medically acceptable, because a medical expert usually studies the longitudinal multimodal patient data before making progression decisions [20]. Cui et al. [21] proposed a CNN-recurrent neural network (RNN) model for AD diagnosis based on MRI time-series data of six time-steps. Most DL models in the AD domain are implemented as binary classifications based on single modality and single task [22].

In this paper, we propose an advanced multimodal multi-task DL model for AD progression detection. We utilize the CNN and RNN to capture local and long-term temporal dependencies, where each time series is separately learned using a pipeline of stacked CNN-BiLSTM blocks. The framework utilizes AD time-series data to concurrently predict the progression status as a four-class classification task and four cognitive scores (e.g., Alzheimer's diseases assessment scale [ADAS], MMSE, functional assessment questionnaire [FAQ], and clinical dementia rating sum of boxes [CDRSB]) from five heterogeneous modalities. This system is medically intuitive, more stable, and more accurate than existing state-of-the-art studies.

2 Methods

In this section, we describe the medical dataset used for this study as well as the architecture of the deep learning model used to predict the AD progression and the critical medical scores related to the disease. Moreover, the methods used for data preparation are described at the end of the section.

2.1 Dataset

The ADNI is the used dataset to prepare this study. The study includes 1536 subjects (54.7% male). Eligible participants were 54 to 91 years of age. There are four classes (cognitively normal (CN), sMCI, pMCI, and AD) categorized according to the baseline diagnosis. There are 419 subjects diagnosed as CN, 473 subjects diagnosed as sMCI, 305 subjects diagnosed as pMCI, and 339 subjects diagnosed as AD. These data are time-series data, where each patient has 15 time-steps (i.e., baseline, month-6, month-12…month-84). Furthermore, the data are grouped in five modalities: MRI (326 features), PET (288 features), cognitive scores data (CSD, nine features), assessment data (neuropsychological battery and cognition report) (ASD, 51 features), and neuropathological data (NPD, seven features). MRI features are collected based on collected features by a team from the University of California at San Francisco (UCSF) using FreeSurfer software. More details on ADNI MRI neuroimaging procedures can be found at (http://adni.loni.ucla.edu). We collected 130 statistical measures for each of these markers and biomarkers, including min, max, etc. In addition, we used a set of static features like age, gender, number of years education, etc. The description of the dataset is shown in Table 1.

2.2 Proposed Model

Figure 1 illustrates the proposed DL model. This model learns AD progression based on a multimodal multitask paradigm. It will predict four regression tasks (i.e., four cognitive scores) and multi-class classification task (i.e., AD progression) based on multivariate time series data. The model optimizes a multi-objective cost function, which makes the model more accurate and more stable. We utilize five time-series modalities with 15 regular time steps (i.e., baseline, M06, M12… M84) and a collection of statistical features. We propose that the stacking of CNN and BiLSTM deep learning models could learn local and temporal features from time-series data. Every modality is learned by a stacked CNN-BiLSTM pipeline. The CNN layer has one Conv1D layer followed by max pooling. The BiLSTM layer has three stacked BiLSTM layers, an L2 regularization layer, and a dropout layer. The decisions of these five CNN-BiLSTM neural networks are fused and learned by a set of dense layers to make the final decisions. For each pipeline,

Table 1. Dataset description.

Feature	Baseline value	Month 84 value
Gender (M/F)	840/696	840/696
Age (years)	73.84 ± 07.17	73.84 ± 07.17
Education (years)	15.85 ± 02.90	15.85 ± 02.90
FAQ	04.71 ± 06.50	09.13 ± 10.15
MMSE	26.48 ± 03.51	24.59 ± 05.55
MoCA	22.21 ± 03.95	20.66 ± 05.40
ADAS 11	11.25 ± 06.84	15.53 ± 11.90
ADAS 13	17.69 ± 09.82	22.90 ± 15.45
Hippo. vol. (/1000)	06.68 ± 01.22	06.31 ± 01.32

Abbreviations: ADAS, Alzheimer's disease assessment scale; FAQ, functional assessment questionnaire; Hippo, hippocampus volume; MoCA, Montreal cognitive assessment.

the CNN network extracts local features in each time series, and them the BiLSTM subnetwork learns the temporal relationships among time-series features. Two dense layers take the Learned BiLSTM features for deeper feature learning. The outputs of the five pipelines are combined using three dense layers. The extracted statistical data are learned by a deep feed-forward neural network, and its learned features are concatenated with the learned features from the five modalities. Learned features from static and time series data are used to predict the final decisions.

