UCSF Department of Radiology & Biomedical Imaging

Berkeley UNIVERSITY OF CALIFORNIA MSSE MOLECULAR SCIENCE & SOFTWARE ENGINEERING

Project Report: Model Details and Analysis

April 26th, 2024

Project Title

Predicting Longitudinal Cognition in Alzheimer's Disease using Deep Learning

Student Name

Amar Jilani (MSSE)

Organization

Brain Networks Laboratory (Raj Lab); University of California, San Francisco (UCSF)

Collaborators

Ashish Raj (UCSF), Daren Ma (UCSF)

Abstract

This project introduces a novel approach to Alzheimer's disease prognosis by developing a Long Short-Term Memory (LSTM) model to longitudinally predict protein progression and baseline cognitive scores using neuroimaging data. Addressing the limitations of current diagnostic methods, which are largely retrospective and invasive, this project leverages PET-derived data for its non-invasive nature and detailed biomarker insights. Diverging from traditional classification-focused machine learning research, this project emphasizes regression tasks to forecast clinical measures critical to Alzheimer's progression and treatment. By targeting the prediction of cognitive decline and biomarker evolution, the model aims to facilitate earlier diagnosis, enable personalized treatment strategies, and contribute to the development of new therapeutic interventions. This work responds to the challenges of existing machine learning tools in handling the complex dimensionality mismatch of neuroimaging and biomarker data, emphasizing the need for a custom, hybrid deep learning approach to improve predictive accuracy in Alzheimer's disease progression. This project yielded a model that demonstrates promising results in accurately forecasting the progression of tau protein concentrations across the brain and cognitive metrics.

Relevant Topics

Neuroscience, Neuroimaging, Deep Learning, Cognition Prediction

Table of Contents

Project Summary

This project introduces a novel approach to Alzheimer's disease prognosis by developing a Long Short-Term Memory (LSTM) model to longitudinally predict cognitive scores and tau protein progression using neuroimaging data. Addressing the limitations of current diagnostic methods, which are largely retrospective and invasive, this project leverages regional biomarker data derived from neuroimages, particularly positron emission tomography (PET). PET is valued for its non-invasive nature and ability to provide extensive insights on biomarker concentrations within the brain [1]. Diverging from traditional classification-focused machine learning research, this project emphasizes regression tasks to forecast clinical measures critical to Alzheimer's progression and treatment. By targeting the prediction of cognitive decline and biomarker evolution, the model aims to facilitate earlier diagnosis, enable personalized treatment strategies, and contribute to the development of new therapeutic interventions. This work responds to the challenges of existing machine learning tools in handling the complex dimensionality mismatch of neuroimaging and biomarker data, emphasizing the need for a custom, hybrid deep learning approach to improve predictive accuracy in Alzheimer's disease progression.

This project involves the implementation of an LSTM model to forecast the progression of cognitive scores over time. An LSTM is a type of recurrent neural network (RNN) that is adept at retaining long-term information, making it highly effective at dealing with sequential data [2]. This model will be trained on all available multimodal biomarkers such as: demographics, regional biomarkers, cognition scores, and diagnosis. The model will utilize a rich time-series of biomarker concentrations at different brain regions. LSTMs are particularly suitable for this task due to the inherently sequential and time-dependent nature of disease progression data. LSTMs may also be able to capture the non-linear nature of cognitive decline in Alzheimer’s. This model will incorporate an autoencoder for dimensionality reduction, allowing the LSTM to handle high dimensional sequential data.

This project is specifically focused on harnessing a time-series dataset of regional biomarkers from 194 patients, capturing multiple timepoints per patient. This data was developed through the processing of PET data through the extended network diffusion model (eNDM) [3]. The regional biomarker vector contains relative levels of relevant Alzheimer’s disease biomarkers, such as tau and amyloid beta, at 86 different brain regions. Leading theories in AD research attribute the spread and increased concentration of these biomarkers to the development and progression of Alzheimer’s disease [4]. Understanding the spread of these biomarkers, as well as predicting their impact on cognition, is crucial for early diagnosis and the development of targeted therapeutic interventions.

Our goal was to develop an LSTM deep learning framework to meticulously analyze rich time-series data, with the aim of accurately predicting Alzheimer's related biomarkers and cognitive scores. This project has culminated in a detailed report that assesses the model's accuracy and efficacy. Our comprehensive analysis demonstrates that our LSTM-based tool performs well compared to a standard fully connected neural network model. This comparison underscores and validates the effectiveness of the LSTM framework in analyzing and predicting disease progression.

Project Methodology

Overview

The primary approach for this project involved the design, training, and iterative refinement of a series of LSTM models. Each version provided enhanced capabilities or improved performance, informed by ongoing discussions and a deepening understanding of the project. The initial implementation was based on standard LSTM architecture, which served as a foundation for subsequent models.

