International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Time Series Analysis of Cognitive Scores for Alzheimer's Prediction: An LSTM Deep Learning Approach

Amuloju Bala Bhavani[1], Sirisha Balla[2], Cherukuri Anusha[3], Dr. Prasad Rayi[4]

Submitted: 07/09/2024 **Revised:** 20/10/2024 **Accepted:** 29/10/2024

Abstract: Timely identification of persons predisposed to Alzheimer's disease (AD) dementia is crucial for the development of disease-modifying therapeutics. This research aims to predict the clinical diagnosis, cognitive function, and ventricular volume of a person at each subsequent month indefinitely, based on multimodal Alzheimer's disease indicators and clinical diagnoses from one or more timepoints. We introduced a recurrent neural network (RNN) model and used it on data from The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge, which includes longitudinal data from 1,677 people (Marinescu et al. 2018) sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We evaluated the efficacy of the RNN model against three baseline algorithms over a forecast period of six years. Most prior research on forecasting Alzheimer's disease development neglects the problem of missing data, a common challenge in longitudinal studies. We examined three distinct ways for addressing missing data. Two of the solutions addressed the missing data as a "preprocessing" concern by imputing the absent data via the prior timepoint ("forward filling") or linear interpolation ("linear filling"). The third technique used the RNN model to complete the missing data during both training and testing, referred to as "model filling." Our findings indicate that the RNN using "model filling" outperformed baseline techniques, such as support vector machines/regression and linear state space (LSS) models. Nonetheless, there was no statistically significant difference between the RNN and LSS in predicting cognition and ventricular volume. Significantly, while using longitudinal data in the training process, our analysis revealed that the trained RNN model had comparable performance whether utilizing either one or four input timepoints, indicating that our methodology may be effective with just cross-sectional data. An previous iteration of our methodology achieved a 5th place ranking among 53 submissions in the TADPOLE competition in 2019. The present methodology is positioned 2nd among 56 submissions as of August 12, 2019.

Keywords: *predisposed, dementia, therapeutics, timepoints, Neuroimaging, methodology.*

Introduction

Alzheimer's disease (AD) dementia is a debilitating neurological condition characterized by an extended prodromal phase and the absence of a treatment. An successful treatment plan should early target persons at risk for Alzheimer's disease (Scheltens et al., 2016). As a result, there is considerable interest in forecasting the longitudinal course of illness in people. A primary challenge is that although Alzheimer's disease often manifests as an amnestic state, there exists considerable variation across people (Murray et al., 2011; Noh et al., 2014; Zhang et al., 2016; Risacher et al., 2017; Young et al., 2018; Sun et al., 2019). Given that Alzheimer's disease dementia is characterized by injuries mediated by beta-amyloid and tau, leading to brain atrophy and cognitive deterioration (Jack et al., 2010, 2013), a multimodal strategy may prove more efficacious than a singular modality in elucidating this heterogeneity and forecasting longitudinal disease progression (Marinescu et al., 2018). This work presents a machine learning method designed to forecast multimodal Alzheimer's disease indicators (e.g., ventricular volume, cognitive scores) and the clinical diagnosis of individual individuals for each month extending up to six years ahead. Most prior research has concentrated on a "static" variant of the problem, wherein the objective is to forecast a singular timepoint (Duchesne et al., 2009; Stonnington et al., 2010; Zhang and Shen, 2012; Moradi et al., 2015; Albert et al., 2018; Ding et al., 2018) or a predetermined set of future timepoints (regularized regression; (Wang et al., 2012; Johnson et al., 2012; McArdle et al., 2016; Wang et al., 2016)). In contrast, our objective is the longitudinal forecasting of clinical diagnoses and multimodal Alzheimer's disease markers at an ostensibly infinite number of future timepoints, as delineated by The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge

[1,2,3,4]*International School Of Technology And Sciences For Women, A.P, India..*

(Marinescu et al., 2018), which is arguably a more pertinent and comprehensive aim for tasks such as prognosis and cohort selection.