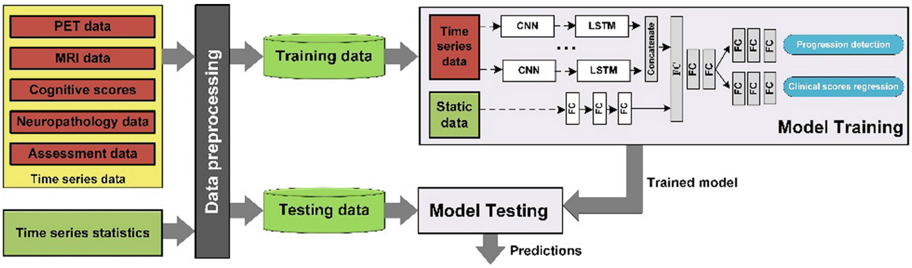

Fig. 1. The proposed DL model for AD progression detection

2.3 Data Preparation

2.3.1 Missing Data Handling

A feature with more than 30% missing is removed. A subject who has missing baseline data is excluded. According to the patient class, for numerical data, we use the mean value, and for categorical features, we use the mode value.

2.3.2 Data Standardization

Features are standardized using the z-score method, i.e., $z_j = (x_j - \mu_j)/\sigma_j)$, where x_j is the participant's original value of a feature j, z_j is the normalized value, μ_j is the feature's mean, and σ_j is the feature's standard deviation.

2.3.3 Feature Reduction

We used principal component analysis (PCA) to reduce the number of features of MRI and PET data. PCA is implemented with a retained variance of more than 91%. PCA reduces the MRI features from 326 to 110. PCA reduces PET features from 288 to 75.

3 Results

In this section, we present the setup and results of our proposed multimodal multitask architecture. Besides, we discuss the results of both the diagnoses and medical scores predictions.

3.1 Setup

The checked the role of time-series data on the model confidence and accuracy as we increase the number of time steps. A total of 15 DL models are trained based on *baseline data* (BL), $BL + M06$, $BL + M06 + M12$,..., $BL + \ldots + M84$. We employed a computer with Intel® Xeon(R) CPU E5-2620 v3 @ 2.40 GHz × 24 with Cuda-10.0 and two GPU GEFORCE GTX TITANx 12 GB; with Python 3.7.3 distributed in Anaconda 4.7.7 (64-bit). The model is implemented using Keras library based on TensorFlow as a backend. The classification task is based on the SoftMax activation function and cross-entropy loss, while the sigmoid activation function with mean square error loss is used for regression tasks. Adam optimizer (lr = 0.0001) is used [23], batch size is 32, and epochs 90. The overfitting is prevented using dropout, L2 regularization, and early stopping mechanisms. The dataset is randomly split in 60/20/20% (922/307/307cases) for training, validation, and testing, respectively. The classification performance will be measured by accuracy, precision, recall, and F1-score metrics. The regression performance is evaluated by the mean absolute error (MAE) measure.

3.2 Multimodal Multitask Experiment

Figure 2 shows the performance of the multiclass classification task. The model has been tested using accumulative time series data. In other words, the model has been tested using only baseline (bl) time step, then based on bl + M06, then bl + M06 + M12, …, bl + M06 + … + M84. On average, this task achieves accuracy (92.62 ± 2.41%), precision (94.02 ± 3.26%), F1-score (92.56 ± 2.38%), and recall (98.42 ± 1.38%). Adding more time steps increases the noise of data; however, we noticed that the performance of the model is improved as we add more time steps. For instance, using baseline data only, the classification accuracy is 89.90%, but with the 15-time steps, the task achieves accuracy

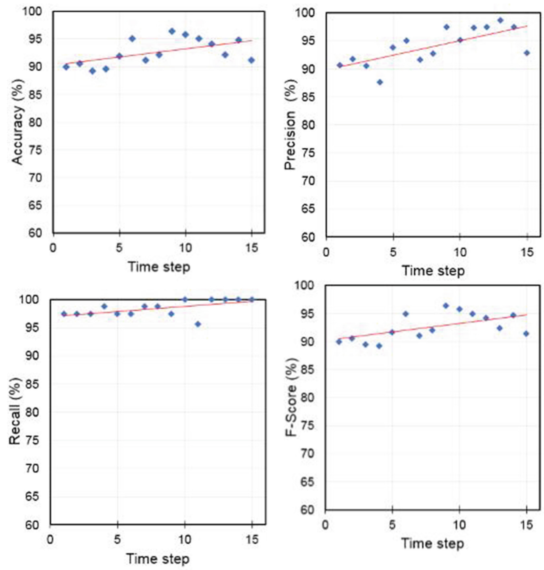

Fig. 2. The multiclass classification task performance. The solid red lines represent regression lines.