A general overview of each version of the model:

Version	Model	Rationale
1	LSTM with direct MRI input	- This initial implementation was part of the exploratory phase, where we investigated potential strategies based on the project goal.
2	LSTM with MRI preprocessing via CNN	- The first version turned out to be infeasible due to the large dimensionality of imaging data, which is generally difficult for LSTMs to handle. We considered the possibility of utilizing a CNN to preprocess the imaging data into a smaller representation, implementing a CNN from scratch as well as looking into pre-trained models.
3	LSTM with regional biomarker (RB) input	- Due to the difficulty and computational extensiveness of dealing with imaging data, the focus shifted from image processing to working with data that has been derived from tau PET scans. Attempting to create a new image processing pipeline would be redundant as the lab already developed tools to generate regional biomarkers from MRI images [5].
4	LSTM with RB input capable of scalar prediction	- Version #3 was successful in predicting regional tau at the next timestep but did not predict cognition. Therefore, the purpose of this version was to modify the architecture to generate scalar prediction, as well as predicting tau at the next timepoint.
5	LSTM with reduced dimensionality RB input (Laplacian)	- The Laplacian of the connectome was utilized to reduce the dimensionality of the input. Reduced dimensional data can train faster and is generally less likely to overfit. Data was reduced from 86 features to 8 features.
6	LSTM with reduced dimensionality RB input (Autoencoder)	- An alternative approach with a similar goal as version #5, however this version utilized an autoencoder model to reduce the dimensionality instead. This process involved training an

			autoencoder to produce an 8-dimension latent representation of the regional biomarkers. Development of this model was discontinued due to the Laplacian version performing significantly better.
7	**LSTM with reduced dimensionality RB input and static demographic inputs (Laplacian)**	-	The final version of this model is to integrate static demographic data with the time-series regional biomarkers. This is done by concatenating each timepoint with the demographic data when passing it through the LSTM model. This is the latest, well-performing version of the model.

Table 1: General overview of model iterative evolution. This table outlines each of the major designs that were developed through the course of the project. Version #7 (**bold**) represents the most up-to-date and well-performed model.

A high-level visualization of the final model architecture (version #7) is displayed in Figure 1. All references to the 'LSTM model' below refer to this model. The key components of this project's implementation are the regional biomarker (RB) input, the static demographics data, the LSTM model, and the model outputs – the projected regional biomarkers at the next timestep (RB_{t+1}), and cognitive scores.

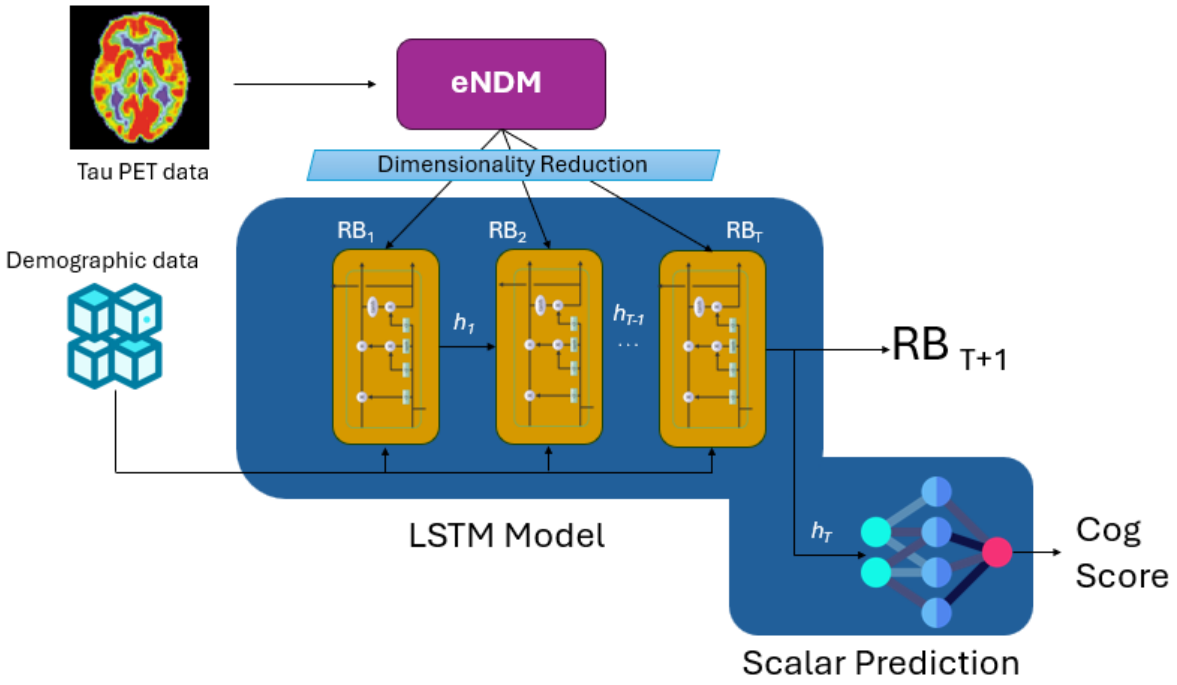

Figure 1: High-level visualization of model architecture. This project is focused on the development of the LSTM model portion (dark green) of the above diagram. The eNDM model [3] generates regional biomarker (RB) data at multiple timepoints from baseline tau PET data. The reduced RB timeseries, along with static demographic data, is fed into the LSTM as input to predict the next timestep. The LSTM's final hidden state is fed into a dense network to predict cognitive score.

Software Components and Dependencies

The data-preprocessing was conducted using of Python's Pandas and NumPy libraries. These tools allowed us to build a streamlined pipeline to process and transform the data so that it was optimally formatted to be input into the model. The scikit-learn library was also utilized to normalize the data and encode categorical features. This process was carried out in Jupyter Notebooks.

The machine learning model was developed primarily through the PyTorch python library. This library provides the tools and features to develop and train the LSTM models. Specifically, the built-in LSTM module was used as a foundation for the LSTM model. The models were developed in a Python file as modular classes, which allowed for them to be imported into any training script.

The training scripts also made use of scikit-learn and PyTorch's data-processing tools, which allowed for the data to split into training, validation, and testing datasets. The training scripts also utilized PyTorch's MSE loss function and Adam optimizer. Visualizations were done using Python's matplotlib library. The training scripts were run in Jupyter Notebooks.

