Predicting alzheimer's disease progression from critical biomarker data using recurrent neural networks
BMIN 521: Project Report

Introduction:

Alzheimer's disease (AD) is characterized by a loss of memory and cognitive function in patients. The progression of the disease and symptoms is gradual, in that it starts slowly with mild cognitive impairment, followed by more severe dementia following disease progression [1]. In patients with AD, there is shrinkage of the cerebral cortex, and an enlargement of ventricles, as evidenced by studies involving MRI and PET scans [2]. However, the exact causes of AD are not that well understood. Predicting the onset of AD is crucial in order to predict the onset of AD in patients so as to help focus treatments to the right cohorts of people. This presents opportunities for implementing machine learning algorithms to learn from data composed of critical biomarkers and accurately predict disease onset in individuals and screen them for further clinical trials. Consequently, the aim of this project is to utilize clinical biomarker data from MRI and PET scans, cognitive tests and risk factors to predict the clinical status of individuals using artificial recurrent neural networks.

Materials and Methods:

A. Biomedical dataset:

All data will be acquired from the TADPOLE challenge, an initiative instituted to predict the longitudinal evolution of AD [3]. The TADPOLE challenge provides a list of 1677 individuals that were recruited to the Alzheimer's Disease Neuroimaging Initiative (ADNI). These individuals have all provided data within earlier ADNI studies and have also agreed to provide follow-up data, also referred to as 'rollovers'. The objective of ADNI was to forecast features of each rollover individual at the time of future data provision. The measures that is of interest for this project are:

1. Clinical status: (Normal, mild cognitive impairment or AD).
2. Ventricle volume relative to the intracranial volume

TADPOLE datasets, in addition to the clinical status and the ventricle volume, contain data composed of several biomarkers, but for the purposes of this project, the biomarkers considered are restricted to those that were suggested by TADPOLE challenge organizers. These are:

1. MRI measures: Hippocampus, WholeBrain, Entorhinal, MidTemp
2. Cognitive tests: CDRSB, ADAS11, ADAS13, FAQ, MMSE, RAVLT_immediate, RAVLT_learning, RAVLT_forgetting
3. PET measures: FDG, AV45

B. Data Extraction:

The TADPOLE data has already been split into four data sets, namely D1, D2, D3, and D4. Dataset D1 is the standard training set, and dataset D2 is the standard prediction set to evaluate

the proposed model. Dataset D3 consists of the same set of participants as D2, but with a more limited set of features. Dataset D4 is the test data set consisting of rollover individuals used for evaluating the forecasts. These datasets were available in the form of CSV files, which were then read into Python using the Pandas framework. The datasets D1 and D2 were combined and then 80% was used as the training data set and 20% was used as the test data set for longitudinal predictions. The performance of the model was then also evaluated on the dataset D4 consisting of rollover ADNI subjects.

C. Data Imputation:

Upon close examination of the data, it was found that there exists a lot of missing data points for most subjects. Missing data is always an issue when dealing with longitudinal data sets, and therefore data imputation is important in order to address the issue of missing data. For the purposes of this project, a linear interpolation scheme was used to impute data at missing time points. For example, if data for feature A was available at the first and seventh month for a particular subject, then the values of feature A for the intermediate months was simply imputed using a linear fit between months 1 and 7. If data was missing at the first month itself, then a population average was used to impute the value of feature A for this month.

D. Machine Learning Model:

Recurrent neural networks (RNNs) are a popular class of supervised machine learning techniques used for both time dependent classification and regression tasks, such as the problems that are presented by the TADPOLE challenge. The biomarker data discussed in the previous section will be used as features to train an RNN model. The RNN architecture was adapted from a minimal RNN model proposed for modeling longitudinal disease progression in the literature [4,5]. Let xt denote all variables observed at time t, comprising the diagnosis st and remaining continuous variables gt. diagnosis was represented using one-hot encoding. Then, the model prediction of observed variables at the next time point was as follows. All the results presented in this report were trained using this RNN model with a hidden layer size of 200 neurons.

$$x_t = [s_t,\ g_t]$$
$$u_t = tanh(W_x x_t)$$
$$f_t = tanh(U_h h_{t-1} + W_u u_t)$$
$$h_t = f_t \odot h_{t-1} + (1 - f_t) \odot u_t$$
$$s_{t+1} = softmax(W_s h_t)$$
$$g_{t+1} = W_g h_t + g_t$$

E. Model Training:

A composite loss function was used to train the model, which was the sum of the cross-entropy loss associated with the classification task, and the mean absolute error associated with the regression task.

$$L = \sum_{t>1} \left(\sum_{j=1}^{3} - s_t^j \log \widehat{s}_t^j + \frac{1}{n_{feat}} \sum_{j=1}^{n_{feat}} |g_t^j - \widehat{g}_t^j| \right)$$

The RNN model was trained using full batch gradient descent on the loss function using the Adam optimizer, using a variable learning rate starting initially from 0.01 to a rate of 0.0001. All training was carried out using the popular Python-based machine learning library PyTorch.

