Title
Applying Deep Leaming Models on the Brain Connectome for Improving Cognitive Predictions of Parkinson's Disease

Author
Nguyen, Alex

Publication Date
2018

Peer reviewed|Thesis/dissertation

Applying Deep Learning Models on the Brain Connectome for Improving Cognitive Predictions of Parkinson's Disease

by

Alex Nguyen

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Imaging

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgments

None of this work would have been possible without the knowledge and expertise of Dr. Ashish Raj and his network diffusion model. I would like to express my sincere gratitude to Dr. Pedro Maia for his guidance and mentorship, to my fellow lab member Gavin Gao for his assistance and insight, along with Sneha Pandya and Yashar Zeighami for sharing valuable data used in this work. I would also thank my committee members Dr. Christopher Hess, Dr. Pratik Mukherjee, and Dr. Duygu Tosun-Turgot for their support and constructive criticism. Finally, I would like to express my gratitude to my loving friends and family for their encouragement and support.

Applying Deep Learning Models on the Brain Connectome for Improving Cognitive Predictions of Parkinson's Disease

by

Alex Nguyen

Abstract

Parkinson's disease (PD) is one of the most prevalent progressive, neurodegenerative disorders that affects motor and cognitive function. It is characterized by tremors, rigidity and bradykinesia and eventually progresses to cognitive decline in late stages. Currently, there is no cure for PD. The standard therapy for treatment merely slows progression, to an extent, and provides temporary symptomatic relief. This is largely due to the lack of clinical biomarkers to successfully identify PD in early stages, resulting in a huge gap of knowledge surrounding the progression and disease stage. Recent advancements in MR clinical imaging have provided substantial anatomical datasets for subsets of populations affected by PD. In addition, there has been increasing focus on the implementation of artificial intelligence within the study of the brain network. In 2012, Raj et. al proposed the network diffusion model (NDM) based on the diffusion equation that provides analytical projections on baseline atrophy rate. In this study, we applied a deep learning model, autoencoder, to predict future motor and cognitive states using features acquired at baseline. We found that by implementing the autocoder with the NDM, prediction accuracy was improved when compared to using stepwise linear regression alone. This novel and innovative approach to neurodegenerative diseases, such as PD, has great potential for enhancing statistical power within the clinic, by providing clinicians to make more informed therapy decisions for PD patients. Not only does it reduce subjectivity, but it allows the clinician to assess motor and cognitive states at any given time point in the future.

Table of Contents

List of Tables

List of Figures

Table of Contents

1. Introduction

1.1 Parkinson's Disease and Neurodegeneration

Parkinson's Disease (PD) is one of the most common neurodegenerative diseases, and is marked by the progressive decay of dopaminergic neurons, the spread-like development of Lewy pathology, and the toxic misfolding of α-synuclein (α-syn) in a "prion-like" fashion. PD ultimately manifests itself into motor deficits including tremors, rigidity, and bradykinesia. Typically, PD is diagnosed through the assessment of motor deficits, which occur during later stages of the disease when 80% of the dopamine concentration in the brain has already been depleted. In addition to motor-control dysfunction, PD patients may also experience cognitive disturbances, for example, depression and anxiety[1]. Brain atrophy has been most commonly identified in the Substantia nigra pars compacta region, causing considerable reduction in striatal dopamine. This reduction is currently believed to be responsible for the previously described symptoms[2]. PD is currently an incurable condition and all current therapies focus on reducing the speed of progression and provide symptomatic relief[2].

The goal of this project is to investigate how accurately a deep-learning model (autoencoder) can predict future motor and cognitive states of a patient using only (i) features collected at baseline, and (ii) analytical predictions given by the network diffusion model developed by Raj et al. (2012). Currently, there are no clear biomarkers for PD diagnosis and prognosis. An accurate prediction of the future motor/cognitive state of a patient may reveal those with higher risk of functional decline, help the patient with life planning, and provide an individualized standard of care. There have been numerous studies on the predictive capabilities of machine learning and deep learning models for Alzheimer's disease with varying success in the recent decade[3]. Contrastingly, the few studies available for PD show several limitations such as (i) a large

variability in the results due to the ill-posed nature of the problem, (ii) a smaller sample size as compared to AD datasets, and (iii) the lack of longitudinal atrophy progression for individual patients.

The spread of Lewy pathology was recently shown to occur along the local and long-range fiber projections, thereby suggesting a process of "network spread"[4]. On this account, the structural connectome of the brain - often referred to as the brain's anatomic connectivity network or *connectome* - surfaced as a powerful tool for understanding the progression of neurological disorders since it is the natural place upon which many of the brain's complex interactions and phenomena take place. This type of spread has been mathematically presented by Raj et al. (2015) and has shown success in predicting longitudinal atrophy patterns in Alzheimer's Disease[5]. The mathematical model is based on dispersion mechanisms arbitrated and bound by the connective network of the brain, described as the "network diffusion" model.

Currently, there are only a handful of studies that have explored the NDM for PD. Remarkably, these studies demonstrate important and innovative potential for understanding disease progression of PD in order to provide more precise treatment for each individual. Therefore, we hypothesize that the implementation of the NDM along with deep learning may predict future cognitive states with higher accuracy than using machine-learning methods alone.

1.2 Neural Networks

The use of artificial intelligence has been tested and proven with relatively high specificity and sensitivity in a number of studies as a way of assisting with a clinical diagnosis[6,7,8]. These studies may also remove some subjectivity from clinician's scores as well as support the

diagnosis by providing high statistical and quantitative data. Overall, not many machine learning studies have been conducted for staging PD or to predict cognitive performance versus the amount of studies used for early diagnosis predictions[9]. To take this a step further, the implementation of a deep learning network may help detect more complex patterns in larger datasets without the need of having to selectively extract meaningful features.

Deep learning models have multiple levels of nonlinear modules that transform raw data into more abstract representations and thus, learn complex functions[10]. Usually, a large amount of well-curated and high-quality data is used as input and the model will learn on that data. Predictors are fed into the neural network and processed by activation functions in the hidden layers and finally linked to the desired output. Weights and biases are adjusted accordingly at each layer to reduce the mean square error. By employing this highly sophisticated deep neural network architecture, we aim to successfully predict future cognitive and motor decline that strongly correlate to changes in the brain network.

1.3 Cognitive and Motor Tests

Often the initial step in the clinical decision making process for PD is a chain of cognitive and behavioral assessments. For the purpose of this study, one cognitive and one motor test are predicted separately by the proposed autoencoder deep learning model.

The Montreal Cognitive Assessment (MoCA) has been shown to be more sensitive compared to the Mini Mental State Examination for determining cognitive symptoms in PD patients[11,12]. It is useful for differentiation between different cognitive stages of PD (no cognitive impairment, mild cognitive impairment, and dementia). Although results via MoCA show promise, this test comes

with some limitations. Previous studies have demonstrated that MoCA is subjective based on the clinician and the participant at the time of testing[11]. This test may also be limited by its cumulative scoring basis of subtests, where biases can occur, however, only the cumulative score is taken into consideration for the prediction model. New studies are being conducted to evaluate the current MoCA scoring with the goal of providing a new scale that better reflects the cognitive score for the person taking the test.

According to the Movement Disorder Society, the Unified Parkinson's Disorder Rating Scale Part III (MDS-UPDRS-III) was developed to provide an efficient and reliable test to evaluate the motor capabilities of PD patients[13]. Greffard et al. (2006) concluded that the MDS-UPDRS-III score is linearly linked to neuronal density which also represents the extent of neuron damage[14].

2. Methods

2.1 Study Data

The data utilized in this study was obtained from the Parkinson's Progression Markers Initiative (PPMI) database, which was established to identify PD progression biomarkers and to help advance the understanding of the disease mechanism[15]. The PPMI database contains numerous clinical, imaging, and biospecimen biomarker tests that are longitudinally collected from a cohort of newly diagnosed PD subjects and healthy control participants located at multiple centers.