Dependency	**Version**
Pandas	2.0.3
NumPy	1.24.4
PyTorch	2.1.0
Matplotlib	3.7.1
Scikit-learn	1.3.0

Table 2: Software dependencies. The table above describes the main Python libraries used during the development of this project, along with the version used.

Data Overview

The initial aim of this project was to utilize MRI scans, which provide structural brain images, to predict future regional biomarker levels and cognitive scores. This data was obtained from the Alzheimer's Disease Neuroimaging Initiative. As the project advanced, we encountered challenges related to the availability of computational resources and the inherent constraints of obtaining longitudinal imaging data. Therefore, a pivot was made to instead focus on utilizing regional biomarker data directly, rather than developing an image processing pipeline. Furthermore, this reduced redundancy as the lab has previously developed a UNet model capable of generating regional biomarkers from MRI data [5]. This allows the model to still be applicable to MRI-derived data in the future.

For this project, the regional biomarker timeseries was derived from the extended network diffusion model (eNDM) [3]. This model utilizes tau PET imaging data to generate regional biomarker timeseries. The tau PET data is obtained from ADNI, along with demographics information for each patient. This neuroimaging-derived data was used in training the LSTM model. The dataset used consists of 196 patients, each with timeseries tau and amyloid beta values over 86 brain regions, cognition scores, genetic information, and general demographic

data (i.e. gender, age, etc.). The two cognitive measures central to this project are the Alzheimer's Disease Assessment Scale (ADAS11) and the Mini-Mental State Examination (MMSE).

Data Pre-processing

For the purposes of this project, we solely focused on tau concentrations, however the same methodology could also apply to amyloid-beta concentrations. The features used for the data consist of the regional tau timeseries, age, gender, years of education, marriage status, and presence of the APOE4 gene. After removing rows containing null values, we were left with 194 unique patients. This final set was split into a training and testing set, which both follow the same pre-processing procedure described below.

Each patient had around 90 timepoints generated from the eNDM. To simulate real world conditions more closely, where patients generally have much fewer timepoints (i.e. brain scans), we divided the timeseries into sequences of 10. This process was done carefully to ensure no overlap or repeated values, thus protecting against data leakage.

To enhance the LSTM's performance and increase training speed, the original 86-dimensional regional tau data was reduced to an 8-dimensional representation. This dimensionality reduction was achieved using the Laplacian matrix of a standard brain connectome, which is a graphical representation of the brain that operates under the assumption that brain connectivity is uniform across individuals. By representing the brain as a graph, the Laplacian matrix was constructed, and the top 8 eigenvectors were employed to transform the data into a more manageable, lower-dimensional form.

Further processing included one-hot encoding the categorical values (marriage status and gender), while the numerical data was normalized against all patients. This led to static data with the dimensionality of 13. The training data was further split up into training and validation sets during model training. Furthermore, k-fold cross validation was utilized to reduce model bias and improve robustness.

Model Details

The final model consists of LSTM components, as well as an additional artificial neural network (ANN) for scalar predictions. The model takes in the time series data as input, along with static demographic data. The model then generates two outputs, the next regional biomarker timestep (i.e. the 11^{th} timepoint), as well as a 2x1 vector of cognition scores, consisting of both ADAS11 and MMSE scores. This project implemented the model with 8-dimensional input, 128-dimension hidden state, 1-dimensional output for the LSTM portion. The model additionally contains the ANN architecture of [128, 64, 32, 16], producing a 2-dimensional output. The input to this model is the final hidden state of the LSTM, as shown in Figure 1.

The model was trained for 500 epochs over 5 folds. The mean-squared error (MSE) loss function was used for both biomarker prediction and cognitive score prediction. The losses were calculated independently for regional biomarker prediction and cognitive scores, and then

combined for optimization. The model used the Adam optimizer with a learning rate of 0.001 and a weight decay of 1e-5.

A simple ANN was also developed as a benchmark to compare cognitive score prediction. This ANN had the layer architecture of [128, 64, 32] and used a similar training process. This model was trained on the same patients as the LSTM model (196 patients), as well as a larger dataset consisting of 819 unique patients.

Model Results

The model has demonstrated impressive performance, particularly in terms of the loss curves depicted in Figure 2. The observed low loss indicates that the model is proficient in learning the underlying patterns in the data.

Research on forecasting the progression of Alzheimer's disease, particularly through the application of LSTMs, remains extremely limited, leaving very few for benchmarks or comparisons. Furthermore, there are very few models that are attempting to predict cognition from neuroimaging. Therefore, the simple benchmark model described above was used as a comparison for the performance of the LSTM model.

