A Mind Without Time: Classifying the Conversion to Alzheimer's Disease

Problem Statement

Alzheimer's Disease (AD) is one of the most prevalent and debilitating forms of dementia. In Canada alone, there are 564,000 people diagnosed with dementia, a number that is expected to increase to nearly a million by 2031[1]. Aside from the impact on an individual, dementia places a large burden on the healthcare system and persons involved with an affected individual. Dementia is currently estimated to cost 10.4 billion dollars in yearly expenses within Canada[1].

Early diagnosis of AD is associated with a higher quality of life and a reduced cost on a healthcare system[2]. However, detecting AD early in the disease progression is difficult due to the multifaceted nature of how neurodegeneration affects the brain, cognitive processing, and behavior[3,4,5]. Clinical evaluation relies on assessment of a myriad of cognitive tests and biomarkers that are not always identifiable in patients with mild cognitive impairment (MCI), a precursor to AD[6].

The multifaceted impact of cognitive impairment and neurodegeneration in MCI and AD suggests that machine learning algorithms such as neural networks may be beneficial in identifying and predicting disease progression. Current studies typically only incorporate one form of data, however, often relying solely on features extracted from structural magnetic resonance imaging (MRI) scans[7]. Other forms of data that show promise in classification with machine learning algorithms include cognitive assessments[8] and the connectivity patterns of resting-state functional networks[9]. This is because spatial and episodic memory, cognitive processes that are typically the first affected in MCI and AD, rely on complex, dynamic interactions of distributed neural networks[10,11] and are therefore susceptible to the impact of neurodegeneration. Critically, there has yet to be an assessment of how machine learning

algorithms perform using features extracted from structural and functional MRI data, as well as cognitive assessments. This project aims to remedy this.

Target audience and use cases:

Healthcare providers. Structural and resting-state functional MRIs are one of easiest and fastest methods of brain imaging. Using them to classify persons at risk or with Alzheimer's Disease would assist in providing targeted treatments.

Dataset:

The dataset is based on the ADNI research initiative.

http://adni.loni.usc.edu/data-samples/adni-data-inventory/

Portions of this initiative have been populated into a single dataframe. This dataframe contains numerous variables of interest:

- Demographic information such as age and gender
- Assessment on cognitive tests
- Volumetric measures of brain regions from structural MRI data
- Measures of functional connectivity in fMRI data
- Diffusion tensor imaging of the hippocampus and entorhinal cortex
- Measures from PET imagining
- Presence of the APOE4 allele

These measures are provided a detailed overview in section 2 of the EDA notebook. Briefly, they are based on research showing that Alzheimer's Disease affects them and therefore may be useful in forecasting patients that convert from cognitively normal or MCI.

Data cleaning:

All data cleaning is contained and summarized in the data_cleaning.ipynb notebook. The steps taken were:

- Rename columns to make them more interpretable and pythonic
- Extract columns of interest from main dataframe
- Merge main dataframe with dataframe containing fMRI measures
- Recode missing values
- Remove or impute missing values
- Reduce the time series column to patient visits with measures of interest
- Deal with missing values in the diagnosis change column
- Identify and clean repeated entries in diagnosis change column

The clean dataset is saved as df_clean.csv in the Github repository.

Exploratory data analysis:

The dataset contains baseline and time series measures for each variable. Each patient has a baseline diagnosis of either cognitively normal (CN), significant memory complaints (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), or Alzheimer's Disease (AD). Demographic information at baseline for the different diagnoses were inspected and visualized.