All of the data extraction, neural network modeling, and statistical analysis are performed with Matlab R2018a (MATLAB, Natick, MA) and especially with the neural network toolbox on an Intel Core i7-2620M CPU @2.7GHz and 8GB of ram. For data extraction, CSV files were downloaded from the database and were imported into the Matlab environment for feature extraction. The utilized baseline-demographic information, biological-specimen features, and imaging data are listed in its entirety in **Table A** in the **Appendix**. Up-to-date biological specimen collection, measurement protocols, and structural magnetic resonance imaging (MRI) acquisition protocols can be found in more detail at ppmi-info.org.

Inclusion criteria for subjects include having all available data for demographics, biological specimens, as well as cognitive and behavioral assessment information at baseline listed in **Table A** in the Appendix. Subjects with one category missing will be included[1]. The averages are calculated separately for the normal cognition patients and the mild cognitively impaired

[1] There are 5 PD patients with missing baseline serum IGF data and 2 PD patients with missing CSF amyloid beta 1-42 data

patients according to their baseline MoCA score. The replacement of missing independent variable data is performed in order to increase the sample data size. To limit the amount of subject dropout for missing data, averages were used as replacements for data. Data replacement with the average is applied as to not affect the overall average. Subjects must also have longitudinal data available for the prediction outcomes of MDS-UPDRS-III and MoCA. When the subject has more than 2 input features missing, that subject was excluded from the study. Overall, 120 PD subjects and 42 healthy controls (HC) at baseline were included in this study with their mean/standard deviation shown in **Table 1.** With the inclusion of subjects at four different time points (one, two, three, and four years past baseline); study subject dropout is observed and the amount of remaining subjects is presented in **Table 2.** MDS-UPDRS-III and MoCA scores are averaged from the remaining subjects at each time point and is displayed in **Figure 1.** Although, there continues to be progression variability among patients, predicting future disease states becomes one of the difficult questions being addressed with this study.

Table 1: Demographic and clinical characteristics of PPMI cohort at baseline.

	Male	**Female**	**Age**	**MoCA**	**MDS-UPDRS-III**
HC (n=42)	27	15	59.3 ± 11.0	28.5 ± 1.2	-
PD (n=120)	77	43	60.1 ± 9.4	27.4 ± 2.1	20.8 ± 8.6

Table 2: Number of PD subjects with available longitudinal test score data past baseline.

	BL	**1 year**	**2 years**	**3 years**	**4 years**
MoCA	120	100	101	103	95
MDS-UPDRS-III	120	93	84	84	76

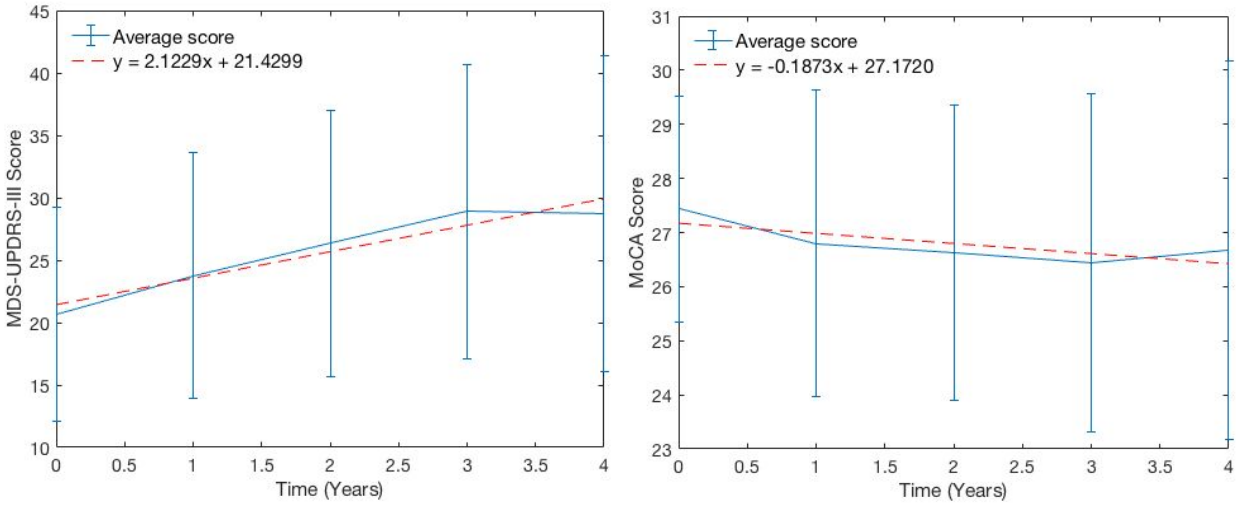

Figure 1: Average change (plus/minus 1 standard deviation shown with the blue error bars) in MDS-UPDRS-III (left) and MoCA (right) scores on the entire dataset. The dashed red line is the best fit line of the data with a corresponding slope and intercept. An average increase of 2 MDS-UPDRS-III points per year and a smaller average decrease in MoCA points is observed.

We bin the MDS-UPDRS-III scores into 3 categories, lightly mild, more mild, and moderate plus severe. These categories and optimal cutoff scores maximize the sample size in each group, and were validated in a study performed by Martínez-Martín et al. (2015)[16]. MoCA scores were similarly binned into 3 categories, normal cognition, borderline mild cognitive impairment, and mild cognitive impairment (MCI). See **Table 3-4** for details on the number of observations in each category.

Table 3: Amount of subjects in each binned MDS-UPDRS-III category with the cutoff values chosen as 21 and below for lightly mild, between 22 and 32 for more mild, and 33 and above for moderate and severe.

	BL	**1 year**	**2 years**	**3 years**	**4 years**
Lightly Mild	69	41	27	24	22
More Mild	39	31	34	23	26
Moderate + Severe	12	21	23	37	28

Table 4: Amount of subjects in each binned MoCA category with the cutoff values chosen as 28 and above for normal, 26 and 27 for borderline MCI, and 25 and lower as MCI and dementia.

	BL	1 year	2 years	3 years	4 years
Normal	64	50	42	44	51
Borderline MCI	37	18	31	24	16
MCI + Dementia	19	32	28	35	28

2.2 Structural MRI Acquisition

Structural MRI data are only acquired at baseline and is used to retrospectively determine whether or not baseline structural atrophy can be used to predict cognitive outcomes in PD patients. According to the PPMI protocol, structural MR images must consist of a 3-dimensional T1-weighted scan, using either a Magnetization Prepared Rapid Gradient Echo (MPRAGE) or a Spoiled Gradient Recalled (SPGR) sequence[15]. The field of view must include the vertex, cerebellum, and pons, and have a slice thickness of no greater than 1.5mm without an interslice gap[15]. The repetition time and echo time are not mandated by PPMI and thus set according to the individual manufacturer's recommendations.

2.3 Deformation-based Morphometry and the Brain Connectome

One set of medical imaging features used for cognitive and motor score prediction is deformation-based morphometry (DBM). DBM is a calculation of regional brain atrophy based on the displacement to a standard parcellation template. These atrophy values are computed by spatial normalization of structural MRI scans to the same MNI152-2009c template, which is provided in thanks to Zeighami et al. (2015)[17]. From the PPMI database, baseline structural MRI scans were analyzed on 232 PD subjects and 117 HC participants and corresponding atrophy

maps are generated. In addition, corresponding baseline atrophy maps could only be matched to a subset of subjects and about half of the DBM atrophy values were excluded from the study. Spatial smoothing is performed on the gray matter maps obtained from the MRI scans, hence, local image intensity reflects gray matter density. The final baseline DBM atrophy map includes 78 brain regions; where atrophy values from cerebellum regions are removed.