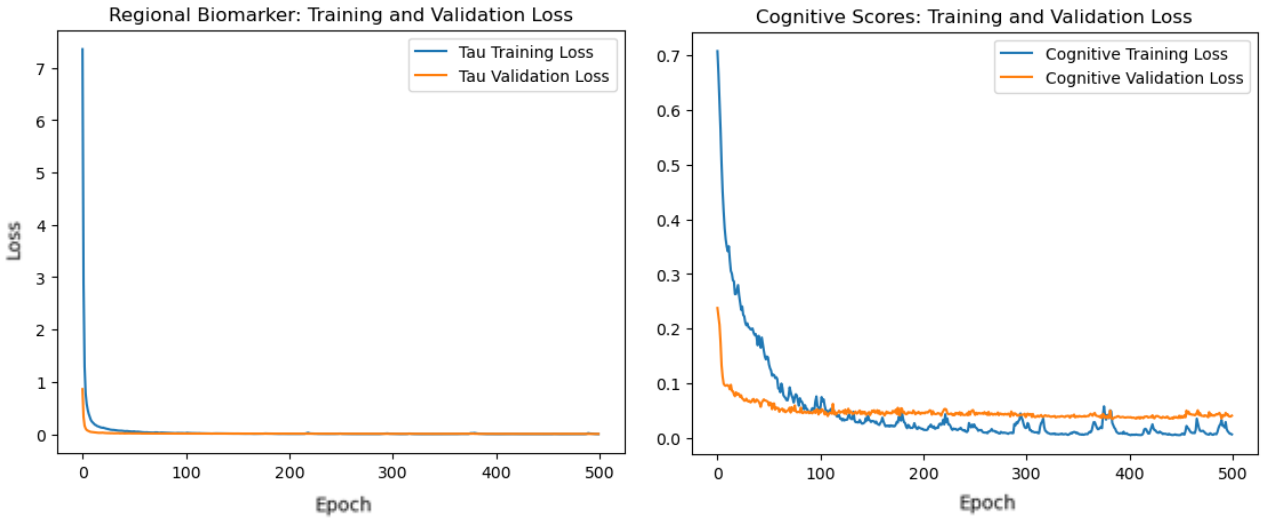

Figure 2: Training and validation loss curves for both outputs of the LSTM model. The plot to the left depicts the LSTM's prediction of the next biomarker time-step given a sequence of biomarker vectors. The plot to the right shows the MSE loss of the model's deep network that allows for prediction of a scalar, which in this case is cognitive score.

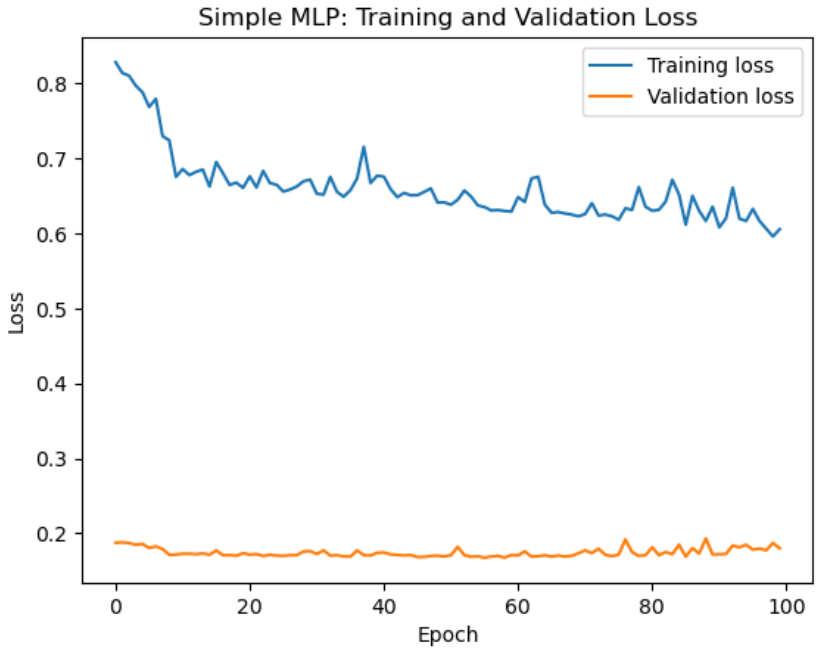

Figure 3: Training and validation loss of simple benchmark MLP model for cognitive score prediction. The plot above demonstrates the training curves for a simple multilayer perceptron (or ANN) for predicting cognitive scores (both MMSE and ADAS11).

As demonstrated by Figure 2, the LSTM model exhibits effective learning ability, evidenced by the low training and validation losses over time. This indicates that the model is fitting well to the training data, as well as generalizing effectively to unseen data. The validation loss for cognitive score prediction does tend to be slightly higher than the training loss, whereas this is not seen for the biomarker prediction.

In contrast to the learning capabilities of the LSTM model, the simple benchmark model shows much higher training losses when predicting cognitive scores (Figure 3). This suggests that the model is not able to capture the underlying complexities of the data, in other words, the model is underfitting. Interestingly, the validation loss is very low, comparable to the LSTM model, despite the training loss being significantly higher.

To evaluate the performance of the model, we utilized the test set and performed Pearson correlation analysis between the true and predicted values. This was done for both the regional tau predictions (Figure 4), as well as for both ADAS11 and MMSE predictions (Figure 5). The performance of the cognitive score prediction can be compared to the benchmark results shown in Figures 6 and 7.