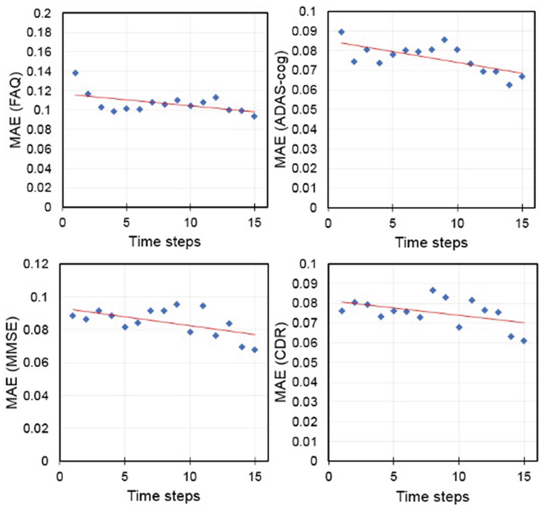

Fig. 3. The four regression tasks for the values of critical cognitive scores at progression time. The solid red lines represent regression lines.

of 91.21%. The same behavior is done with the precision, F1-score, and recall where adding of 15-time steps improves them by 2.16%, 1.39%, and 2.56%, respectively.

It is worth noting that the multimodal multitask model achieved better results compared to the single-task classification model, where the average accuracy, F1-score, and recall are improved by 5.98%, 5.78%, and 7.5%, respectively. The resulting multitask learning is more confident and more stable, as asserted by the literature [16, 24]. By utilizing the 15 time-steps, the system achieves accuracy (91.21%), F1-score (91.36%), and recall (99.99%).

We have four regression tasks corresponding to the prediction of the regression variables of MMSE, ADAS, FAQ, and CDR values at progression time. Figure 3 shows the performance of the model based on MAE. The system achieves an average MAE of (0.107 ± 0.01), (0.076 ± 0.01), (0.075 ± 0.01), and (0.085 ± 0.01), for the FAQ, ADAS, CDR, and MMSE tasks, respectively.

The simultaneous training of related multitasks improves the performance of all tasks and make the model more stable. In addition, as shown in Fig. 3, we find that adding more time steps improves the performance of all regression tasks. The system approximately achieves the lowest error rate when it uses the 15-time steps (i.e. MAE rates are 0.094, 0.067, 0.061, and 0.068 for FAQ, ADAS, CDR, and MMSE tasks, respectively).

Table 2. Correlation coefficient for the 15-time steps.

Time step	FAQ	ADAS-cog	CDR	MMSE
1	0.799	0.869	0.834	0.732
2	0.818	0.820	0.815	0.725
3	0.860	0.809	0.813	0.759
4	0.863	0.813	0.815	0.746
5	0.881	0.817	0.857	0.775
6	0.888	0.809	0.851	0.748
7	0.868	0.819	0.847	0.746
8	0.875	0.823	0.835	0.747
9	0.883	0.808	0.848	0.735
10	0.891	0.824	0.885	0.790
11	0.894	0.854	0.865	0.784
12	0.884	0.846	0.870	0.802
13	0.898	0.860	0.895	0.816
14	0.911	0.881	0.914	0.864
15	0.909	0.820	0.869	0.774

On average, the best performing task is the ADAS ($P < 0.007$), and the worst one is FAQ ($P < 0.0001$). The predicted and original scores have a high correlation in the 15-time steps. The average correlation coefficient for the FAQ, ADAS, CDR, and MMSE features are 0.875 ± 0.031, 0.832 ± 0.024, 0.854 ± 0.030, 0.770 ± 0.037, respectively. The average real and predicted scores are (9.56 ± 9.54, 9.69 ± 8.23) for FAQ, (23.57 ± 14.09, 24.54 ± 12.81) for ADAS, (3.53 ± 3.60, 3.93 ± 3.30) for CDR, and (24.61

± 5.11, 23.95 ± 4.78) for MMSE, respectively. We notice that as we add a new time step, the model becomes more confident and the correlation coefficient increases for all tasks, as shown in Table 2. The best correlation is achieved at time step 14 (i.e., at M78), i.e., it is 0.911, 0.881, 0.914, and 0.864 for FAQ, ADAS, CDR, and MMSE scores, respectively; the predicted and real values for test data at time 14 can be seen Fig. 4 and Fig. 5.