Due to recent advances in diffusion tensor imaging, the human brain connectome has seen much utilization and impact in the neuroscience community. The connectome can be in the form of a map or matrix representing neural connections within the brain. It can help us understand the interactions that are going on inside the brain and can add to our knowledge of understanding the brain and how cognitive changes occur[18]. An anatomical connectome is generated by using a fully automated fiber tractography algorithm[19]. After the construction of the brain connectome, an anatomical connection strength (ACS) map is derived by counting the number of nodes connecting two regions from each healthy control and taking an average[19]. This is an estimation of the potential flow of information being sent between the two regions. The ACS map is used to calculate the predicted rate of atrophy using the network diffusion model.

2.4 Network Diffusion Model

Various studies on neurodegenerative diseases have demonstrated that certain proteins such as tau, beta amyloid, and alpha-synuclein have a prion-like spread[20,21]. More specifically, studies on PD progression reveals spread of misfolded proteins to neocortical and cortical regions of the brain through the brain network[22,23]. The Network Diffusion Model (NDM) proposed by Raj. et al (2012) predicts atrophy rate based on the anatomical connectivity strength between brain

regions[24]. The NDM is a deterministic mathematical model that predicts regional atrophy rate based on the heat diffusion equation and regulated by the brain network connectivity.

For this study, a simplified linear baseline atrophy slope is calculated by first calculating the Laplacian matrix, *L*, given by **Equation 1**. *A* is the adjacency matrix or ACS map in this case. *D* is the diagonal matrix derived from the column-wise sum of the ACS map. A normalized Laplacian is calculated with **Equation 2**. Lastly, in order to obtain our analytical baseline atrophy rate, the normalized Laplacian, which is representative of the heat diffusion equation, is multiplied with the baseline DBM atrophy values.

$$L = D - A \quad \textbf{(1)}$$

$$L_{norm} = D^{-1/2} L D^{-1/2} \quad \textbf{(2)}$$

2.5 Stepwise Linear Regression

We implement a classical statistical technique as a starting point in order to compare the proposed deep learning model on this specific application. Stepwise Linear Regression (SLR) fits a dependent prediction output vector by adding or subtracting a linear combination of independent input predictor variables based on the added value it provides to the model while maintaining statistical significance.

SLR is performed with all baseline features including demographic information, biomarker tests, DBM atrophy values and with or without the NDM to test if it will improve the prediction accuracy. First, we separate the entire dataset into a 90% training set and a 10% hold-out

testing set. The 10% dataset is set aside to be used as a type of gold standard to be used to test the final prediction model. A 10-fold cross validation is performed on the remaining 90% training dataset where the observations were split into 90% training and 10% validation. This process is repeated ten times where each 10% validation set is different for each iteration. Averages of the ten prediction accuracies is calculated at each future time point and is shown in **Table 6-7**. The implemented deep learning model is also validated using 10-fold cross validation.

2.6 Autoencoder

Currently, we are researching and implementing the autoencoder as our unsupervised learning model. Our autoencoder is fully connected and utilizes backpropagation and scaled conjugate gradient descent for training. The training phase stops when either the maximum number of iterations is complete or when a local minimum is found. This type of neural network is able to encode or compress an input vector into a smaller representation of data. The decoder function can approximately expand the reduced vector back into its original size. The encoder and decoder transfer functions are set accordingly in order to minimize the mean square error of the reconstructed vector and the original input vector. To achieve this, the logistic sigmoid function is chosen, which transforms data nonlinearly as opposed to the positive saturating linear function. The latter transforms data linearly, which is not desirable for minimizing error on our application. Given these two transfer function options, it can be concluded that the logistic sigmoid function provides greater optimization. The global error cost function of the proposed autoencoder is based on the adjusted mean square error function. See **Figure 2** for architectural details.

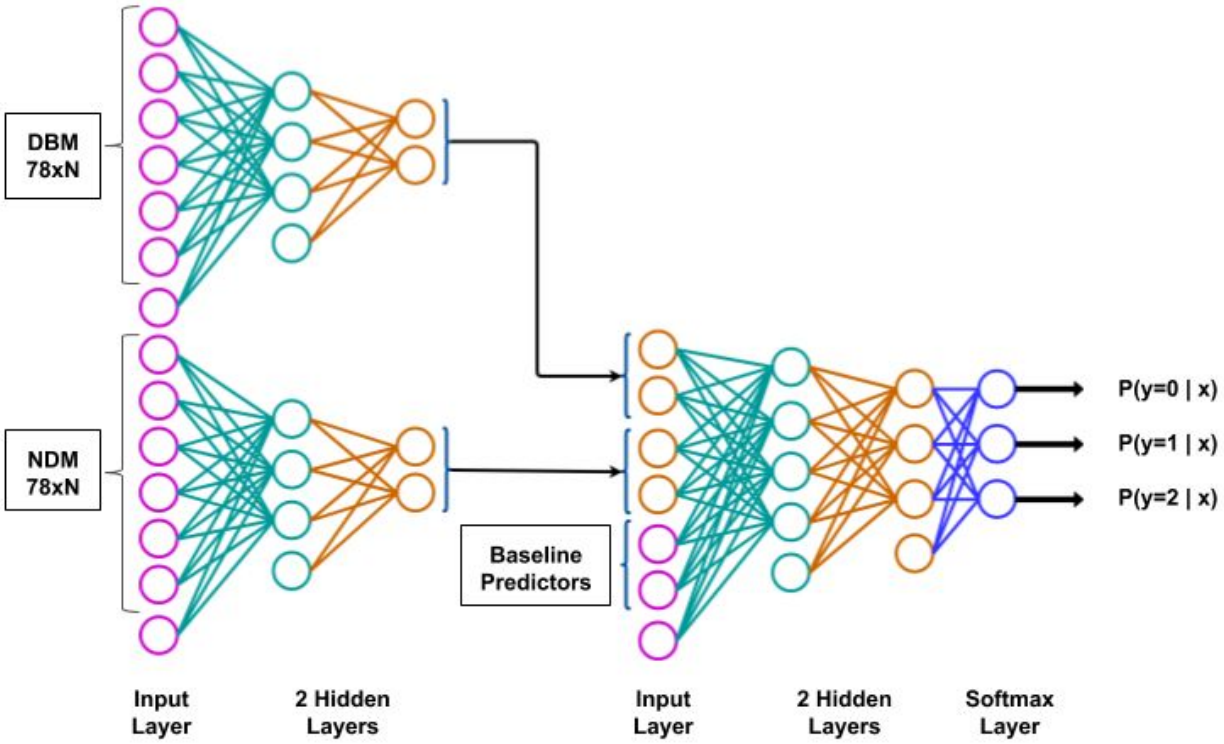

Figure 2: Architecture of the implemented deep autoencoder neural network. First, the 78 brain atrophy regions and the 78 predicted atrophy slopes were reduced down to a chosen size i.e. 20, 10, or 5. The reduced features were added onto another input layer with 36 baseline predictors. The features are fed into the network and into the softmax classification layer.

First, the 78 brain atrophy regions and the 78 predicted atrophy rates are reduced down to a smaller representation of data with two hidden layers. This forces the autoencoder to learn a smaller representation of the atrophy values. The reduced imaging features with and without the addition of the analytical projections obtained from the NDM are combined with baseline demographics and biospecimen measurements. Stacking of the individual sparse autoencoders at each layer creates a deep autoencoder network. A supervised softmax classification layer is used to to link the input to the desired output. The softmax layer generates a probability matrix for each possible class. The entire deep autoencoder model is finely tuned at the end so that it

only takes one iteration to improve all of the weights simultaneously. To fine tune the model, we implement backpropagation on the entire stacked autoencoder network by retraining it on the training data in a supervised fashion.

