HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 2 (2023)

Using Machine Learning to Forecast Progression from Cognitively Normal to Alzheimer's Disease

Ann Song

Choate Rosemary Hall

ABSTRACT

Alzheimer's Disease (AD) affects approximately 50 million individuals worldwide and is estimated to rise to 152 million by 2050. There is currently no treatment for AD that halts the progression from cognitively normal (CN) and/or mild cognitive impairment (MCI) to AD. The ability to predict disease progression will allow for early treatment. While Machine Learning (ML) has been successful in diagnosing the cognitive state, further improvement is necessary for predicting progression. In this study, Random Forest and Bagging Decision Tree Recursive Feature Elimination (RFE) was utilized to ascertain the cognitive state and forecast progression. Clinical diagnoses, demographics, and post-processed PET and MRI scans used in this study were obtained from the Open Access Series of Imaging Studies (OASIS). The findings suggest that aging and lower levels of education are associated with higher risk. The study found that ML using post-processed MRI and PET scans, particularly RFE ML, is effective in diagnosing cognitive states with 90% accuracy. It can predict progression from CN to MCI or AD with 85% accuracy, which is significantly higher than the average reported in literature. Patients with progression from CN to AD were distinguished by elevated amyloid deposition, hippocampus and amygdala atrophy, left accumbens atrophy, thinning of the left hemisphere temporal, and enlarged inferior lateral ventricles. The study demonstrated that RFE ML is effective in diagnosing and predicting the progression of AD. Future studies will concentrate on identifying the specific regions of amyloid plaque that have the most significant impact on cognitive state and progression.

Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia, characterized by a gradual and progressive deterioration of the nervous system. There are around 50 million AD patients worldwide and this number is projected to double every 5 years. By 2050, there will be 152 million patients globally [1] [2]. Individuals with AD experience a decline in cognitive function, which is characterized by symptoms such as memory loss, difficulty concentrating and thinking, and impaired reasoning and decision-making abilities. Unfortunately, there is currently no cure for AD, and the FDA has not approved any medications for treating AD that have been demonstrated to delay or prevent its progression [3]. One of the main reasons that finding treatment for AD is challenging is because AD cannot be easily identified at early stages of development [4] [5].

AD is characterized by neuritic plaques and neurofibrillary tangles [6]. To diagnose AD, several tests need to be performed, including neurological examination, Magnetic Resonance Imaging (MRI) for neurons, laboratory examinations, and analysis of medical and family history. Recently, amyloid Positron Emission Tomography (PET) imaging uses agents to bind to insoluble fibrillar forms of Aβ 40 and Aβ 42 deposits, a major component of compact neuritic plaques and vascular deposits, has been utilized extensively to characterize the deposition of amyloid plaque [7]. A probable diagnosis of AD can be established with a confidence of $> 90\%$ by the methods mentioned above [6]. However, a method to predict progression from cognitively normal (CN) to mild cognitive impairment (MCI) or then to AD is still not available.

The increase in processing power of Graphics Processing Units (GPUs) has enabled the development of cutting-edge deep learning algorithms, which have been applied to the diagnosis of AD patients with great success in recent years [8]. Machine learning studies have reported average accuracies of 92%, 83%, 80%, and 79% for distinguishing between CN and AD, MCI and AD, CN and MCI, and multi-class (CN, MCI, AD) groups, respectively [8] [9] [10] [11]. While some studies have focused on using machine learning to predict progression to AD, the majority have concentrated on forecasting progression from MCI to AD, with an accuracy of 76% [8]. Few or no studies have attempted to predict the progression from CN to AD. Therefore, further research is needed to improve the accuracy of predicting progression from MCI to AD, or from CN to MCI or AD. In a study by Albright [12], neural network machine learning was utilized to predict the progression from MCI or CN to AD with an accuracy of 86.6%. The study incorporated the 13-item Alzheimer's Disease Assessment Scale and several verbal learning tests as primary features for prediction, with ventricular volume serving as the sole biomarker. It provided an effective method for predicting the progression, but it was focused mainly on the outcome with minimal information on the underlying physiological mechanisms that led to the behavioral differences. The application of machine learning to AD biomarkers can enhance our understanding of changes in brain structure and physiology, facilitating early treatment as well as providing a foundation for doctors and researchers to develop effective medicines or therapies for early intervention.

This study uses Recursive Feature Elimination (RFE) machine learning to identify the biomarkers that can be used to diagnose the cognitive state (CN, MCI and AD) and forecast the progression from CN to AD. RFE machine learning is effective at selecting features in a training dataset that are the most relevant in predicting the target variable [13]. It is hypothesized that machine learning can be applied to MRI and PET scans of the brain to determine cognitive state and assess the risk of progression from CN to AD.

Method

This study utilized brain volume and thickness measurements of various brain regions based on MRI images, along with global amyloid burden data based on amyloid PET imaging, obtained from the Open Access Series of Imaging Studies (OASIS-3) to train Random Forest and Bagging Decision Tree (RFE) machine learning models. The overview of the method is demonstrated in Figure 1, including three steps: pre-processing, modeling and testing. Through pre-processing, the relevant OASIS-3 data was imported and subjected to dimension reduction to eliminate strongly correlated features and reduce the risk of overfitting. All the data was then merged, cleaned, and partitioned to prepare the data for modeling, which would identify the important features and predict the diagnosis of the training set. Finally, the model was tested by applying it to the test set for accuracy.