A prevalent method for addressing this longitudinal prediction challenge is mixed-effect regression modeling, which characterizes the longitudinal trajectories of Alzheimer's disease biomarkers using linear or sigmoidal curves (Vemuri et al., 2009; Ito et al., 2010; Sabuncu et al., 2014; Samtani et al., 2012; Zhu and Sabuncu, 2018). Nonetheless, this modeling technique requires previous knowledge of the geometries of the biomarker trajectories. Moreover, whereas biomarker trajectories may exhibit linear or sigmoidal patterns when averaged across individuals (Caroli and Frisoni, 2010; Jack et al., 2010; Sabuncu et al., 2011), individual patients may substantially diverge from the anticipated parametric forms.

Methods

Problem setup

The issue configuration adheres to that of the TADPOLE challenge (Marinescu et al. 2018). Utilizing multimodal Alzheimer's disease markers and the diagnostic status of a participant from one or more timepoints, we aim to forecast cognition (as assessed by ADAS-Cog13; Mohs et al., 1997), ventricular volume (as evaluated by structural MRI), and clinical diagnosis of the participant for each month indefinitely into the future.

Data

We used the data supplied by the TADPOLE challenge (Marinescu et al., 2018). The dataset included 1,677 participants from the ADNI database (Jack et al., 2008). Every subject had scanning at many time intervals. The mean number of timepoints was 7.3 ± 4.0 (Figure 1A), while the mean duration from the first timepoint to the last timepoint was 3.6 ± 2.5 years (Figure 1B).

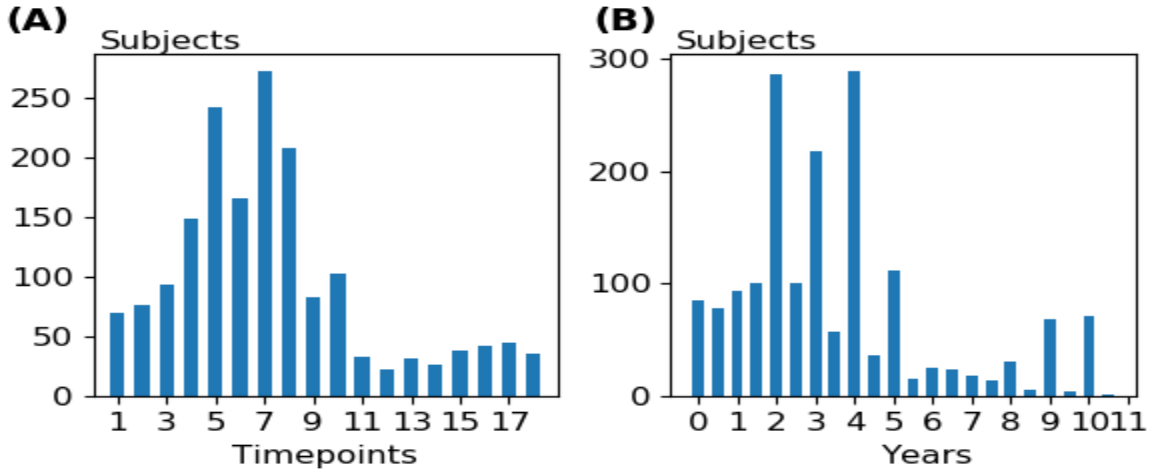

Figure 1. (A) Distribution of the number of timepoints for all subjects in the dataset. (B) Distribution of the number of years between the first and last timepoints for all subjects in the dataset.

To ensure consistency, we used the same collection of 23 variables suggested by the TADPOLE challenge, including diagnosis, neuropsychological test scores, anatomical characteristics obtained from T1 magnetic resonance imaging (MRI), positron emission tomography (PET) metrics, and cerebrospinal fluid (CSF) biomarkers (Table 1). The diagnostic classifications were normal control (NC), moderate cognitive impairment (MCI), and Alzheimer's disease (AD). We arbitrarily partitioned the data into training, validation, and test subsets. The proportion of participants in the training, validation, and test sets was 18:1:1. The training dataset was used to train the model. The validation set was used to determine the hyperparameters. The test set was used to assess the models' performance. In the validation and test sets, the first half of the timepoints for each subject was used to forecast the subsequent half of the timepoints for the same subject. All variables, with the exception of the diagnostic category, which was categorical rather than continuous, were z-normalized. Z-normalization was conducted on the training set. The mean and standard deviation from the training set were then used to z-normalize the validation and test sets. The data was randomly divided into training, validation, and test sets.