The starting weights and biases are set randomly at the start of each model training with a normal distribution of a small value close to, but not equal to zero. Weights and biases are learnable parameters of the autoencoder model and are adjusted towards values that have correct output. When developing a deep neural network, certain hyperparameters must first be declared before training a model. Optimal hyperparameters are not known beforehand and different values are systematically experimented in a sequential search by changing one variable at a time[25]. Performance results from the validation set is recorded and the best hyperparameters are saved for the final model.

In order to prevent overfitting of the training data, we declare specific parameters for the L2 weight regularization, sparsity regularization, and sparsity proportion coefficients terms onto the cost function at the start of model training. L2 regularization helps drive outlier weights closer to zero. By regularizing or shrinking the L2 weight coefficient, prediction accuracy may be improved and the variance may be reduced. The larger the sparsity regularization parameter, the greater its impact on activated training data. The sparsity proportion specifies the amount of training examples a neuron reacts with, whereas the sparsity regularization term controls the impact of sparsity for faster optimization and evaluation of the model. By implementing this restriction on sparsity, this pressures the neural network to reduce and store only the essential features of the data. The investigated range of hyperparameters is shown in **Table 5**.

Table 5: Researched autoencoder hyperparameters and optimal parameters were observed for different prediction outcomes.

Hyperparameters	Feature Reduction Size	Hidden Layer Neuron Size	L2 weight Regularization	Sparsity Regularization	Sparsity Proportion
Test Range	5xN - 20xN	20/10, 20/5	0.001 - 0.009	2 - 6	0.05 - 0.2
Optimal for MDS-UPDRS-III	20xN	20/10	0.005	3	0.2
Optimal for MoCA	5xN	20/10	0.006	4	0.1

The deep autoencoder network trains with all baseline features. Initially, the model was optimized on a 80/10/10 split of the data where 80% of the data is used for training, 10% is used for developing and improving performance, and the final 10% is tested for validation. In order to verify the developed model, 10-fold cross validation is applied on the training data set. The average prediction accuracy is then calculated and is shown in **Table 8-9**.

2.7 Statistical Analysis

Using the Pearson's linear correlation coefficient, each baseline input predictor is correlated to the binned outcome of MDS-UPDRS-III or MoCA. Values of $P < 0.05$ are considered statistically significant. The range of the correlation coefficient span from -1 to +1, where a value of -1 is a perfect negative correlation and a value of +1 is a perfect positive correlation. All statistically significant input predictors are presented in **Table B and C in the Appendix**. More input variables are shown to be correlated to MoCA than with MDS-UPDRS-III as well as having a stronger relationship strength.

3. Results

In this study, we predict a variety of outcomes including raw future scores, binned future scores, raw change in scores, and binned change in scores. The best prediction accuracy results came from predicting future scores binned into three categories. For a standard comparison, research on a classical regression method is performed to evaluate the performance between a classical regression method and an emerging and novel deep learning model. The best hyperparameters for the autoencoder model to predict future motor and cognitive decline are detailed in **Table 5**.

3.1 Predicting Motor Scores

For predicting future motor scores, prediction outcome data from MDS-UPDRS-III are classified into three categories consisting of lightly mild, more mild, and moderate. The mild category is split into two in order to equalize the histogram of the amount of subjects in each bin. Prediction accuracy results are represented in **Table 6** as the average of 10 iterations from 10-fold cross validation, as detailed in section 2.5. The results show a decrease in prediction accuracy as the time difference increases after baseline. Efficiency, of both the SLR and autoencoder model with the NDM, decreases more notably when predicting at year 3 and 4 past baseline versus the earlier years.

The addition of the network diffusion model did not significantly improve the prediction accuracy in either models, there was a slight improvement for SLR with NDM at year 4. In general, the average accuracy did improve when using the Autoencoder model at all time points. When implementing the autoencoder model, an improvement of about 20-30% is observed. A distribution of the prediction accuracy of only the autoencoder is shown in **Figure 3** with the

center point as the average and the error bars as one standard deviation. The autoencoder model has the potential to accurately predict future motor decline with up to the 80% accuracy.

Table 6: Accuracy results of MDS-UPDRS-III prediction for two tested models both predicting 3 categories of MDS-UPDRS-III which are lightly mild, more mild, and moderate.

	Stepwise Linear Regression without NDM	**Stepwise Linear Regression with NDM**	**Autoencoder without NDM**	**Autoencoder with NDM**
1 year	53.10%	52.10%	72.12%	69.10%
2 years	54.42%	45.78%	72.68%	70.25%
3 years	41.53%	34.86%	59.73%	56.96%
4 years	34.73%	39.68%	62.84%	59.34%

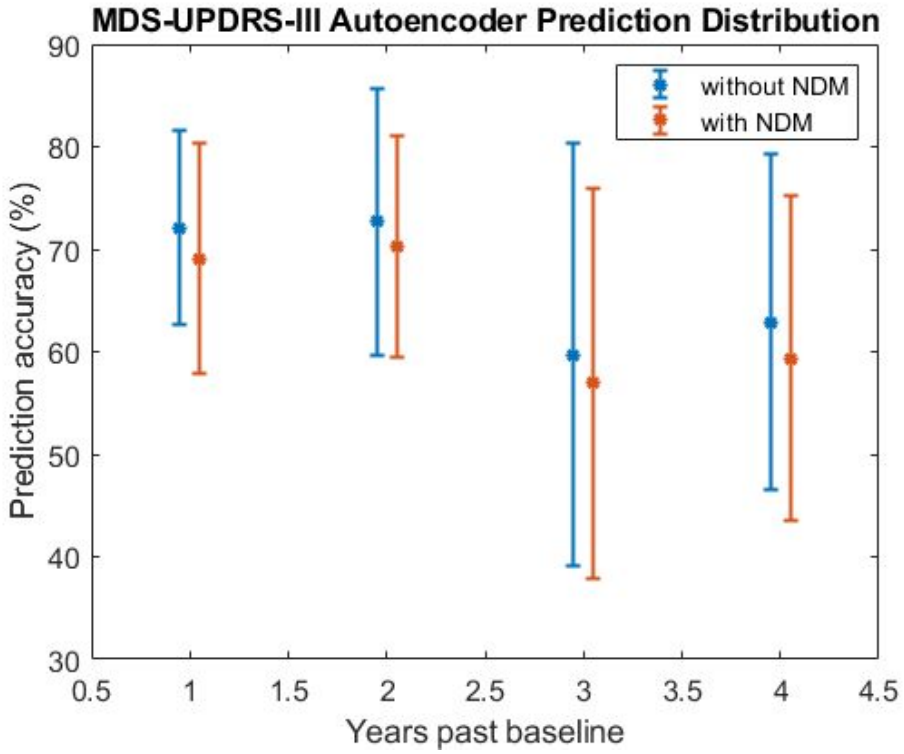

Figure 3: Autoencoder prediction accuracy distribution plot of MDS-UPDRS-III from 10-fold cross-validation.

3.2 Predicting Cognitive Scores

We separated the prediction of future categories of MoCA into three categories for the same reason as MDS-UPDRS-III. Ideally, we aim to normalize the amount of subjects across all three categories. These three categories are normal cognition, borderline mild cognitive impairment, and mild cognitive impairment. Prediction accuracy results for MoCA are represented in **Table 7** as the average of 10 iterations from 10-fold cross validation. The efficiency of the SLR and autoencoder model varies across all future time points where poorer results are observed at years 2 and 3, with better results seen at opposite ends of the predicted time frame.

By comparing the performance of the SLR model with and without the NDM, results indicate the addition of the NDM improves the prediction accuracy at year 3 and 4 but not at the earlier time points. When the autoencoder model was trained with and without the NDM, the addition of the NDM improved the prediction accuracy at all future time points except at year 2. The 1% difference in average prediction accuracy at year 2 with and without the NDM is not significant and may be a result of poorly optimized hyperparameters at each future time point.