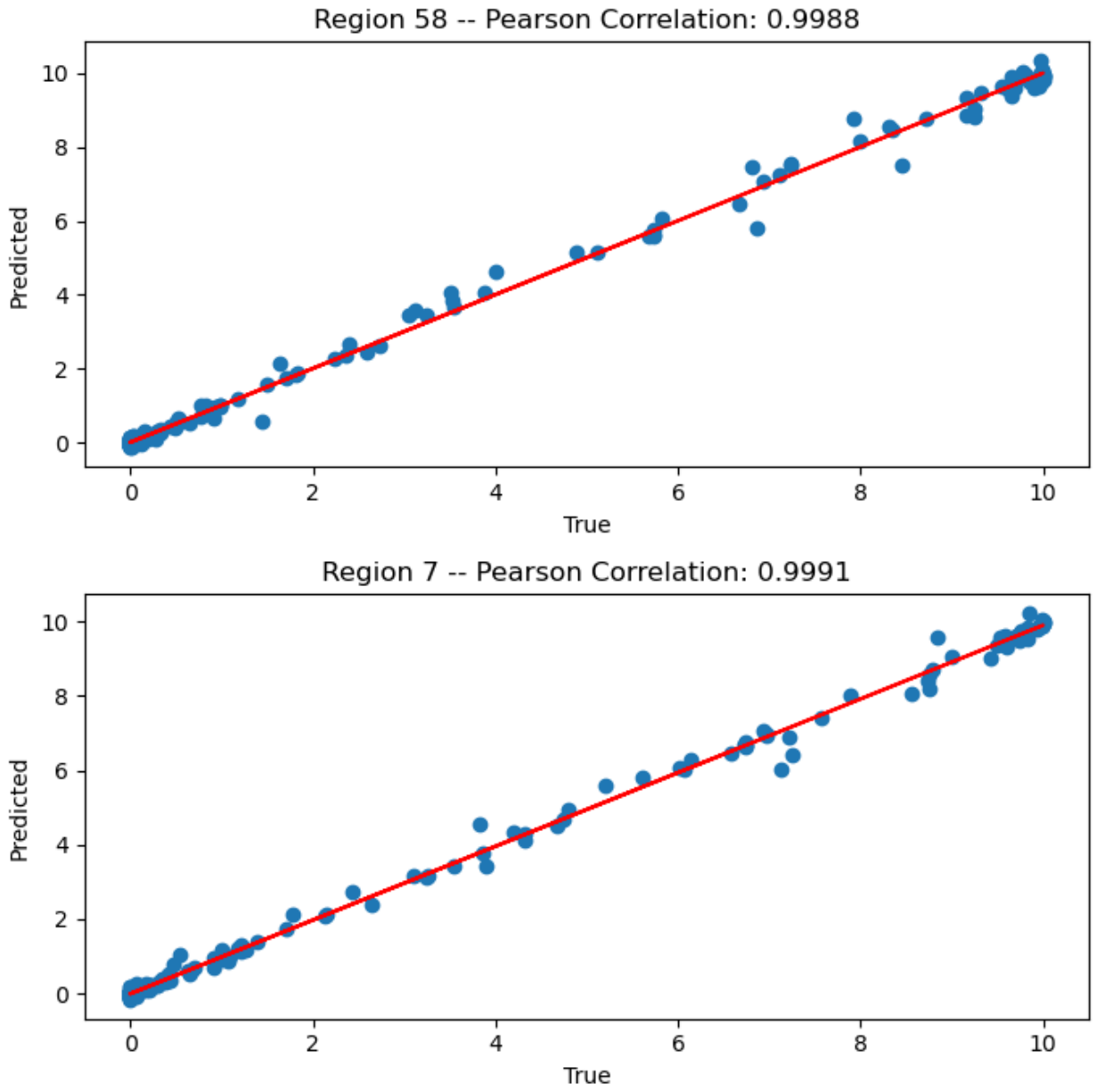

Figure 4: Prediction vs. ground truth plot for regional tau prediction in two brain regions. The model was evaluated on the test set to predict the next timepoint of the regional tau biomarkers in 86 different brain regions. The comparison between the true and predicted values are shown in this plot. The red line represents the line of best fit. These plots specifically demonstrate two sample regions, 58 (rostral anterior cingulate cortex in the left hemisphere) and 7 (inferior temporal cortex in the right hemisphere).