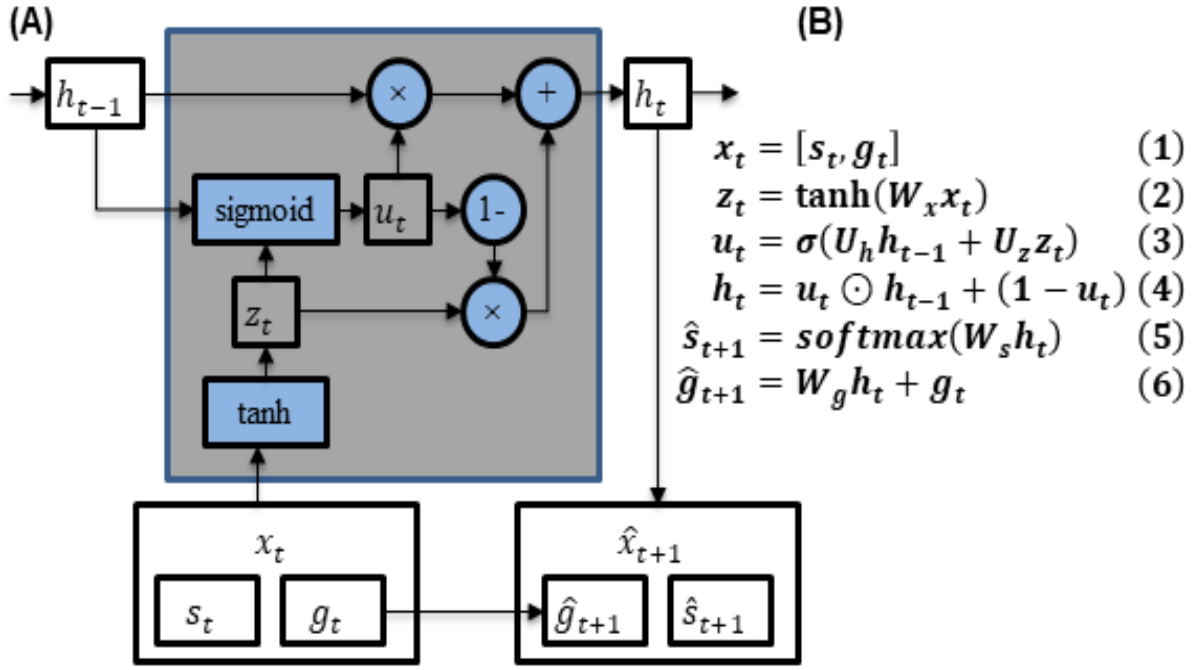

Figure 2. (A) MinimalRNN. (B) MinimalRNN update equations.

$\boldsymbol{st}$ and $\boldsymbol{gt}$ represent categorical (i.e., diagnostic) and continuous variables, respectively (Table 1). The input $\boldsymbol{xt}$ to each RNN cell consisted of the diagnostic $\boldsymbol{st}$ and continuous variables $\boldsymbol{gt}$ (Eq. 1). Observe that $\boldsymbol{st}$ was shown with one-hot encoding. The concealed state $\boldsymbol{ht}$ was an amalgamation of the preceding hidden state $\boldsymbol{ht}-\mathbf{1}$ and the altered input $\boldsymbol{zt}$ (Eq. 4). The forget gate $\boldsymbol{ut}$ evaluated the influences of the preceding hidden state $\boldsymbol{ht}-\mathbf{1}$ and the current transformed input $\boldsymbol{zt}$ on the present hidden state $\boldsymbol{ht}$ (Eq. 3). The model forecasted the subsequent month's diagnostic $\hat{\boldsymbol{s}}\boldsymbol{t}+\mathbf{1}$ and continuous variables $\hat{\boldsymbol{g}}\boldsymbol{t}+\mathbf{1}$ using the hidden state $\boldsymbol{ht}$ (Eqs. 5 and 6). $\odot$ and $\boldsymbol{\sigma}$ represent the element-wise product and the sigmoid function, respectively. We modified the minimalRNN (Chen, 2017) for forecasting illness progression. The model architecture and update equations are shown in the figure. 1. Let $\boldsymbol{xt}$ represent all variables observed at time t, including the diagnosis $\boldsymbol{st}$ and the other continuous variables $\boldsymbol{gt}$ (Eq. 1 in Figure 2B). Diagnosis was expressed using one-hot encoding. In other words, the diagnosis was shown as a vector of length three. If the first input was one, then the person was classified as a normal control. If the second input was one, then the person had minor cognitive impairment. If the third entry was one, then the subject had Alzheimer's disease dementia. a. Currently, we presume that all variables were recorded at every timepoint; the problem of missing data will be discussed in Section 2. At each time point, the transformed input $\boldsymbol{zt}$ (Equation 2 in Figure 2) and the preceding hidden state $\boldsymbol{ht}-\mathbf{1}$ were used to update the hidden state $\boldsymbol{ht}$ (Equations 3 and 4 in Figure 2B). The concealed state may be seen as encapsulating all knowledge about the issue up to that specific moment. The concealed state $\boldsymbol{ht}$ was then used to forecast the observations at the subsequent timepoint $\boldsymbol{xt}+\mathbf{1}$ (Eqs. 5 and 6 in Figure 1B). Data in the ADNI database were gathered at a minimum period of six months. Data may, in reality, be gathered at an unanticipated period (e.g., month 8 instead of month 6). The interval between timepoints t and t+1 in the RNN was established as 1 month. th