Overall, the average accuracy increased when using the Autoencoder model at all time points by about 10-20%. A distribution for the prediction accuracy of MoCA using only the autoencoder model is shown in **Figure 4** with the center point as the average and the error bars as one standard deviation. The autoencoder model is proven to be the better method for predicting the future cognitive state of PD patients.

Table 7: Accuracy results of MoCA prediction for two tested models both predicting 3 categories of MoCA which are normal cognition, borderline mild cognitive impairment, and mild cognitive impairment.

	Stepwise Linear Regression without NDM	**Stepwise Linear Regression with NDM**	**Autoencoder without NDM**	**Autoencoder with NDM**
1 year	52.00%	46.00%	50.00%	54.00%
2 years	36.64%	35.82%	48.44%	47.44%
3 years	37.73%	40.73%	42.01%	43.81%
4 years	41.22%	43.22%	52.67%	58.12%

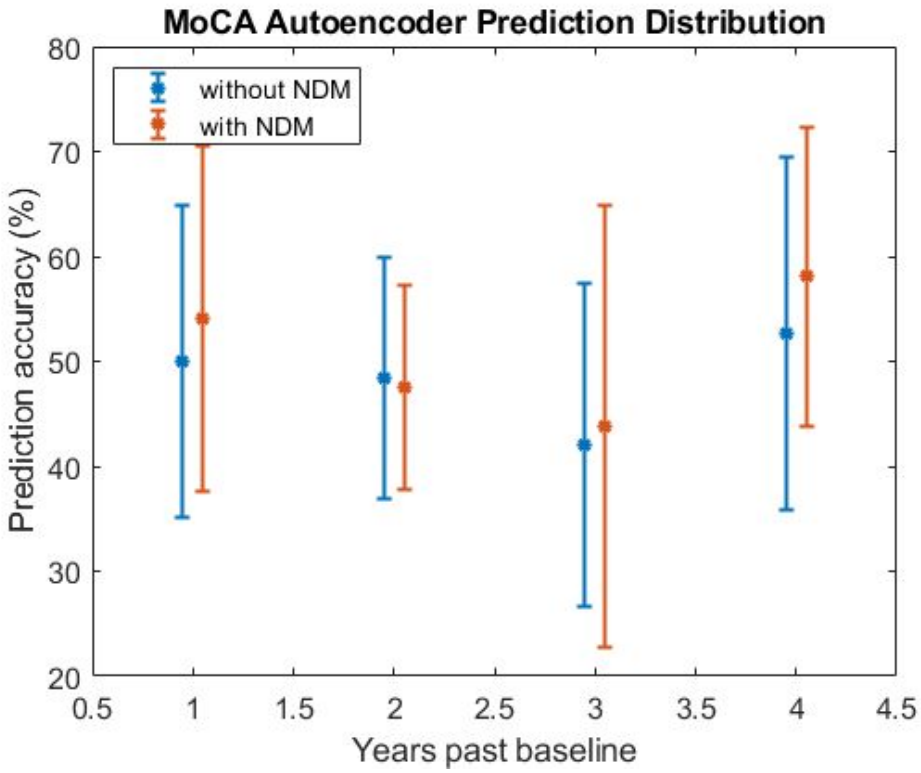

Figure 4: Autoencoder prediction accuracy distribution plot of MoCA from 10-fold cross-validation.

4. Discussion

The goal of this study is to investigate the precision how precisely a deep learning model can predict future motor and cognitive states of a patient using only (i) features collected at baseline, and (ii) analytical predictions given by the network diffusion model developed by Raj et al. (2012). We implement an autoencoder deep learning model, which shows an improvement in prediction accuracy over classical stepwise linear regression and provides further evidence for using deep learning to increase predictive performance. We test a range of hyperparameters and optimal values are selected depending on the anticipated outcome. The network diffusion model was also assessed to determine if it added value to the predictive model.

The stepwise linear regression model did worse overall when compared to the performance of the autoencoder deep learning model. This is due to PD progression variability among every patient and the non linearity of neurodegeneration. Thus, a linear regression model may perform poorly on non-linear neurodegeneration patterns. This study provides validation that deep learning models may be better suited for predicting nonlinear future outcomes.

4.1 Comparison between SLR and Autoencoder

The addition of the predicted atrophy slope rate from the NDM improved the prediction accuracy for future MoCA but not for future MDS-UPDRS-III scores. For MDS-UPDRS-III predictions, both models did not benefit from the additional analytical projections provided by the NDM. This may be due to the fact that motor control is mostly in the deep subcortical brain regions such as the basal ganglia and additional data on the predicted atrophy rate of other brain regions became noise[14]. To address this issue, feature extraction of only brain regions associated with motor control may be used for predicting MDS-UPDRS-III rather than using all brain regions.

Predicting future cognition is improved when we add the NDM as additional predictors for both SLR and the autoencoder at year 3 and 4. Studies have reported that cognitive function of PD patients strongly correlates with the overall normalized brain size. Additionally, investigators observe atrophy in many cortical and subcortical brain regions[26]. A higher percentage of the 78 atrophy values now have more impact on the prediction model due to the addition of cortical brain regions such as the frontal and parietal lobes.

The prediction models benefit more from having more statistically significant brain regions with a P-value of less than 0.05 meaning that there is stronger evidence supporting our hypothesis. Referring to **Tables B and C in the Appendix**, more DBM atrophy values and predicted NDM atrophy rates show correlation to the future MoCA categories than compared with MDS-UPDRS-III. The addition of more predictors improved the performance of the autoencoder model.

4.2 Justification and Rationale

Studies show that DBM is the optimal method to detect sub cortical irregularities when applied to temporal lobe epilepsy when compared to voxel-based morphometry (VBM)[27]. Based off of the data shared with our lab from Zeighami and group, we decided against using VBM because it does not accurately reflect all of the MRI data and it is noted that VBM may be less sensitive to subcortical atrophy[17]. However, DBM is still not a fully accurate representation and errors may be introduced into the atrophy maps due to technical limitations of image processing pipelines and hardware limitations of imaging modalities.

For this study, adopting the autoencoder deep learning model is promising for its ability to reduce dimensionality without reducing the meaningful data qualities as well as removing noise from data. Dimensionality reduction is particular helpful for decreasing the size of the dataset in order to improve the efficiency of developing and training a model. This model also learns only on the input data and does not require any type of feature engineering or extraction. Despite all of the freely accessible and different types of autoencoders, we choose to work with the stacked autoencoder that forms a neural network, however, there are many other approaches to deep learning and this is just the starting point for our specific usage case. Future work is being done on predicting two categories instead of three. Our study data set only consists of newly diagnosed PD patients whom many patients do not suffer from severe motor or cognitive decline. It may be more relevant to be able to predict if a patient will convert from mild to moderate motor dysfunction or from normal to mild cognitive impairment for cognitive decline. By splitting the mild category for MDS-UPDRS-III and splitting the normal cognition category for MoCA, both predictive models generally show low specificity and low sensitivity. The range of the middle categories is slim and the model has a difficult time with classification.

4.3 Limitations

The PPMI database is one of the largests of its kind, however, it still lacks in the amount of subjects as well as longitudinal data. Validation of the NDM and autoencoder model was limited to the PPMI database size and by the narrow timespan range of 1 to 4 years past baseline. Missing baseline data from some subjects, such as serum IGF and CSF amyloid beta measurements, causes the dataset to shrink in size if too many baseline predictors are missing. The small size of the data set may not be optimal for deep learning applications and machine

learning may be used instead. Deep learning models improve in skill as more data is given for training.

The NDM outputs a linear atrophy rate and this limits the researched prediction model's ability to account for variability in disease progression where atrophy can be faster or slower at different stages of the disease. This can not be independently verified due to the fact that structural MRI were only imaged at baseline and at no other future timepoints. There is no ground truth to compare the network diffusion model to for this specific application.