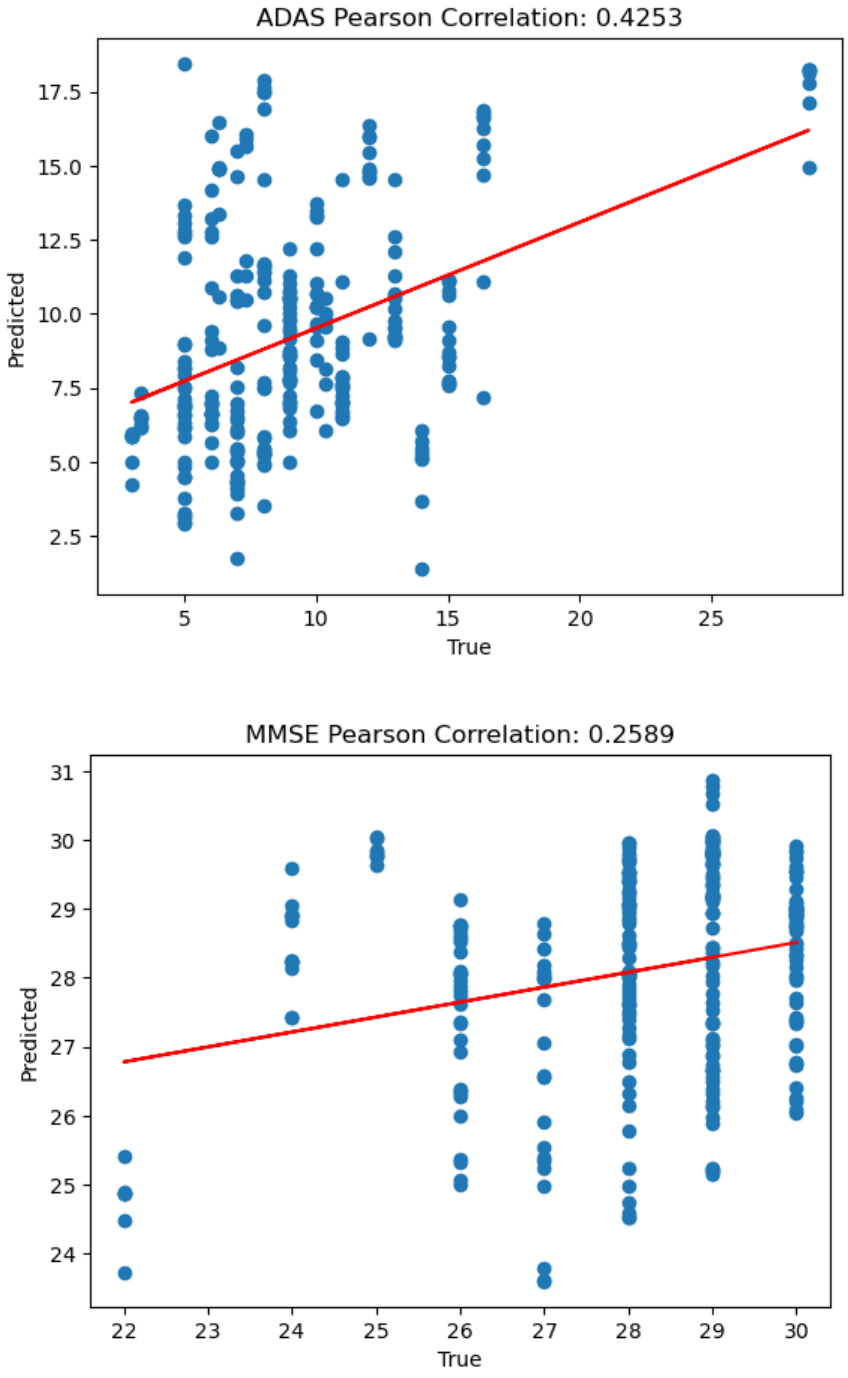

Figure 5: Prediction vs. ground truth for ADAS11 and MMSE predictions. The model was evaluated on the test set to predict two cognitive metrics, ADAS11 and MMSE. The plots above show the correlation between the prediction and the true values.

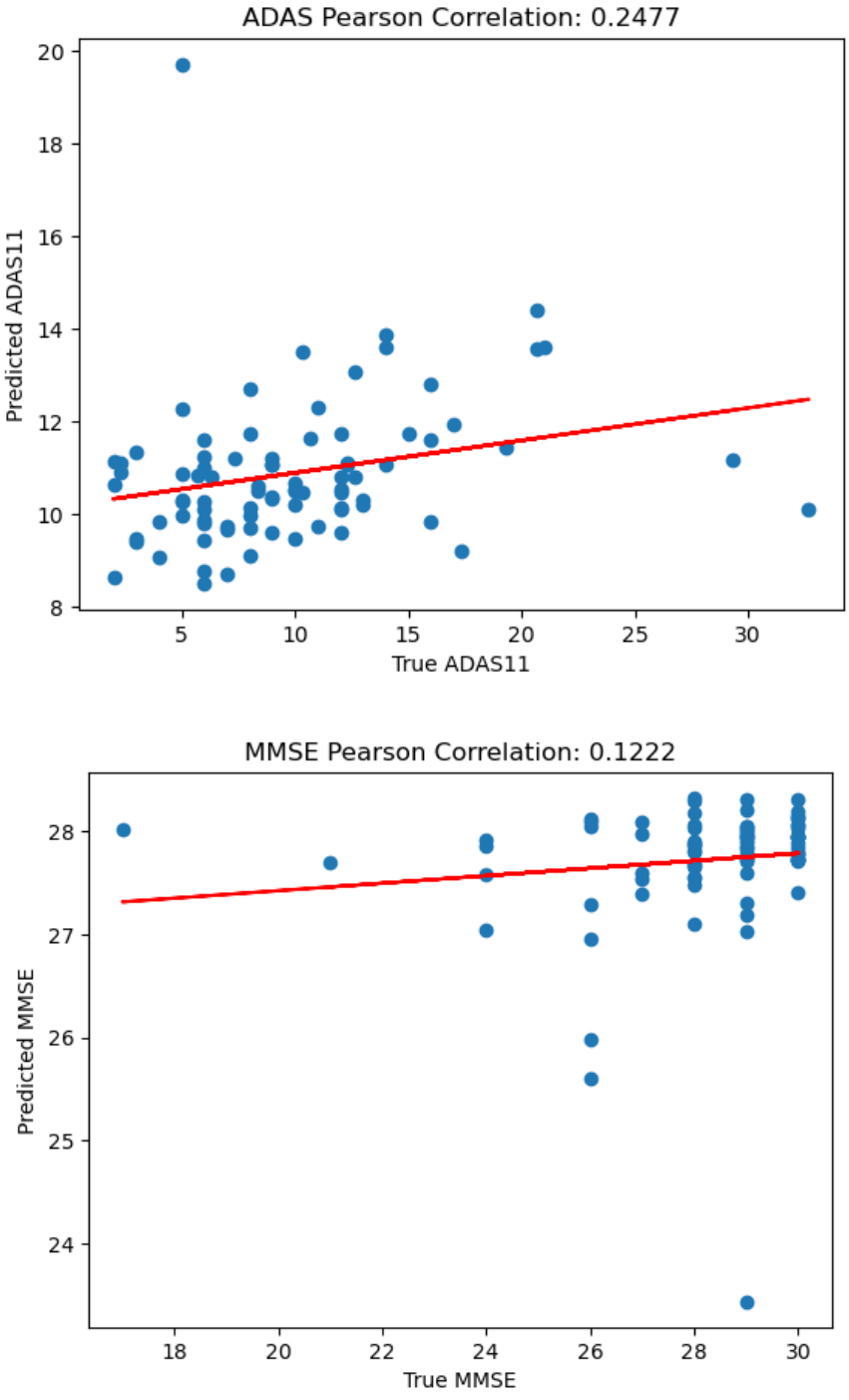

Figure 6: Benchmark – Prediction vs. ground truth plot for ADAS11 and MMSE predictions. The benchmark model was trained and evaluated on the original dataset (196 patients). The plots above show the correlation between the prediction and the true values.