Results

Overall performance

Figure 3 depicts the test performance of minimalRNN alongside three baseline models: SVM/SVR, constant prediction, and LSS. For clarity, we only presented RNN with mixed filling (RNN–MF), LSS with mixed filling (LSS–MF), and SVM/SVR with a single input timepoint, since they produced the most favorable outcomes within their respective model categories. The test performance of all models (RNN, SVM/SVR, constant

prediction, and LSS) across the three ways for handling missing data.

We conducted statistical analyses comparing the three RNN variations (RNN–FF, RNN–LF, and RNN–MF) against all other baseline methods (LSS, constant prediction, SVM/SVR). Multiple comparisons were adjusted using a false discovery rate (FDR) of $q < 0.05$. In clinical diagnostic prediction, RNN–MF exhibited superior performance and was statistically more effective than other baseline methods (LSS, constant prediction, SVM/SVR). Regarding ADAS-Cog13 and ventricular volume, RNN-MF exhibited superior performance and was statistically more effective than other baseline methods, with the exception of LSS with model filling (LSS–MF; $p = 0.59$).

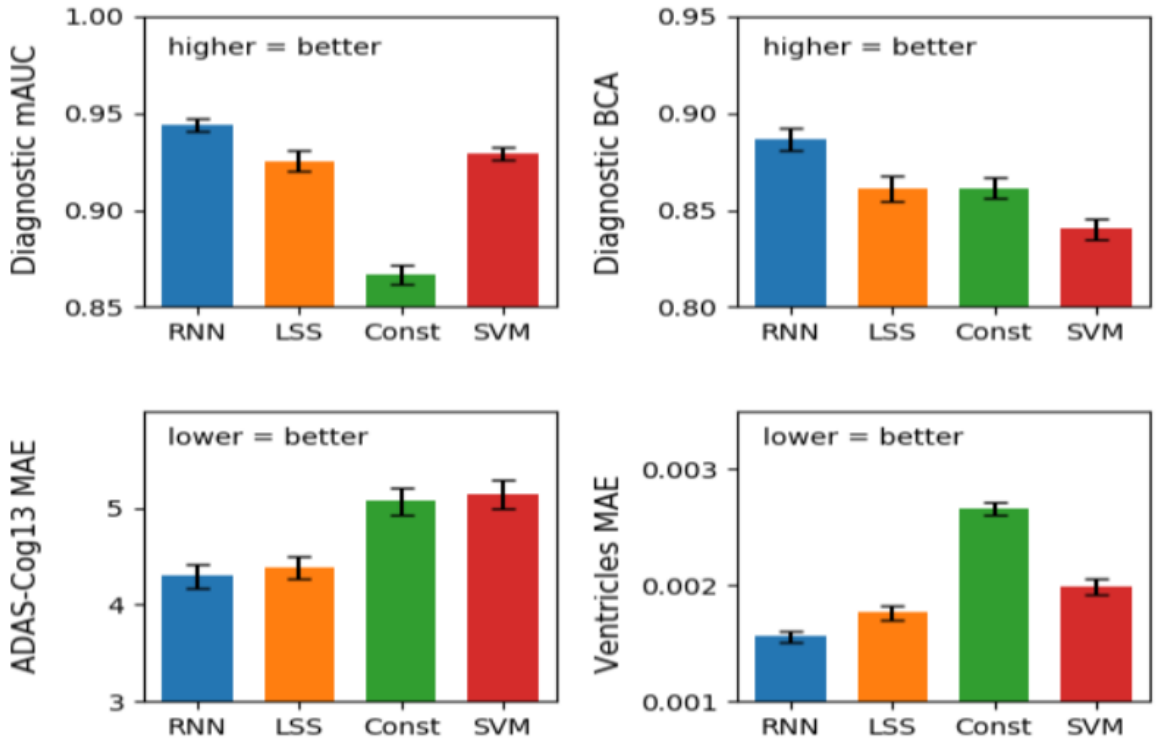

Figure 3. Performance of the best models from each model class averaged across 20 test sets.