F. Model Evaluation:

Finally, the model was evaluated using the evaluation metrics used in the TADPOLE challenge. For the classification tasks, the multiclass area under the curve (mAUC) and the balanced classification accuracy (BCA) were used to evaluate the classification accuracies. For the regression task of predicting the ventricular volume, the mean absolute error (MAE) between the true and predicted values was used as the metric for accuracy.

Results:

A. Model Training:

As discussed earlier, the model training was carried out using full-batch gradient descent using an Adam optimizer on the composite loss function at each epoch. The results of the training is shown as a plot of the loss function versus the number of epochs in Figure 1 for the first 100 training epochs.

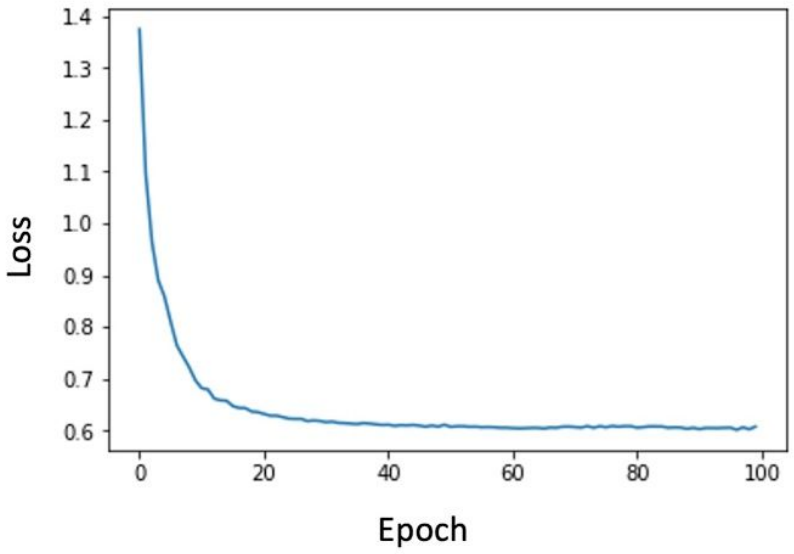

Figure 1: The compound loss versus the number of training iterations during RNN training

The model takes ~3 minutes of training time per epoch, resulting in ~5 hours of training time for 100 epochs on Google Colaboratory with GPU capability enabled. Five different neural networks were trained using the same architectures and activation functions, but for different 80-20 splits of the dataset into training and testing sets, resulting in a net training time of ~1 day.

B. Model Performance on the Test Data:

Upon training, the model was evaluated on the test data set. In order to evaluate the performance on the test subjects, the data in the first half of the time points was assumed to be known for each

subject, and then the data for the next half up until the final time point was evaluated using the model predictions. The model predictions were then used to evaluate the model based on the TADPOLE challenge evaluation metrics and are reported in Table 2.

Evaluation Metric	Value
Clinical diagnosis mAUC	0.923 ± 0.011
Clinical diagnosis BCA	0.823 ± 0.025
Ventricles MAE	0.0012 ± 0.00015

Table 2: Performance of the RNN model averaged across 6 different test sets

It is observed that the model has a really good prediction accuracy on the test data set, as evidenced by the clinical diagnosis mAUC and BCA values being very close to 1, and a really low ventricles MAE. This suggests that the model does not overfit, and performs really good short-term prediction (1-5 years) of disease progression.

C. Model Performance on the Rollover Data D4:

Considering the excellent performance of the model on both the classification and the regression test accuracies of the model on the test data set, the model was now applied to the rollover ADNI data set D4, where data on rollover subjects was collected ~10 years after the first data collection.