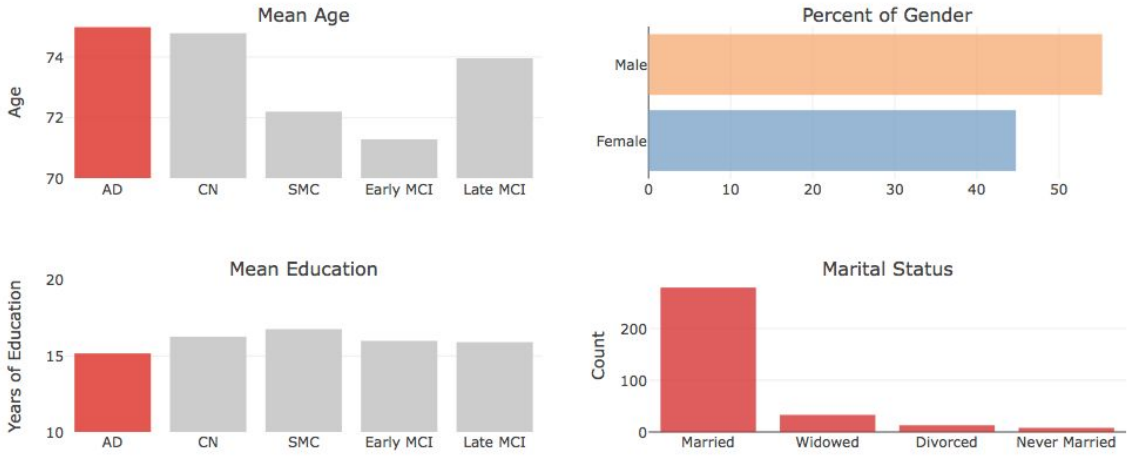


Figure 1. Demographic information for patients with Alzheimer's Disease at baseline.

The average patient with AD at baseline was older than the other diagnosis groups, male, had less education and were married.

The different variables of interest were also visualized and inspected. Variables that did not appear to show differences between AD and the other groups were flagged for removal. The rationale was that if a variable did not differ between AD and the other groups, it would not contain information that is useful in classifying patients with AD. Overall, the cognitive tests and volumetric brain measures appear the most useful, with some measures from DTI and PET imaging showing group differences as well. The presence of the APOE4 allele, especially when there are two copies present, appears useful as well.

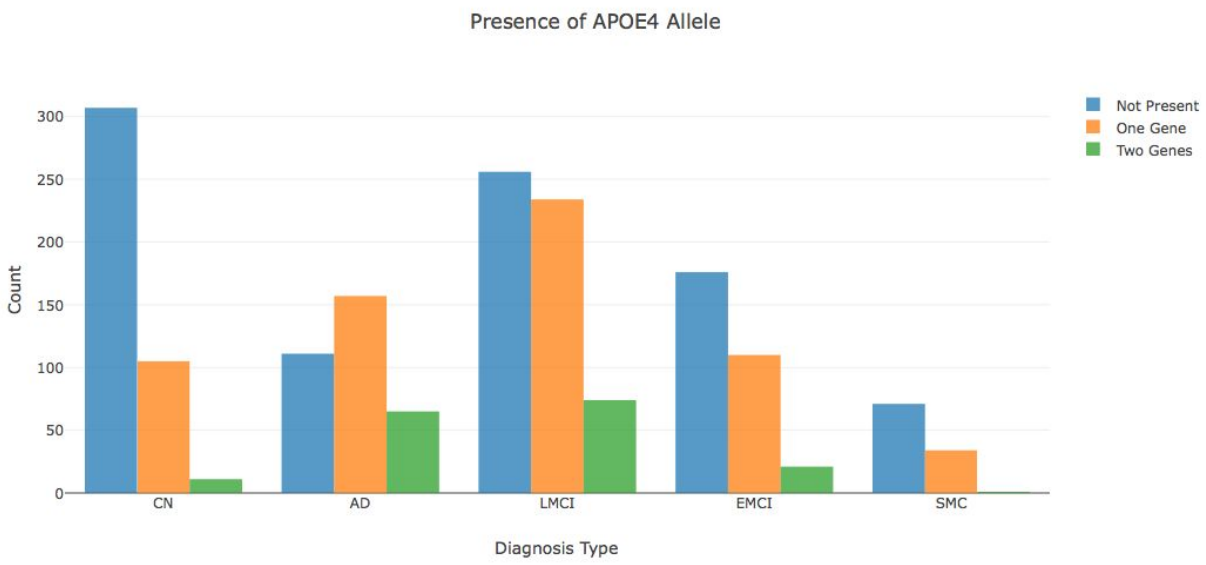


Figure 2. Frequency of APOE4 allele by different diagnosis groups.