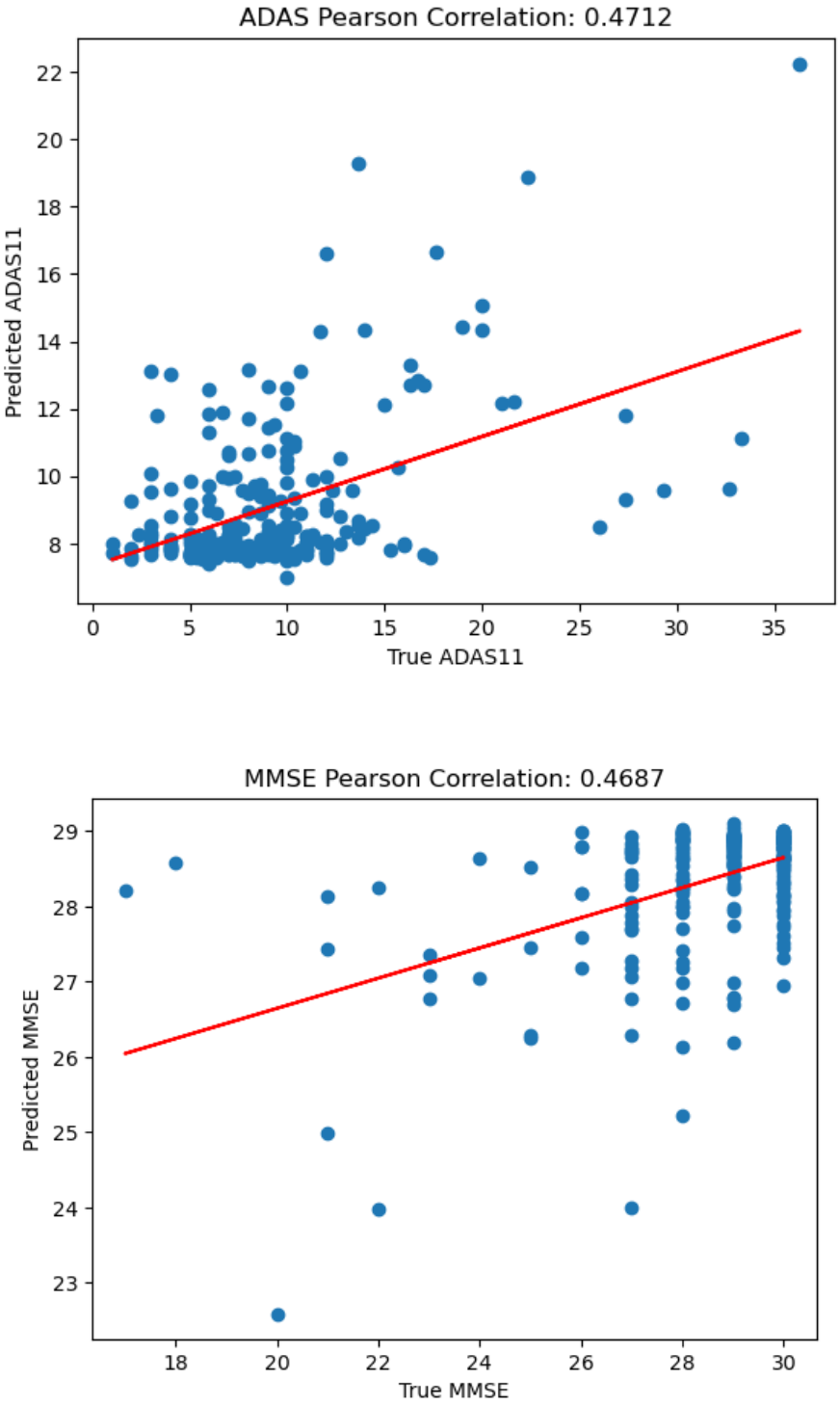

Figure 7: Benchmark (large dataset) – Prediction vs. ground truth plot for ADAS11 and MMSE predictions. The benchmark model was trained and evaluated on the original dataset (819 patients). The plots above show the correlation between the prediction and the true values.

Figures 4 and 5 above aim to demonstrate the capabilities of the LSTM model in predicting tau progression, as well as predicting cognitive scores. Figure 4 demonstrates the model's strong ability to accurately predict the next timepoint of the regional tau concentrations. Figure 4 arbitrarily chose two brain regions out of 86, however similar results were seen across all brain regions, with all brain regions exhibiting Pearson correlation coefficient of over 0.98.

In contrast, the model exhibited some limitations in accurately predicting cognitive score, which is the much more complex task. The model generally performed better when predicting ADAS11, with a Pearson correlation coefficient of 0.43, whereas the MMSE correlation coefficient is 0.26. These results demonstrate that there is a weak positive correlation between the predictions and the true values. To compare the effectiveness of this architecture, the results can be compared to the results of the benchmark model. The results demonstrate the performance of the simple ANN trained and evaluated on both the on the same dataset (196 patients), and an extended dataset (819 patients).