There is a wide variability in atrophy patterns as well as cognitive and motor tests between each PD patient. Specifically, there may be some clinician subjectivity for MoCA and MDS-UPDRS-III tests. There may be variability between different subjects from different study centers with different test proctors. Total scores were only considered for this study. Bias may be introduced if a patient performed satisfactory overall but perhaps very poorly in a select group of sub categories. The "gold standard" diagnosis and prognosis of PD is a more comprehensive, formal testing by a neuropsychologist. The addition of that high quality testing data may help with the training of the neural network.

Hyperparameters could have been explored in more detail. A wider range as well as a finer step size within the tested hyperparameters can be investigated, however, this will increase the time and computational power required for training and validation.

4.4 Future work

One way to potentially improve the proposed deep learning model is to add more variables as predictors, where a fuller dataset can be used by the model to learn a better representation of the data. This can be done by extracting more data from the PPMI database such as the extensive amount of available genetic factors that may benefit more from deep learning which can extract more meaningful representations of the data when compared to machine learning techniques. As an ongoing PPMI study, more patients will be added to the study leading to an increase of potential subjects for a more substantial data set. Additional subjects have been added recently and longitudinal data is still in the process of collection and organization.

Due to the smaller starting data set, machine learning applications may be better suited. Leave one out validation may also be performed but it is more computationally expensive to retrain a model the amount of times there are for the number of observations. Other deep learning models that were considered at the start of the study are the deep boltzmann machines, convolutional neural networks, and generative adversarial networks. Additional research is required for further validation of the network diffusion model as well as its implementation with Parkinson's disease.

References

1. S. Emrani, A. Mcguirk, and W. Xiao, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 17 1457 (2017).
2. Prashanth, R., & Roy, S. D. (2018). Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning. Neurocomputing, 305, 78-103. doi:10.1016/j.neucom.2018.04.049
3. Falahati, F., Westman, E., & Simmons, A. (2014). Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. Journal of Alzheimer's Disease,41(3), 685-708. doi:10.3233/jad-131928
4. Raj, A., & Powell, F. (2018). Models of Network Spread and Network Degeneration in Brain Disorders. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3(9), 788-797. doi:10.1016/j.bpsc.2018.07.012
5. Raj, A., Locastro, E., Kuceyeski, A., Tosun, D., Relkin, N., & Weiner, M. (2015). Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer’s Disease. Cell Reports, 10(3), 359-369. doi:10.1016/j.celrep.2014.12.034
6. Kamagata, K., Zalesky, A., Hatano, T., Biase, M. A., Samad, O. E., Saiki, S., . . . Pantelis, C. (2018). Connectome analysis with diffusion MRI in idiopathic Parkinson's disease: Evaluation using multi-shell, multi-tissue, constrained spherical deconvolution. NeuroImage: Clinical, 17, 518-529. doi:10.1016/j.nicl.2017.11.007
7. Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., . . . Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. Journal of Neuroscience Methods, 222, 230-237. doi:10.1016/j.jneumeth.2013.11.016
8. Hirschauer, T. J., Adeli, H., & Buford, J. A. (2015). Computer-Aided Diagnosis of Parkinson’s Disease Using Enhanced Probabilistic Neural Network. Journal of Medical Systems, 39(11). doi:10.1007/s10916-015-0353-9

9. Grover, S., Bhartia, S., A., Yadav, A., & K.r., S. (2018). Predicting Severity Of Parkinson's Disease Using Deep Learning. Procedia Computer Science, 132, 1788-1794. doi:10.1016/j.procs.2018.05.154

10. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. doi:10.1038/nature14539

11. Fengler, S., Kessler, J., Timmermann, L., Zapf, A., Elben, S., Wojtecki, L., . . . Kalbe, E. (2016). Screening for Cognitive Impairment in Parkinson's Disease: Improving the Diagnostic Utility of the MoCA through Subtest Weighting. Plos One, 11(7). doi:10.1371/journal.pone.0159318

12. Hoops, S., Nazem, S., Siderowf, D., Duda, J. E., Xie, S. X., Siderowf, A. D., . . . Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. Neurology, 73. doi:10.1212/WNL.0b013e3181c34b47

13. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. (2003). Movement Disorders, 18(7), 738-750. doi:10.1002/mds.10473

14. Greffard, S., Verny, M., Bonnet, A., Beinis, J., Gallinari, C., Meaume, S., . . . Duyckaerts, C. (2006). Motor Score of the Unified Parkinson Disease Rating Scale as a Good Predictor of Lewy Body–Associated Neuronal Loss in the Substantia Nigra. Archives of Neurology,63(4), 584. doi:10.1001/archneur.63.4.584

15. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., . . . Taylor, P. (2011). The Parkinson Progression Marker Initiative (PPMI). Progress in Neurobiology, 95(4), 629-635. doi:10.1016/j.pneurobio.2011.09.005

16. Martínez-Martín, P., Rodríguez-Blázquez, C., Alvarez, M., Arakaki, T., Arillo, V. C., Chaná, P., . . . Merello, M. (2015). Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. Parkinsonism & Related Disorders, 21(1), 50-54. doi:10.1016/j.parkreldis.2014.10.026

17. Zeighami, Y., Ulla, M., Iturria-Medina, Y., Dadar, M., Zhang, Y., Larcher, K. M., . . . Dagher, A. (2015). Network structure of brain atrophy in de novo Parkinson's disease. ELife, 4. doi:10.7554/elife.08440

18. Sporns, O., Tononi, G., & Kötter, R. (2005). The Human Connectome: A Structural Description of the Human Brain. PLoS Computational Biology, 1(4). doi:10.1371/journal.pcbi.0010042

19. Iturria-Medina, Y., Canales-Rodríguez, E., Melie-García, L., Valdés-Hernández, P., Martínez-Montes, E., Alemán-Gómez, Y., & Sánchez-Bornot, J. (2007). Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. NeuroImage, 36(3), 645-660. doi:10.1016/j.neuroimage.2007.02.012

20. Sweeney, P., Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., . . . Hodgson, R. (2017). Protein misfolding in neurodegenerative diseases: Implications and strategies. Translational Neurodegeneration, 6(1). doi:10.1186/s40035-017-0077-5

21. Brundin, P., Melki, R., & Kopito, R. (2010). Prion-like transmission of protein aggregates in neurodegenerative diseases. Nature Reviews Molecular Cell Biology, 11(4), 301-307. doi:10.1038/nrm2873

22. Yau, Y. H., Zeighami, Y., Baker, T., Larcher, K., Vainik, U., Dadar, M., . . . Dagher, A. (2017). Network Connectivity Determines Cortical Thinning In Early Parkinsons Disease Progression. doi:10.1101/147611

23. Biundo, R., Weis, L., & Antonini, A. (2016). Cognitive decline in Parkinson's disease: The complex picture. Npj Parkinson's Disease, 2(1). doi:10.1038/npjparkd.2016.18

24. Raj, A., Kuceyeski, A., & Weiner, M. (2012). A Network Diffusion Model of Disease Progression in Dementia. Neuron, 73(6), 1204-1215. doi:10.1016/j.neuron.2011.12.040

25. Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics, 5(1). doi:10.1007/s13721-016-0125-6

26. Prakash, K. G., Bannur, B. M., Chavan, M. D., Saniya, K., Sailesh, K. S., & Rajagopalan, A. (2016). Neuroanatomical changes in Parkinson's disease in relation to cognition: An update. Journal of Advanced Pharmaceutical Technology & Research, 7(4), 123–126. doi: 10.4103/2231-4040.191416

27. Scanlon, C., Mueller, S., Tosun, D., Cheong, I., Garcia, P., Barakos, J., . . . Laxer, K. (2011). Impact of Methodologic Choice for Automatic Detection of Different Aspects of Brain Atrophy by Using Temporal Lobe Epilepsy as a Model. American Journal of Neuroradiology, 32(9), 1669-1676. doi:10.3174/ajnr.a2578