The target variable in the project is forecasting the conversion to AD. As such, the diagnostic change variable that codes for each visit whether a patient was stable or converted to a different diagnosis group was inspected.

There are a total of 331 patients out of 1382 patients (24% of sample) that entered the study without AD that subsequently converted to AD at some point. There are 330 patients with AD at the onset of the study.

Figure 3. Overview of demographic information for patients that convert to Alzheimer's Disease during the study.

The average person who converts to AD enters the study with a diagnosis of late MCI, is male, older on average with a university degree, is married and not hispanic/latino.

Differences on variables of interest were also investigated by comparing diagnosis group patients who convert to AD during the study with those that don't. Three diagnosis groups were used: cognitively normal, early MCI, and late MCI. There was only one patient with SMC who converted, and therefore this diagnosis group was omitted from analysis.

Table 1. Results from statistical tests for cognitive tests

Category	Feature	Omnibus	DX baseline	Mean difference
Cognitive Test				
	CDRSB	F = 490*		
			CN	0.02
			EMCI	0.8*
			LMCI	0.52*
	ADAS13	F = 299.2*		
			CN	1.02
			EMCI	3.89*
			LMCI	5.19*
	MMSE	F = 128.8*		
			CN	0.55*
			EMCI	-0.23
			LMCI	-0.72*
	MOCA	F = 53.38*		
			CN	-1.46
			EMCI	-1.23*
			LMCI	-2.34*

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Table 2. Results from statistical tests for volumetric brain measures

Category	Feature	Omnibus	DX baseline	Mean difference
Brain volume				
	Ventricles	F = 18.43*		
			CN	168.4
			EMCI	10421.41*
			LMCI	2273.45*
	Whole brain	F = 20.39*		
			CN	-58638.92*
			EMCI	546.96
			LMCI	-22196.44*
	ICV	F = 9.27*		
			CN	-43421.67
			EMCI	59481*
			LMCI	3742.1
	L HC	F = 100.1*		
			CN	-286.6*
			EMCI	-282.95*
			LMCI	-345.37*
	R HC	F = 96.33*		
			CN	-340.83*
			EMCI	-305.7*
			LMCI	-313.17

L EC	F = 59.64*		
		CN	-149.47
		EMCI	-104.73
		LMCI	-193.39*
R EC	F = 43.59*		
		CN	-149.26
		EMCI	-136.47*
		LMCI	-196.8*
L EC Thick	F = 77.06*		
		CN	-0.13
		EMCI	-0.22*
		LMCI	-0.23*
R EC Thick	F = 79.05*		
		CN	-0.2*
		EMCI	-0.3*
		LMCI	-0.29*

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Table 3. Results from statistical tests for electroencephalography tests

Category	Feature	Omnibus	DX baseline	Mean difference
Ecog Tests				
	Memory	F = 71.01*		
			CN	0.5*
			EMCI	0.21
			LMCI	0.19
	Visuospatial	F = 18.87*		
			CN	-0.1
			EMCI	0.17
			LMCI	0.14

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Table 4. Results from statistical tests on fMRI measures

Category	Feature	Omnibus	DX baseline	Mean difference
fMRI				
	ADMNRV	F = 0.58		
			CN	NA
			EMCI	NA
			LMCI	NA
	PDMNRV	F = 0.6		
			CN	NA
			EMCI	NA
			LMCI	NA
	DMNRVR	F = 0.92		
			CN	NA
			EMCI	NA
			LMCI	NA