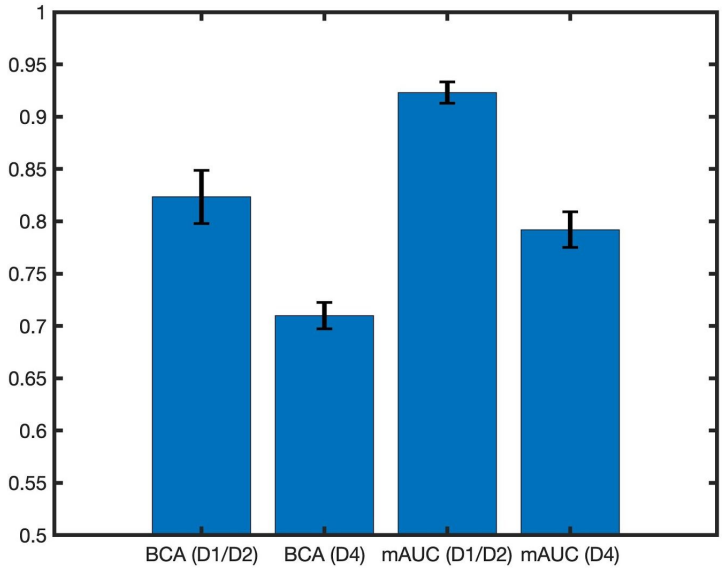

Figure 2: Performance of the model in predicting clinical diagnosis on the test data set and the rollover data set D4. A lower BCA and mAUC indicate better performance.

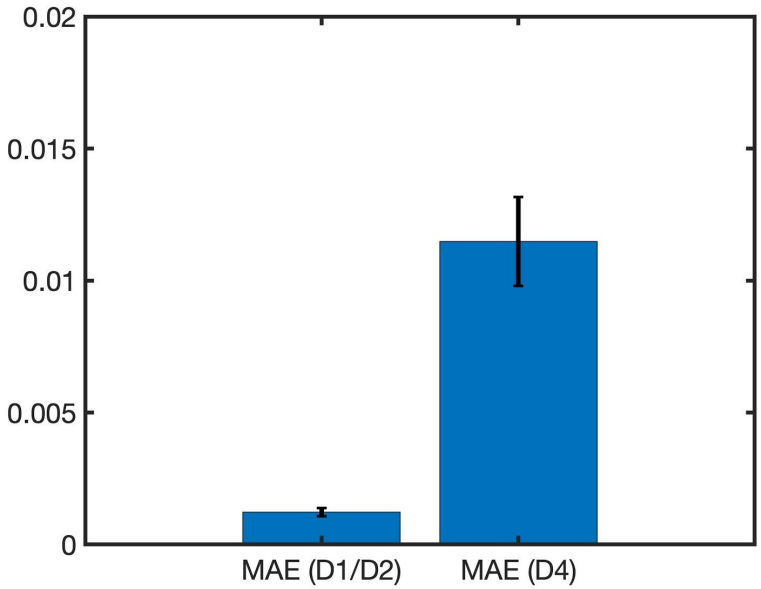

Figure 3: Performance of the model in predicting the ventricular volume relative to the intracranial volume. A higher MAE indicates better performance.

The prediction accuracies of the model on dataset D4 was compared with the test accuracy of the model in the previous section. These observations are reported in Figure 2 for the classification tasks and in Figure 3 for the regression task. It is observed that not surprisingly, the RNN model predictions become less accurate later in the future. This is a trend seen with both the classification accuracy, which drops from an mAUC of ~0.923 to a value of ~0.732, and the MAE for the ventricular volume which increases from an MAE of ~0.001 to an MAE of ~0.01.

Discussion:

In summary, an RNN model was developed to predict the longitudinal AD progression in subjects, given clinical biomarker data, cognitive test scores, and clinical diagnosis. The RNN model that was proposed seemed to provide excellent accuracies for near-term prediction within ~5 years. However, it is noted that any effective AD dementia treatment probably has to begin early in the disease process, potentially at least a decade before the emergence of behavioral symptoms. However, the model prediction performance of clinical diagnosis dropped from an mAUC of 0.923 in the first ~5 years to an mAUC of 0.732 for rollover subjects in D4, while ventricular volume MAE increased from 0.001 in the first ~5 years to 0.01 for rollover subjects D4. Thus, significant improvement is needed for clinical utility. A possible approach could be to append the feature set with data from electronic health records of subjects, which may tend to be more easily accessible than the features considered in this study, and thereby yield better prediction accuracies.

References:

1. Burns A, Iliffe S. Dementia. Bmj. 2009 Feb 5;338.

2. Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL, Fischl B. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain. 2009 Aug 1;132(8):2048-57.

3. Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, Barkhof F, Fox NC, Klein S, Alexander DC. Tadpole challenge: Prediction of longitudinal evolution in Alzheimer's disease. arXiv preprint arXiv:1805.03909. 2018 May 10.

4. Chen M. MinimalRNN: Toward more interpretable and trainable recurrent neural networks. arXiv preprint arXiv:1711.06788. 2017 Nov 18.

5. Nguyen M, He T, An L, Alexander DC, Feng J, Yeo BT, Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using deep recurrent neural networks. NeuroImage. 2020 Nov 15;222:117203.