Appendix

Table A: List of all baseline input predictors

Demographics
Sex
Age
Years of formal education
Biospecimens
Serum Insulin-like Growth Factor-1
CSF α-synuclein
CSF Aβ42
CSF Total tau
CSF Phosphorylated tau
Cognitive and Behavioral Assessments
MoCA
Hopkins Verbal Learning Test (3 sections)
Immediate Recall, Delayed Recognition Hits, Delayed Recognition False Alarms
Benton Judgement of Line Orientation
Semantic Fluency Total Score
Symbol Digit Modalities Score
UPSIT
Geriatric Depression Scale (GDS-15)
Impulse control - QUIP (7 sections)
+ (Gambling, Sex, Buying, Eating, Hobbies, Punding, Walking or Driving)
SCOPA-AUT
MDS-UPDRS (5 sections)
Part I, Part 1 Patient Questionnaire, Part II Patient Questionnaire, Part III, Total
Hoehn & Yahr
Modified Schwab England ADL
Imaging
Striatal Binding Ratio (4 regions)
Left caudate, right caudate, left putamen, right putamen
78 deformation-based morphometry atrophy values
78 analytical atrophy rate predictions given by the network diffusion model

Table B: Pearson's linear correlation coefficients of baseline predictors at four time points past baseline to MDS-UPDRS-III.

	Predictor	R value	P value
1 year	Hoehn and Yahr	0.3424	0.0008
	IGF	0.2836	0.0059
	SCOPA-AUT	0.2267	0.0289
	SBR Right Putamen	-0.2039	0.0451
	DBM Region 34	-0.2050	0.0487
	NDM Region 34	-0.2051	0.0486
	Symbol Digit Modalities Score	-0.3087	0.0026
	Modified Schwab England ADL	-0.3450	0.0007
2 years	Hoehn and Yahr	0.3848	0.0003
	NDM Region 68	0.3040	0.0049
	DBM Region 68	0.3023	0.0052
	NDM Region 67	0.2686	0.0135
	DBM Region 67	0.2635	0.0155
	QUIP Positive Sex	0.2600	0.0169
	DBM Region 47	0.2233	0.0411
	NDM Region 47	0.2196	0.0448
	DBM Region 30	0.2150	0.0495
	NDM Region 2	-0.2188	0.0455
	NDM Region 40	-0.2232	0.0413
	NDM Region 36	-0.2243	0.0403
	DBM Region 45	-0.2262	0.0386
	NDM Region 45	-0.2558	0.0188
	Symbol Digit Modalities Score	-0.2754	0.0112
	HVLT Immediate Recall	-0.2850	0.0086
	HVLT Delayed Recognition Hits	-0.3313	0.0021
	DBM Region 8	-0.4030	<0.0001
	NDM Region 8	-0.4134	<0.0001
	Modified Schwab England ADL	-0.4134	<0.0001
3 years	DBM Region 33	0.2674	0.0139
	NDM Region 33	0.2620	0.0160
	Hoehn and Yahr	0.2588	0.0175
	NDM Region 69	0.2479	0.0230
	DBM Region 69	0.2426	0.0262
	NDM Region 21	0.2305	0.0349
	DBM Region 21	0.2246	0.0399
	Alpha-Synuclein	-0.0699	0.0171
	SBR Right Caudate	-0.0858	0.0471
	SBR Left Caudate	-0.0939	0.0164
	SBR Left Putamen	-0.1192	0.0412
	SBR Right Putamen	-0.1282	0.0489
	DBM Region 20	-0.2276	0.0373
	NDM Region 20	-0.2899	0.0075
	NDM Region 49	-0.2918	0.0071
	Modified Schwab England ADL	-0.4134	0.0065

	Predictor	R value	P value
4 years	Age	0.3701	0.0010
	SCOPA-AUT	0.3121	0.0061
	NDM Region 68	0.2733	0.0170
	NDM Region 67	0.2615	0.0225
	DBM Region 68	0.2588	0.0234
	Hoehn and Yahr	0.2561	0.0256
	DBM Region 67	0.2280	0.0476
	Semantic Fluency Total Score	-0.2273	0.0483
	NDM Region 78	-0.2281	0.0475
	Modified Schwab England ADL	-0.2348	0.0412
	DBM Region 78	-0.2349	0.0411
	DBM Region 63	-0.2356	0.0405
	SBR Left Putamen	-0.2449	0.0330
	DBM Region 24	-0.2557	0.0258
	NDM Region 77	-0.2628	0.0218
	SBR Right Putamen	-0.2629	0.0218
	DBM Region 77	-0.2681	0.0192
	NDM Region 24	-0.2753	0.0161
	HVLT Delayed Recognition Hits	-0.2809	0.0140
	SBR Right Caudate	-0.3100	0.0064
	SBR Left Caudate	-0.3104	0.0064
	Symbol Digit Modalities Score	-0.3121	0.0061
	UPSIT Total Score	-0.3121	0.0061
	Benton Judgement of Line	-0.3769	0.0008

Table C: Pearson's linear correlation coefficients of baseline predictors at four time points past baseline in relation to MoCA.

	Predictor	R value	P value
1 year	Age	0.4325	<0.0001
	NDM Region 42	0.3758	<0.0001
	MDS-UPDRS-I	0.3212	0.0011
	DBM Region 42	0.2623	0.0084
	MDS-UPDRS-Total	0.2614	0.0086
	HVLT Delayed Recognition False Alarms	0.2554	0.0103
	NDM Region 28	0.2353	0.0185
	SCOPA-AUT	0.2295	0.0216
	NDM Region 48	0.2188	0.0287
	MDS-UPDRS-III	0.2012	0.0447
	NDM Region 40	-0.2080	0.0378
	NDM Region 36	-0.2093	0.0366
	DBM Region 73	-0.2181	0.0293
	DBM Region 1	-0.2228	0.0259
	DBM Region 4	-0.2289	0.0219
	ABeta 1-42	-0.2290	0.0219
	DBM Region 40	-0.2331	0.0196
	DBM Region 35	-0.2420	0.0153
	NDM Region 74	-0.2513	0.0117
	NDM Region 77	-0.2591	0.0092
	DBM Region 36	-0.2602	0.0089
	NDM Region 78	-0.2630	0.0082
	UPSIT	-0.2633	0.0081
	DBM Region 74	-0.2648	0.0077
	DBM Region 78	-0.2725	0.0061
	DBM Region 77	-0.2750	0.0056
	HVLT Delayed Recognition Hits	-0.2779	0.0051
	Symbol Digit Modalities Score	-0.3820	<0.0001
	Semantic Fluency Total Score	-0.4330	<0.0001
	HVLT Immediate Recall	-0.4546	<0.0001
2 years	Age	0.3880	<0.0001
	MDS-UPDRS-I	0.2992	0.0024
	NDM Region 42	0.2547	0.0102
	HVLT Delayed Recognition False Alarms	0.2463	0.0130
	GDS Score	0.2334	0.0188
	SCOPA-AUT	0.2255	0.0234
	QUIP Positive Walking or Driving	0.1971	0.0481
	Benton Judgement of Line	-0.1995	0.0455
	NDM Region 75	-0.2081	0.0368
	DBM Region 34	-0.2085	0.0364
	DBM Region 73	-0.2118	0.0335
	ABeta 1-42	-0.2265	0.0227
	DBM Region 38	-0.2333	0.0189
	NDM Region 76	-0.2363	0.0174
	NDM Region 36	-0.2365	0.0173
	DBM Region 75	-0.2432	0.0143
	DBM Region 76	-0.2561	0.0097
	NDM Region 35	-0.2631	0.0079
	NDM Region 77	-0.2710	0.0061