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Table 5. Results from statistical tests on DTI measures

Category	Feature	Omnibus	DX baseline	Mean difference
DTI				
	FA L HC	F = 1.63		
			CN	NA
			EMCI	NA
			LMCI	NA
	FA R HC	F = 1.1		
			CN	NA
			EMCI	NA
			LMCI	NA
	MD L HC	F = 7.3*		
			CN	0
			EMCI	0
			LMCI	0
	MOCA	F = 53.38*		
			CN	-1.46
			EMCI	-1.23*
			LMCI	-2.34*

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Table 6. Results from statistical tests on PET measures

Category	Feature	Omnibus	DX baseline	Mean difference
PET				
	AV45	F = 49*		
			CN	-0.01
			EMCI	0.18*
			LMCI	0.22*
	FDG	F = 60.63*		
			CN	-0.05
			EMCI	-0.11*
			LMCI	-0.08*

* denotes statistical significance. Mean differences calculated using Tukey's HSD

The summary of the omnibus and post-hoc tests is that the fMRI and fractional anisotropy measures do not differentiate the converters from non-converters in any of the diagnosis groups. As such, these are dropped from further analysis. The remaining features statistically differentiate converters in at least one diagnosis group, with the cognitive tests and volumetric brain measures being the most prominent.

Another feature that was explored based on preliminary analysis was the correlation between two pairs of cognitive tests, namely the CDRSB and ADAS13, as well as the MMS and MOCA. The initial inspection showed that there may be differences between diagnosis groups in how correlated performance on these assessments are. To explore this, difference scores between the test pairs were calculated and compared, testing putative differences between converters and non-converters.

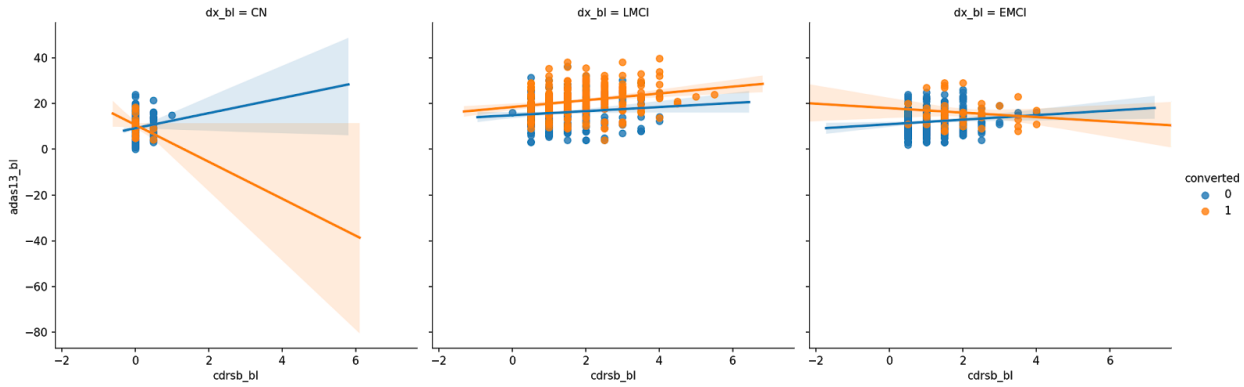


Figure 4. Scatterplots for relationship between CDRSB and ADAS13 scores at baseline for converters vs. non-converters.