As shown by Figure 6, the benchmark performs relatively worse compared to the LSTM model when trained and tested on the same dataset, attaining significantly lower Pearson correlation coefficients of 0.25 and 0.12 for ADAS11 and MMSE, respectively. However, it is important to note that when trained on the larger dataset containing 819 patients, the simple model's predictive performance surpassed the LSTM's performance, as shown in Figure 7. Training on the larger dataset leads to Pearson correlation coefficients of 0.47 for both ADAS11 and MMSE when using the simple ANN model. This improvement in performance when using the larger dataset emphasizes the important that data volume plays in enhancing predictive accuracy. Timeseries data was not available for the larger dataset, therefore we could not compare the effectiveness of the LSTM after being trained on more data.

Verification and Validation

The goal of this project was to predict the progression of Alzheimer's disease longitudinally, utilizing neuroimaging data. Throughout the development phase, we successfully designed and trained a model capable of forecasting changes in biomarker concentrations within specific brain regions over time, leveraging tau PET-derived data. Additionally, this model is moderately capable of predicting scalar values that correspond to the clinical manifestations of Alzheimer's disease, such as cognitive scores. The model clearly aligns with the initial goals set out for the project. Despite minor modifications to the model's functionality, it continues to satisfy the project's requirements effectively, as it can predict regional tau and baseline cognition.

Figure 2 demonstrates the proficiency with which the LSTM model was able to learn from the timeseries data. This project was focused on longitudinal data analysis, which is pivotal for studying disease progression. LSTMs have been specifically designed to extract meaningful features from longitudinal data and make meaningful predictions over time. The loss curves (Figure 2) show that model was able to learn extremely well from the timeseries data, emphasizing its suitability for this project. The model's strong performance for predicting the next timestep of regional tau concentrations across the brain validates that this architecture was the right choice for addressing this problem. When trained on the same data, the LSTM also outperforms a simple ANN for predicting cognitive scores. However, when trained on the larger

dataset, the ANN was able to outperform the LSTM. Therefore, we cannot determine with confidence whether an LSTM is necessarily the best architecture for predicting cognition. It is also important to acknowledge that alternative architectures may also offer promising results for timeseries analysis and predictions, such as transformer-based approaches or gated recurrent units.

Next Steps

The software developed for this project is presently hosted on a private GitHub repository, accessible exclusively to members of the Raj Lab team. The parameters for the best performances (i.e. lowest validation losses), along with the datasets are also made available on GitHub and can be loaded into the model using built-in PyTorch tools.

This project demonstrates the potential of combining traditional mathematical modeling with advanced deep learning methods for disease prediction. The lab aims to combine deep learning techniques with their established models, opening new avenues for understanding Alzheimer's disease progression. The model presented in this paper can be used as a foundation for future models, or its performance can be compared to alternative methods to determine the most optimal architecture. The model's predictive performance can also be evaluated on more unseen data to further analyze its capabilities.

There are several potential improvements for the model. The existing model can be improved through hyper-parameter tuning, which could increase learning performance, thus improving predictive power and generalizability. The model can also be trained on a larger dataset, which would also improve generalizability and predictive capabilities, as seen with the benchmark model. Similarly, the model can also be trained on data with more features, including more genetic data, or additional imaging data, such as MRI.

This model is capable of accurately predicting the future time-points of tau protein across 86 brain regions. Furthermore, it can also currently be used for the prediction of cognitive score based on data derived from neuroimaging. While the model is in its initial stages and requires further refinement for real-world application, it clearly highlights the significant potential of deep learning tools in advancing disease prognosis. Real world applications of this model involve rapid deduction of cognitive scores of patients based on brain scans. Cognitive testing often involves extensive administration time, excluding the waiting periods for appointments and additional administrative delays. This model lays the groundwork for the ability to instantly predict accurate cognitive scores based on a brain scan, saving significant time for both health practitioners and patients. Moreover, forecasting biomarker concentrations over time enhances our understanding of disease progression, facilitating more tailored planning and personalized treatment based on the projected progression rate and severity. Gaining insights into the progression of biomarkers across various brain regions and their correlation with cognitive scores can elucidate the complex nature of Alzheimer's pathology. The ability to accurately model biomarker progression can also help with early detection of the disease. This increased understanding can also help in the development of more effective treatments, pinpointing the critical brain areas to target for slowing disease progression and mitigating cognitive decline.

References

[1] G. Cassinelli Petersen et al., “Overview of tau pet molecular imaging,” Current Opinion in Neurology, vol. 35, no. 2, pp. 230–239, Feb. 2022. doi:10.1097/wco.0000000000001035

[2] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A search space odyssey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017. doi:10.1109/tnnls.2016.2582924

[3] C. Anand, J. Torok, F. Abdelnour, P. D. Maia, and A. Raj, Selective vulnerability and resilience to alzheimer’s disease tauopathy as a function of genes and the connectome, Mar. 2024. doi:10.1101/2024.03.04.583403

[4]W. Gulisano et al., “Role of amyloid-β and tau proteins in alzheimer’s disease: Confuting the amyloid cascade,” Journal of Alzheimer’s Disease, vol. 68, no. 1, pp. 415–415, Mar. 2019. doi:10.3233/jad-189015

[5] D. Ma et al., “Multi-task learning and ensemble approach to predict cognitive scores for patients with alzheimer’s disease,” Medical Image Analysis (2023), 2022. doi:10.21203/rs.3.rs-1663817/v1