	Predictor	R value	P value
2 years	NDM Region 78	-0.2767	0.0051
	DBM Region 36	-0.2806	0.0045
	DBM Region 78	-0.2845	0.0039
	DBM Region 77	-0.2877	0.0035
	Semantic Fluency Total Score	-0.3127	0.0015
	DBM Region 35	-0.3303	0.0001
	Symbol Digit Modalities Score	-0.3889	<0.0001
	HVLT Immediate Recall	-0.4023	<0.0001
3 years	Age	0.4250	<0.0001
	NDM Region 42	0.3318	<0.0001
	NDM Region 27	0.2634	0.0072
	MDS-UPDRS-III	0.2348	0.0170
	NDM Region 49	0.2302	0.0193
	DBM Region 42	0.2253	0.0221
	Sex	0.2036	0.0391
	NDM Region 4	-0.1956	0.0477
	NDM Region 34	-0.2021	0.0407
	DBM Region 63	-0.2024	0.0404
	NDM Region 39	-0.2030	0.0397
	Benton Judgement of Line	-0.2140	0.0300
	DBM Region 34	-0.2187	0.0265
	DBM Region 1	-0.2234	0.0233
	DBM Region 33	-0.2240	0.0230
	Modified Schwab England ADL	-0.2253	0.0221
	NDM Region 75	-0.2263	0.0216
	DBM Region 2	-0.2272	0.0210
	NDM Region 77	-0.2443	0.0129
	DBM Region 59	-0.2510	0.0105
	NDM Region 36	-0.2623	0.0074
	NDM Region 74	-0.2655	0.0067
	DBM Region 39	-0.2673	0.0063
	DBM Region 37	-0.2679	0.0062
	DBM Region 4	-0.2712	0.0056
	NDM Region 76	-0.2724	0.0054
	NDM Region 40	-0.2753	0.0049
	DBM Region 77	-0.2831	0.0038
	DBM Region 75	-0.2839	0.0037
	NDM Region 35	-0.2885	0.0031
	DBM Region 38	-0.2930	0.0027
	DBM Region 74	-0.2963	0.0024
	NDM Region 73	-0.3036	0.0018
	DBM Region 76	-0.3065	0.0016
	NDM Region 78	-0.3135	0.0013
	UPSIT	-0.3144	0.0012
	DBM Region 40	-0.3156	0.0012
	DBM Region 73	-0.3312	<0.0001
	DBM Region 78	-0.3319	<0.0001
	DBM Region 36	-0.3365	<0.0001
	HVLT Delayed Recognition Hits	-0.3656	<0.0001
	DBM Region 35	-0.3888	<0.0001
	HVLT Immediate Recall	-0.4410	<0.0001

	Predictor	R value	P value
3 years	Symbol Digit Modalities Score	-0.4839	<0.0001
	Semantic Fluency Total Score	-0.5617	<0.0001
4 years	Age	0.5087	<0.0001
	SCOPA-AUT	0.3401	<0.0001
	NDM Region 42	0.2886	0.0046
	MDS-UPDRS-Total	0.2784	0.0063
	HVLT Delayed Recognition False Alarms	0.2766	0.0067
	MDS-UPDRS-III	0.2423	0.0180
	MDS-UPDRS-I Patient Questionnaire	0.2412	0.0185
	DBM Region 39	-0.2046	0.0467
	NDM Region 76	-0.2170	0.0347
	NDM Region 78	-0.2230	0.0298
	DBM Region 37	-0.2241	0.0290
	DBM Region 75	-0.2284	0.0260
	NDM Region 34	-0.2341	0.0224
	DBM Region 40	-0.2344	0.0222
	Benton Judgement of Line	-0.2375	0.0205
	DBM Region 34	-0.2391	0.0196
	DBM Region 78	-0.2432	0.0176
	NDM Region 74	-0.2454	0.0165
	NDM Region 73	-0.2459	0.0163
	DBM Region 76	-0.2492	0.0149
	DBM Region 74	-0.2683	0.0086
	DBM Region 73	-0.2699	0.0082
	NDM Region 77	-0.2732	0.0074
	UPSIT	-0.2749	0.0070
	DBM Region 38	-0.2798	0.0060
	NDM Region 36	-0.2872	0.0048
	DBM Region 77	-0.2981	0.0033
	DBM Region 36	-0.3350	<0.0001
	HVLT Delayed Recognition Hits	-0.3392	<0.0001
	NDM Region 35	-0.3667	0.0003
	DBM Region 35	-0.3998	<0.0001
	HVLT Immediate Recall	-0.5056	<0.0001
	Symbol Digit Modalities Score	-0.5229	<0.0001
	Semantic Fluency Total Score	-0.5301	<0.0001

Table D: Brain region labels

1	Hippocampus (right)	40	Pallidum (right)
2	Hippocampus (left)	41	Precentral gyrus (left)
3	Amygdala (right)	42	Precentral gyrus (right)
4	Amygdala (left)	43	Straight gyrus (left)
5	Anterior temporal lobe, medial part (right)	44	Straight gyrus (right)
6	Anterior temporal lobe, medial part (left)	45	Anterior orbital gyrus (left)
7	Anterior temporal lobe, lateral part excluding superior temporal gyrus (right)	46	Anterior orbital gyrus (right)
8	Anterior temporal lobe, lateral part excluding superior temporal gyrus (left)	47	Inferior frontal gyrus (left)
9	Parahippocampal and ambient gyri (right)	48	Inferior frontal gyrus (right)
10	Parahippocampal and ambient gyri (left)	49	Superior frontal gyrus (left)
11	Superior temporal gyrus, central part (right)	50	Superior frontal gyrus (right)
12	Superior temporal gyrus, central part (left)	51	Postcentral gyrus (left)
13	Middle and inferior temporal gyrus (right)	52	Postcentral gyrus (right)
14	Middle and inferior temporal gyrus (left)	53	Superior parietal gyrus (left)
15	Fusiform (lateral occipitotemporal) gyrus (right)	54	Superior parietal gyrus (right)
16	Fusiform (lateral occipitotemporal) gyrus (left)	55	Lingual gyrus (left)
17	Insula (left)	56	Lingual gyrus (right)
18	Insula (right)	57	Cuneus (left)
19	Lateral remainder of occipital lobe (left)	58	Cuneus (right)
20	Lateral remainder of occipital lobe (right)	59	Medial orbital gyrus (left)
21	Cingulate gyrus, anterior part (left)	60	Medial orbital gyrus (right)
22	Cingulate gyrus, anterior part (right)	61	Lateral orbital gyrus (left)
23	Gyrus cinguli, posterior part (left)	62	Lateral orbital gyrus (right)
24	Gyrus cinguli, posterior part (right)	63	Posterior orbital gyrus (left)
25	Middle frontal gyrus (left)	64	Posterior orbital gyrus (right)
26	Middle frontal gyrus (right)	65	Subgenual frontal cortex (left)
27	Posterior temporal lobe (left)	66	Subgenual frontal cortex (right)
28	Posterior temporal lobe (right)	67	Subcallosal area (left)
29	Inferolateral remainder of parietal lobe (left)	68	Subcallosal area (right)
30	Inferolateral remainder of parietal lobe (right)	69	Pre-subgenual frontal cortex (left)
31	Caudate nucleus (left)	70	Pre-subgenual frontal cortex (right)
32	Caudate nucleus (right)	71	Superior temporal gyrus, anterior part (left)
33	Nucleus accumbens (left)	72	Superior temporal gyrus, anterior part (right)
34	Nucleus accumbens (right)	73	Left Red Nucleus
35	Putamen (left)	74	Right Red Nucleus
36	Putamen (right)	75	Left Substantia Nigra
37	Thalamus (left)	76	Right Substantia Nigra
38	Thalamus (right)	77	Left Subthalamic Nucleus
39	Pallidum (left)	78	Right Subthalamic Nucleus

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

9/7/18

Date