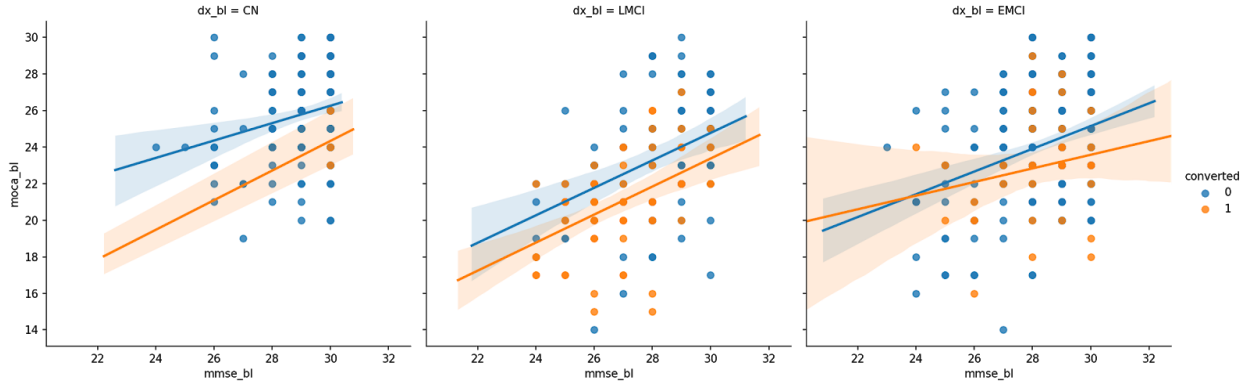


Figure 4. Scatterplots for relationship between MMSE and MOCA scores at baseline for converters vs. non-converters.

Table 6. Results from statistical tests on CDRSB vs ADAS13 difference scores

Category	Feature	Omnibus	DX baseline	Mean difference
Test differences				
	CDRSB vs ADAS13	F = 221.2*		
			CN	-0.1
			EMCI	-3.09*
			LMCI	-4.66*
	MMSE vs MOCA	F = 20.24*		
			CN	2.43
			EMCI	1.04*
			LMCI	1.15*

* denotes statistical significance. Mean differences calculated using Tukey's HSD

Based on these analyses, the difference scores between tests appear to be a useful measure in differentiating patients that convert in the MCI groups.

The resulting dataframe that is used for modelling is available as df_final.csv

Generating a classification model:

The central aim of the project is to develop a classification model that is able to identify persons without AD that are at high risk of developing it. To build the model, the first step is to define a feature set and to encode a binary target label that represents either patients who convert to AD at some point in the study or patients who don't.

In order to retain as many observations as possible, the feature set used included:

- **Demographic information**:
    - Years of education

    - Gender
    - Age at start of study
- **Cognitive assessments**:
    - CDRSB, ADAS13, MMSE, and the difference score between the CDRSB and the ADAS13
- **Structural MRI volumetric measures**:
    - Volume of the left/right hippocampus and entorhinal cortex generated from a cross sectional template
    - Cortical thickness estimates for the left/right entorhinal cortex also generated from a cross sectional template
- **Genetic information**:
    - Number of APOE4 alleles

This feature set represented the data categories that showed promise in differentiating AD from other cognitive types while retaining the most data (the MOCA, for example, was only administered on about half the patient sample).

Converters were identified as patients that converted from either cognitively normal or MCI to AD at some point in the study. There was one patient with significant memory complaints that was dropped as well as six patients who at the baseline visit were diagnosed with AD.

The models were generated by including data points for each feature across all visits in the dataset. For converters, this means that visits before, at, and after conversion to AD were labeled as converted.

The classification models that were evaluated were:

- Logistic regression
- Random Forests
- XGBoost
- Support vector machine

These models encompass simple linear classification, as well as more modern approaches of bagging and boosting.