Our model has been compared with the regular machine learning techniques. These models are used to predict each task separately by using the static and time series statistic data. The stratified 10-fold cross-validation was used to train and validate these models, and each experiment is repeated ten times.

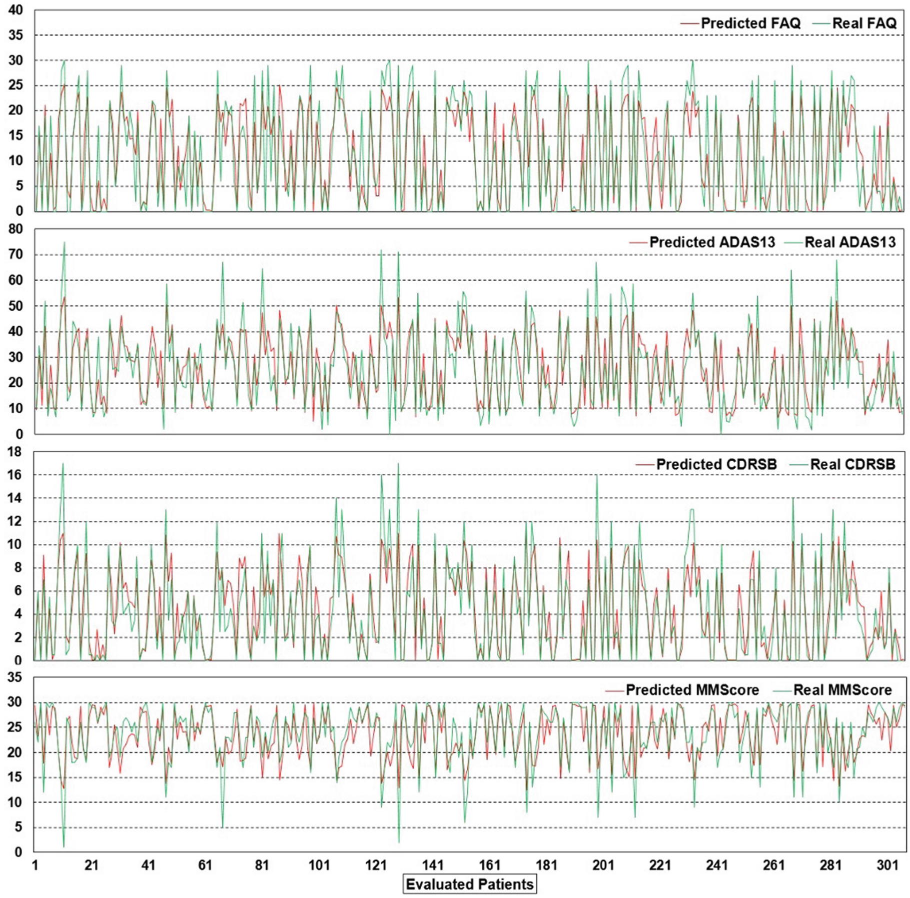

Fig. 4. Correlation between predicted and real cognitive scores for time step 14.

Regarding the classification task, the proposed model has been compared with eight state-of-the-art classification algorithms, i.e., SVM, logistic regression (LR), random forest (RF), k-nearest neighbor (KNN), decision tree (DT), extreme gradient boosting