Error bars represent the standard error across the test sets. In clinical diagnostics, elevated mAUC and BCA values signify superior performance. A reduced MAE for ADAS-Cog13 and Ventricles signifies superior performance. The RNN model is equivalent to RNN–MF as shown in Table 4. The In both RNN and LSS, mixed filling outperformed forward filling and linear filling, particularly in the prediction of ADAS-Cog13 and ventricular volume (Table 4). Notably, an increased number of input timepoints does not inherently enhance prediction accuracy in the context of SVM/SVR. The SVM/SVR model using a single timepoint demonstrated superior numerical performance compared to SVM/SVR models employing multiple timepoints, however the differences were minimal.

Figure 4 illustrates the disaggregation of the predictive performance from Figure 8 in annual intervals extending up to six years into the future. The performance of all algorithms deteriorated for forecasts extending deeper into the future. The constant baseline exhibited strong competitiveness versus other models in the first year; however, performance declined rapidly in later years. The RNN model was either equivalent to or superior to all baseline methods over all years.

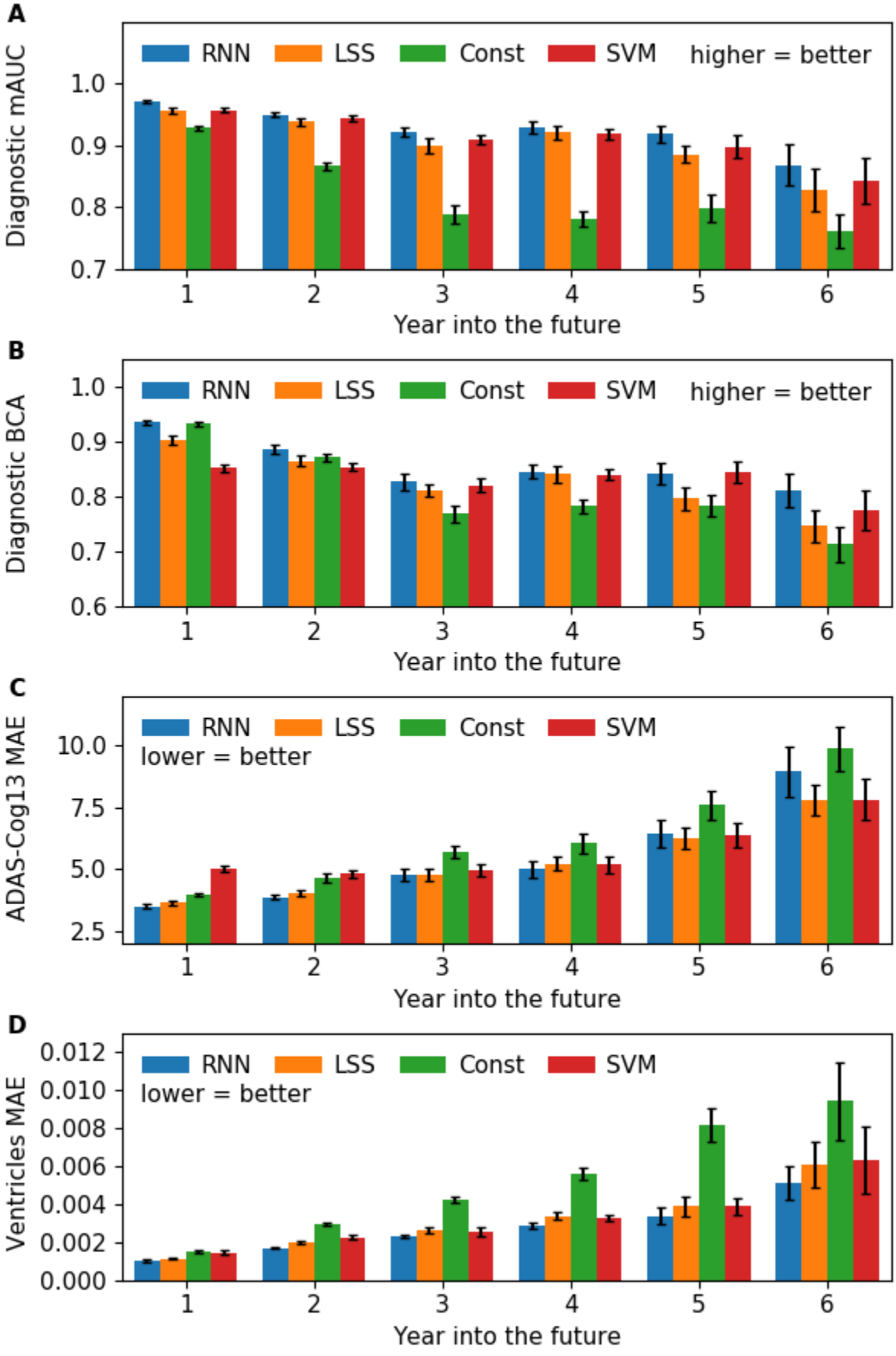
A
Diagnostic mAUC
RNN
LSS
Const
SVM
higher = better
1.0
0.9
0.8
0.7
1
2
3
4
5
6
Year into the future
B
Diagnostic BCA
RNN
LSS
Const
SVM
higher = better
1.0
0.9
0.8
0.7
0.6
1
2
3
4
5
6
Year into the future
C
ADAS-Cog13 MAE
RNN
LSS
Const
SVM
lower = better
10.0
7.5
5.0
2.5
1
2
3
4
5
6
Year into the future
D
Ventricles MAE
RNN
LSS
Const
SVM
lower = better
0.012
0.010
0.008
0.006
0.004
0.002
0.000
1
2
3
4
5
6
Year into the future