All models were evaluated using three metrics:

1. Accuracy score on the test data
2. Area under the ROC curve
3. Average precision

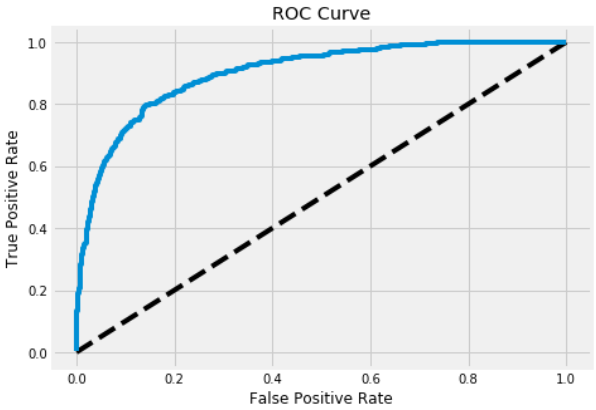
Hyperparameters were optimized using a 5-fold stratified gridsearch that preserved the proportion of convert/non-convert in each fold.

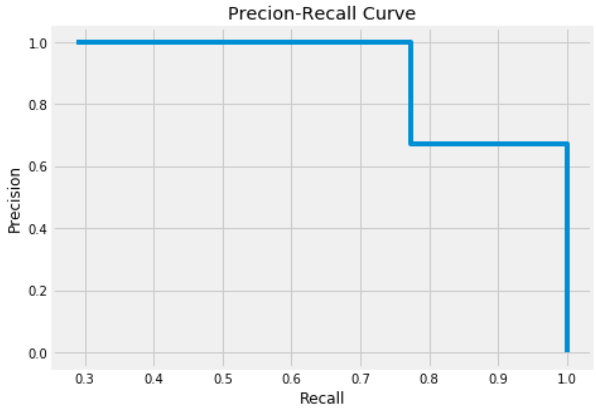
Model evaluation is outlined in the *model.ipynb* notebook.

Classification model performance:

The first model evaluated was logistic regression. After cross validation, the hyperparameters were tuned to

- C = 0.1
- L2 penalty



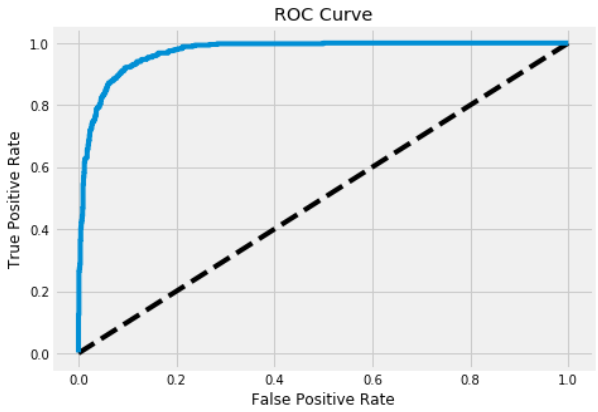


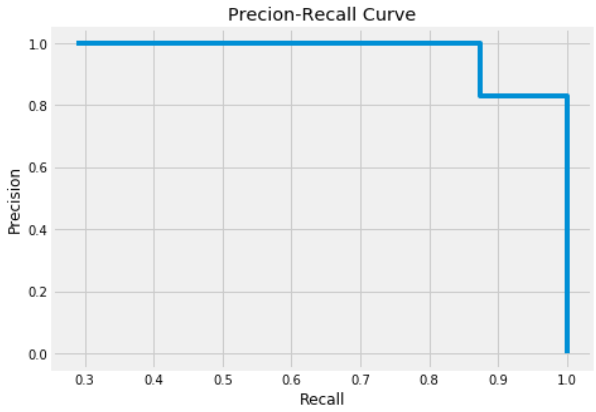
- Accuracy on the test data is 0.848
- The model AUC for ROC curve of the test data is 0.901
- Average precision is 0.613

The model performed reasonably well but had a low average precision. Clinician accuracy in diagnosing AD is around 90% so ideally the final model would perform better than that, although the task isn't to diagnose AD but to identify persons at risk of developing it at some point in the future. As such, the precision-recall curve is particularly important as it incorporates information about false negatives, which in the present study are patients who do convert to AD but the model labeled as a non-converter.