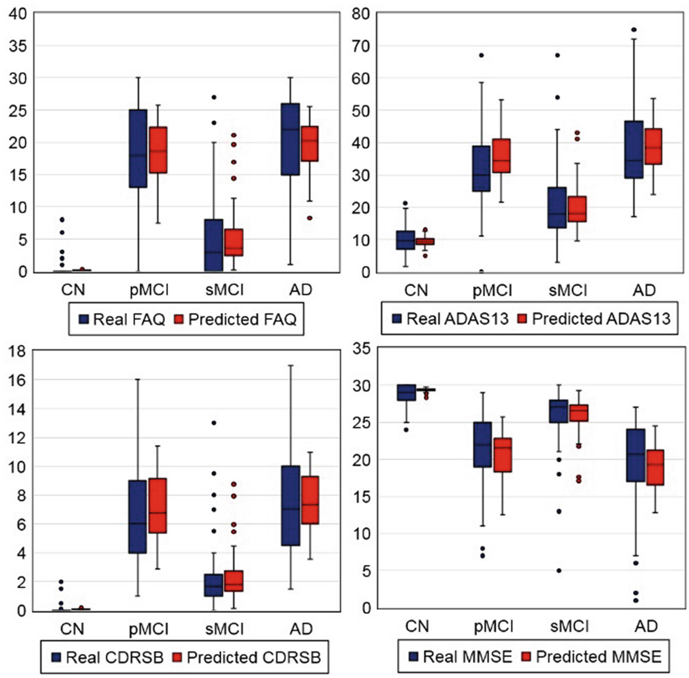

Fig. 5. Box plots for the distribution of predicted and real cognitive scores.

(XGBoost), naive Bayes (NB), and multilayer perceptron (MLP), as benchmark methods. Our DL model achieved much better performance than all conventional ML techniques. SVM achieved the highest performance compare to other regular ML techniques (i.e., accuracy = 85.55 ± 1.07, precision = 8616 ± 0.81, recall = 8570 ± 1.30, and F1-score = 8592 ± 0.84). The proposed model had an average of 7.07%, 7.86%, 12.72%, and 6.64% performance gain in comparison with SVM 620 for accuracy, precision, recall, and F1-score, respectively.

Regarding regression tasks, we compared our model with RF, SVM, Bayesian ridge (BR), lasso, DT, and gradient boosting regression (GBR). BR had the best results for FAQ regression task (MAE of 0.113 ± 0.020) and CDR task (MAE of 0.104 ± 0.015), and RF had the best result for ADAS (MAE of 0.113 ± 0.017) and MMSE (MAE of 0.120 ± 0.019). Our model had lower MAE for all regression tasks by 0.006 for FAQ, 0.037 for ADAS, 0.029 for CDR, and 0.035 for MMSE. On average, the MAE for RF, SVM, BR, lasso, DT, and GBR were 0.117 ± 0.019, 0.159 ± 0.025, 0.119 ± 0.019, 0.154 ± 0.017, 0.144 ± 0.019, and 0.137 ± 0.017, respectively.

The proposed model is medically intuitive because it is based on five time-series modalities data, which provide complementary information to make an accurate decision. The model is based on a novel stacked CNN-LSTM deep learning model. The majority of state-of-the-art studies are based on baseline MRI data only [7]. Even though they achieved good results, these types of studies are not medically acceptable because they did not mimic the real environments. Our study has several limitations that will be handled in the future including: (1) the inclusion of more data about patient's history such as medications and comorbidities, (2) the explainability of the given decisions, and (3) the enhancement of the performance by using ensemble models.

4 Conclusion

This paper proposed a multimodal multitask DL model for AD progression detection. The model is based on five stacked CNN-BiLSTM models trained concurrently based on five time-series data. The model has been implemented and tested using the ADNI time-series datasets. The model achieved state-of-the-art results for both classification and regression tasks.

Acknowledgment. This work was supported by National Research Foundation of Korea-Grant funded by the Korean Government (Ministry of Science and ICT)-NRF-2017R1A2B2012337.

This research was funded by the Galician Ministry of Education, University and Professional Training and by the European Regional Development Fund (ERDF/FEDER program) under grants ED431C 2018/29 and ED431G 2019/04.

References

1. Alberdi, A., Aztiria, A., Basarab, A.: On the early diagnosis of Alzheimer's disease from multimodal signals: a survey. Artif. Intell. Med. **71**, 1–29 (2016). https://doi.org/10.1016/j.artmed.2016.06.003
2. Li, H., Habes, M., Wolk, D.A., Fan, Y.: A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal MRI 1–12 (2019). https://doi.org/10.1016/j.jalz.2019.02.007
3. Ramírez, J., Górriz, J.M., Ortiz, A., Martínez-Murcia, F.J., Segovia, F., Salas-Gonzalez, D., Castillo-Barnes, D., Illán, I.A., Puntonet, C.G.: Ensemble of random forests one vs. rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares. J. Neurosci. Methods **302**, 47–57 (2018). https://doi.org/10.1016/j.jneumeth.2017.12.005
4. Yao, D., Calhoun, V.D., Fu, Z., Du, Y., Sui, J.: An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. J. Neurosci. Methods **302**, 75–81 (2018). https://doi.org/10.1016/j.jneumeth.2018.03.008
5. Bucholc, M., Ding, X., Wang, H., Glass, D.H., Wang, H., Prasad, G., Maguire, L.P., Bjourson, A.J., McClean, P.L., Todd, S., Finn, D.P., Wong-Lin, K.F.: A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert Syst. Appl. **130**, 157–171 (2019). https://doi.org/10.1016/j.eswa.2019.04.022
6. Polikar, R., Topalis, A., Parikh, D., Green, D., Frymiare, J., Kounios, J., Clark, C.M.: An ensemble based data fusion approach for early diagnosis of Alzheimer's disease. Inf. Fusion. **9**, 83–95 (2008). https://doi.org/10.1016/j.inffus.2006.09.003
7. Liu, M., Zhang, J., Adeli, E., Shen, D.: Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Trans. Biomed. Eng. **1** (2018). https://doi.org/10.1109/TBME.2018.2869989
8. Ruiz, E., Ramírez, J., Górriz, J.M., Casillas, J.: Alzheimer's disease computer-aided diagnosis: histogram-based analysis of regional MRI volumes for feature selection and classification. J. Alzheimer's Dis. **65**, 819–842 (2018). https://doi.org/10.3233/JAD-170514

9. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage **56**, 766–781 (2011). https://doi.org/10.1016/j.neuroimage.2010.06.013
10. Wang, T., Qiu, R.G., Yu, M.: Predictive modeling of the progression of Alzheimer ' s disease with recurrent neural networks. Sci. Rep. 1–12 (2018). https://doi.org/10.1038/s41598-018-27337-w
11. Donnelly-Kehoe, P.A., Pascariello, G.O., Gómez, J.C.: Looking for Alzheimer's disease morphometric signatures using machine learning techniques. J. Neurosci. Methods **302**, 24–34 (2018). https://doi.org/10.1016/j.jneumeth.2017.11.013
12. Qiu, S., Chang, G.H., Panagia, M., Gopal, D.M., Au, R., Kolachalama, V.B.: Fusion of deep learning models of MRI scans, mini–mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimer's Dement. Diagn. Assess. Dis. Monit. **10**, 737–749 (2018). https://doi.org/10.1016/j.dadm.2018.08.013
13. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage **104**, 398–412 (2015). https://doi.org/10.1016/j.neuroimage.2014.10.002
14. Liu, F., Zhou, L., Shen, C., Yin, J.: Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. IEEE J. Biomed. Heal. Inform. **18**, 984–990 (2013). https://doi.org/10.1109/jbhi.2013.2285378
15. Duchesne, S., Caroli, A., Geroldi, C., Collins, D.L., Frisoni, G.B.: Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. Neuroimage **47**, 1363–1370 (2009). https://doi.org/10.1016/j.neuroimage.2009.04.023
16. Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G., Galstyan, A.: Multitask learning and benchmarking with clinical time series data. Sci. data. **6**, 96 (2019). https://doi.org/10.1038/s41597-019-0103-9
17. Lu, D., Popuri, K., Ding, W., Balachandar, R., Beg, M.F.: Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Sci Rep. **8**, 5697 (2018). https://doi.org/10.1038/s41598-018-22871-z
18. Choi, H., Jin, K.H.: Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behav. Brain Res. **344**, 103–109 (2018). https://doi.org/10.1016/j.bbr.2018.02.017
19. Spasov, S., Passamonti, L., Duggento, A., Lio, P., Toschi, N.: A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease **189**, 383687 (2018). https://doi.org/10.1101/383687
20. Ding, X., Bucholc, M., Wang, H., Glass, D.H., Wang, H., Clarke, D.H., Bjourson, A.J., Dowey, L.R.C., O'Kane, M., Prasad, G., Maguire, L., Wong-Lin, K.: A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. Sci. Rep. **8**, 1–10 (2018). https://doi.org/10.1038/s41598-018-27997-8
21. Cui, R., Liu, M., Initiative, N.: RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput. Med. Imaging Graph. (2019). https://doi.org/10.1016/j.compmedimag.2019.01.005
22. Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., Guaragnella, C., Bellotti, R., Tangaro, S.: Deep learning reveals Alzheimer's disease onset in MCI subjects: results from an international challenge. J. Neurosci. Methods **302**, 3–9 (2018). https://doi.org/10.1016/j.jneumeth.2017.12.011
23. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization, PP. 1–15 (2014)
24. Ruder, S.: An Overview of Multi-Task Learning in Deep Neural Networks. arXiv Prepr. arXiv1706.05098 (2017). https://doi.org/10.1109/CVPR.2015.7299170