Figure 4

Discussion

In this study, we modified a simple RNN model to forecast longitudinal progression in Alzheimer's disease dementia. Our methodology demonstrated superior performance relative to baseline techniques, including SVM/SVR and LSS models. Nonetheless, we see that there was no statistical difference between the minimalRNN and LSS in predicting ADAS-Cog13 and ventricular volume, despite earlier research indicating advantages of modeling non-linear interactions among characteristics (Popescu et al., 2019).

When establishing the SVM/SVR baseline models (Section 2.5.2), several edge situations must be addressed to modify a "static" prediction method (e.g., SVM/SVR) for the more "dynamic" longitudinal prediction issue examined here. For instance, data is always squandered since static methods often presume that participants possess a same quantity of input timepoints. Consequently, for the SVM/SVR models using four input timepoints, we ultimately retained just 1454 people from the initial 1677 participants. This may elucidate why the SVM/SVR model using one input timepoint performed better compared to the SVM/SVR model employing four input timepoints (Table 4). Additionally, we constructed many distinct SVM/SVR models to forecast a certain number of future timepoints and executed interpolation at intermediate timepoints. In contrast, state-based models (e.g., minimalRNN or LSS) exhibit more elegance since they accommodate participants with varying numbers of timepoints and, in theory, can forecast an endless number of future timepoints.

Although the ADNI dataset included individuals with several timepoints, for the algorithm to be therapeutically effective, it must proficiently manage missing data and persons with just a single input timepoint. Our findings indicate that the "integrative" method of using the model for data imputation (i.e., model filling) is superior than "preprocessing" techniques, such as forward filling or linear interpolation. Nonetheless, it is conceivable that more advanced "preprocessing" techniques, such as matrix factorization (Mazumder et al., 2010; Nie et al., 2017; Thung et al., 2016) or wavelet interpolation (Mondal and Percival, 2010), might provide superior outcomes. Our model filling methodology may be seen as a variant of matrix completion, since the RNN (or LSS) was trained to minimize prediction loss, which corresponds to maximizing the probability of the training data. Matrix completion often presumes that the training data may be expressed as a matrix that is factorable into low-rank or other specifically structured matrices.