The next model evaluated was a Random Forest classifier. After cross validation, the hyperparameters were tuned to

- n_estimators = 1788
- Min_samples_split = 2
- Min_samples_leaf = 1
- Max_depth = 20
- Bootstrap = False



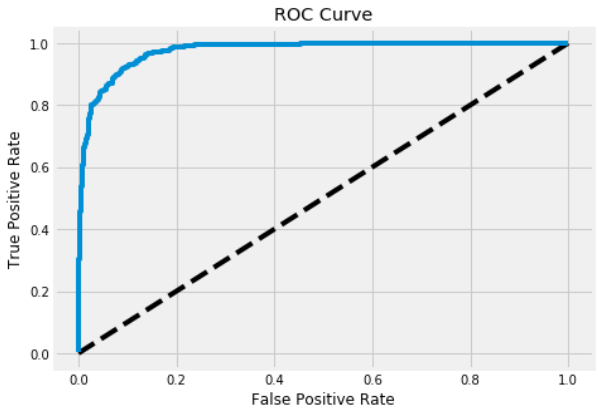


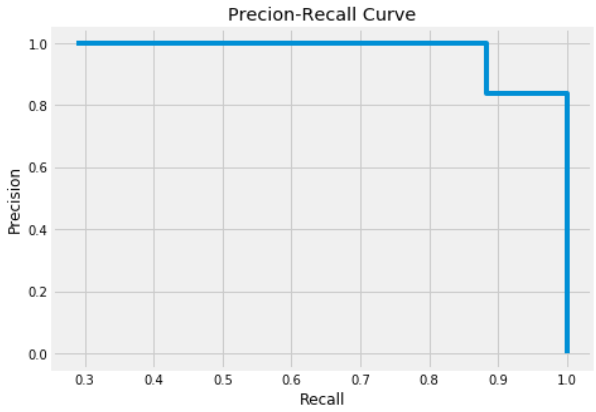
- Accuracy on the test data is 0.916
- The model AUC for ROC curve of the test data is 0.971
- Average precision is 0.773

This model performed much better than the logistical regression one across all three metrics. However, the accuracy on the training data was 100% indicating that the model overfit the training data.

Next, a popular boosting algorithm XGBoost was used. After cross validation, the hyperparameters were tuned to

- Colsample_bytree = 0.7
- Learning_rate = 0.09
- Max_depth = 6
- Min_child_weight = 11
- N_estimators = 1000
- Objective = binary:logistic
- Silent = 1
- Subsample = 0.8



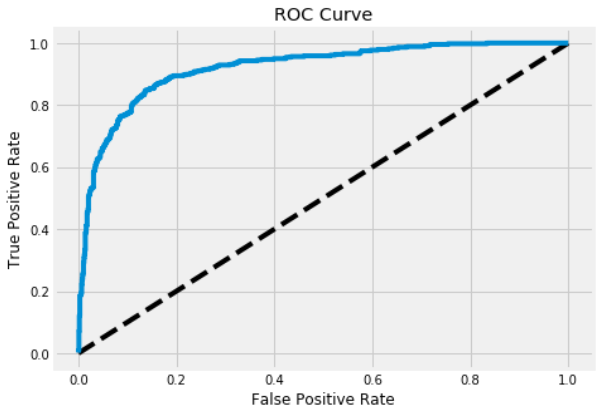


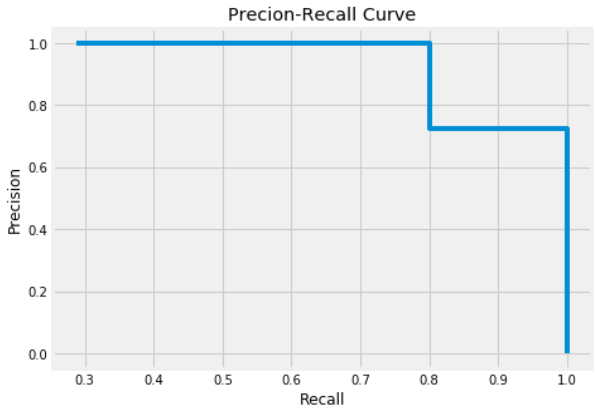
- Accuracy on the test data is 0.921
- The model AUC for ROC curve of the test data is 0.975
- Average precision is 0.787

Overall, the model performs better than the Random Forest classifier across the three metrics. However, the accuracy score on the training data was also 100% indicating that the model was overfit to the training data which may decrease its generalizability.

Lastly, a model using a support vector machine was used. After cross validation, the hyperparameters were tuned to

- C =
- Gamma =
- Class weight =





- Accuracy on the test data is 0.868
- The model AUC for ROC curve of the test data is 0.92
- Average precision is 0.659

The model did not perform as well as the XGBoost classifier across any of the three metrics.

Summary

The best performing model used the XGBoost classifier. This model had a high accuracy score of 92%, with a large area under the curve (0.975) and a fairly high average precision of 0.787. Importantly, the precision recall curve shows that it has high recall as well, indicating that there are few false negatives, patients that the classifier mislabeled as non-converts.

The use case of the model is to assist physicians in identifying patients at high risk for developing AD. This will allow early intervention and preventative therapies to by used by the patient to potentially decrease the likelihood of converting to AD.

Next steps

The final model was trained on a limited set of features in order to maximize the observations during training. However, it may be that some of the omitted features have high influence in classifying converts from non-converts. Model evaluation should be performed on a reduced dataset that includes PET, fMRI and the MOCA, and feature importance looked at to assess whether there is need to collect additional data from the data category.

Second, the two top performing models also overfit the training data, potentially limiting the ability to generalize to the test data and other unseen data. In order to prevent overfitting, additional features such interactions between feature sets may be included, or a different score metric for the gridsearch cross validation to use.

Third, the model currently indicates whether a person is at risk for converting to AD at some point. It may be possible to forecast the timeframe of a conversion. This would not impact the amount of time between assessment and conversion to AD, but would give an indication of the urgency for preventative therapies that may be useful to hospitals or care givers that have limited resources.

References

[1]http://alzheimer.ca/en/Home/Get-involved/Advocacy/Latest-info-stats

[2]Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2016). Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. *Journal of Alzheimer's Disease, 49(3),* 617-631.

[3]Conde-Sala, J. L., Garre-Olmo, J., Vilalta-Franch, J., Llinàs-Reglà, J., Turró-Garriga, O., Lozano-Gallego, M., et al. (2012). Predictors of cognitive decline in Alzheimer's disease and mild cognitive impairment using the CAMCOG: a five-year follow-up. *International Psychogeriatrics, 24(06)*, 948–958.

[4]Matsuda, O., & Saito, M. (2009). Multiple cognitive deficits in patients during the mild cognitive impairment stage of Alzheimer's disease: how are cognitive domains other than episodic memory impaired? *International Psychogeriatrics, 21(05)*, 970-976.

[5]Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., et al. (2009). Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer's Disease. *Journal of Neuroscience, 29(6)*, 1860–1873.

[6]Morris, J.C., & Cummings, J. (2005). Mild cognitive impairment (MCI) represents early-stage Alzheimer's disease. J*ournal of Alzheimer's Disease, 7*, 235–239.

[7]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.

[8]Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage, 104*, 398-412.

[9]Khazaee, A., Ebrahimzadeh, A. & Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer’s disease. *Brain Imaging and Behavior, 10(3),* 799-817.

[10]Arnold, A.E.G.F., Ekstrom, A., & Iaria, G. (2018). Dynamic neural network reconfiguration during the generation and reinstatement of mnemonic representations. *Frontiers in Human Neuroscience, 12,* 292.

[11]Ekstrom, A., Arnold, A.E.G.F., & Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. *Frontiers in Human Neuroscience, 8,* 803.