Conclusion

Utilizing 1677 people from the ADNI database, we demonstrated that the minimalRNN model outperformed alternative baseline algorithms in the longitudinal prediction of multimodal Alzheimer's disease biomarkers and clinical diagnoses of patients up to six years ahead. We examined three distinct techniques to address the problem of missing data in longitudinal studies. The RNN model can independently address the missing data problem, so offering a comprehensive technique for managing data deficiencies. Additionally, we discovered that after training with longitudinal data, the trained RNN model demonstrates commendable performance with a single input timepoint, indicating that this methodology may also be applicable to cross-sectional data.

References

[1] Aksman, L.M., Scelsi, M.A., Marquand, A.F., Alexander, D.C., Ourselin, S., Altmann, A., for ADNI, 2019. Modeling longitudinal imaging biomarkers with parametric Bayesian multi-task learning. Hum. Brain Mapp. https://doi.org/10.1002/hbm.24682

[2] Albert, M., Zhu, Y., Moghekar, A., Mori, S., Miller, M.I., Soldan, A., Pettigrew, C., Selnes, O., Li, S., Wang, M.-C., 2018. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. Brain J. Neurol. 141, 877–887. https://doi.org/10.1093/brain/awx365

[3] Azur, M.J., Stuart, E.A., Frangakis, C., Leaf, P.J., 2011. Multiple Imputation by Chained Equations: What is it and how does it work? Int. J. Methods Psychiatr. Res. 20, 40–49. https://doi.org/10.1002/mpr.329

[4] Bouckaert, R.R., Frank, E., 2004. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms, in: Dai, H., Srikant, R., Zhang, C. (Eds.), Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 3–12.

[5] Byun, M.S., Kim, S.E., Park, J., Yi, D., Choe, Y.M., Sohn, B.K., Choi, H.J., Baek, H., Han, J.Y., Woo, J.I., Lee, D.Y., Initiative, A.D.N., 2015. Heterogeneity of Regional Brain Atrophy Patterns Associated with Distinct Progression Rates in Alzheimer's Disease. PLOS ONE 10, e0142756. https://doi.org/10.1371/journal.pone.0142756

[6] Caroli, A., Frisoni, G.B., 2010. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. Neurobiol. Aging 31, 1263–1274. https://doi.org/10.1016/j.neurobiolaging.2010.04.024

[7] Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y., 2018. Recurrent Neural Networks for Multivariate Time Series with Missing Values. Sci. Rep. 8, 6085. https://doi.org/10.1038/s41598-018-24271-9

[8] Chen, M., 2017. MinimalRNN: Toward More Interpretable and Trainable Recurrent Neural Networks. ArXiv171106788 Cs Stat.

[9] Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

[10] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J., 2016. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, in: JMLR Workshop and Conference Proceedings. NIH Public Access, p. 301.

[11] Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., Lituiev, D., Copeland, T.P., Aboian, M.S., Mari Aparici, C., Behr, S.C., Flavell, R.R., Huang, S.-Y., Zalocusky, K.A., Nardo, L., Seo, Y., Hawkins, R.A., Hernandez Pampaloni, M., Hadley, D., Franc, B.L., 2018. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology 290, 456–464. https://doi.org/10.1148/radiol.2018180958

[12] Duchesne, S., Caroli, A., Geroldi, C., Collins, D.L., Frisoni, G.B., 2009. Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. NeuroImage 47, 1363–1370. https://doi.org/10.1016/j.neuroimage.2009.04.023

[13] Eriksson, D., Bindel, D., Shoemaker, C., 2015. Surrogate optimization toolbox (pysot).

[14] Esteban, C., Staeck, O., Baier, S., Yang, Y., Tresp, V., 2016. Predicting clinical events by combining static and dynamic information using recurrent neural networks, in: 2016 IEEE International Conference on Healthcare Informatics (ICHI). Ieee, pp. 93–101.

[15] Ferreira, D., Verhagen, C., Hernández-Cabrera, J.A., Cavallin, L., Guo, C.-J., Ekman, U., Muehlboeck, J.-S., Simmons, A., Barroso, J., Wahlund, L.-O., Westman, E., 2017.