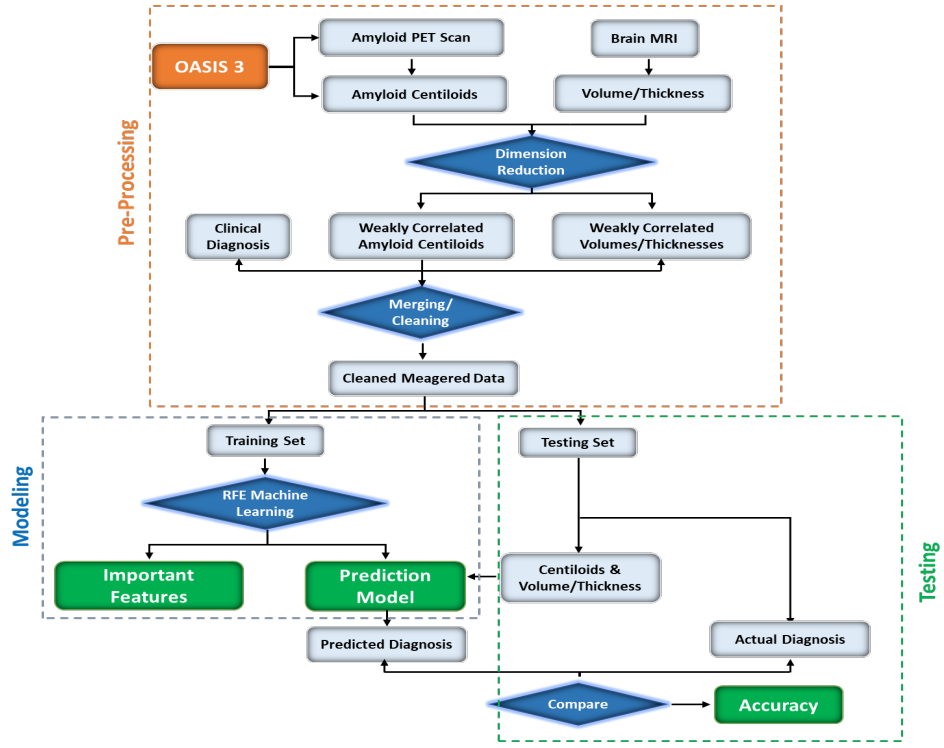


Figure 1. Overview of the method.

About Open Access Series of Imaging Studies (OASIS)

OASIS-3

All the data was obtained from OASIS by the Washington University Knight Alzheimer Disease Research Center over the course of 15 years. There are five independent datasets in OASIS websites. The data used in this study was OASIS-3 data, which is longitudinal and includes 2842 MR sessions, 2157 PET sessions, and 1472 CT sessions from 1379 participants. The participants include 755 cognitively normal adults and 622 individuals at different stages of cognitive decline, ranging in age from 42 to 95 years. For more detailed statistics of OASIS-3, please refer to the literature [14]. The OASIS-3 data is hosted on XNAT Central (https://central.xnat.org), a publicly accessible data-repository.

MRI Images and FreeSurfer Post-Processing

In the OASIS-3 dataset, MR images were post-processed by cortical reconstruction and volumetric segmentation of T1-weighted images using the FreeSurfer image analysis suite, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The technical details of these procedures are described in previous publications [15] [16] [17]. Data included in OASIS-3 was processed using an XNAT pipeline for the FreeSurfer image analysis suite using Dell PowerEdge 1950 servers with Intel Xeon processors running CentOS 5.5 Linux [14].

PET Scan, PUP Post-Processing, and Centiloid Standardization

Participants underwent Positron Emission Tomography (PET) on one of three different Siemens scanners: ECAT HRplus 962 PET scanner, Biograph 40 PET/CT scanner, and Biograph mMR PET-MR. Pittsburgh Compound B ([^{11}C]PIB or PIB) and Florbetapir [^{18}F] (^{18}F-AV-45 or AV45), were used to investigate β-amyloid (Aβ) deposits in the brain. PET imaging analyses were performed using the PET unified pipeline (PUP) (https://github.com/ysu001/PUP) via XNAT [18] [19]. More details about post-processing of PET images, including PET to MR registration is summarized in OASIS-3 literature [14]. The Centiloid scale was developed to standardize the quantification of amyloid by converting amyloid PET data to a 0-100 scale [20]. The Centiloid scale has been used to calibrate both PIB and Florbetapir imaging, with equations presented in literature [21].

Data Cleaning and Exploration

Based on the goal of the study, four datasets from OASIS-3 were used:

- Demographics: this file includes patients' basic information including patient ID, Age at Entry, Gender, Years of Education, and preferred used hand (handedness). These were explored to understand the correlation between the cognitive state and patient demographics information.
- Clinician Diagnosis of Cognitive Status and Dementia: the data includes patients' age and the clinical diagnosis results at different visits. One patient may have multiple visits. The first entry was recorded to be Day 0, and the dates for the remaining visits were recorded to be the number of the days from the first visit. The dataset includes detailed information about the cognitive state of the patient at each visit with options including Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Dementia (DMN). For those who had dementia, types of dementia were further categorized. To simplify the analyses, assuming all the types of dementia were categorized to be Alzheimer's Disease (AD).
- Global Amyloid Burden: this file includes the mean cortical binding potential (BP) or mean cortical standard uptake ratio (SUVR) that was calculated from the arithmetic mean of BP or SUVRs of four critical regions from PET scans with and without partial volume correlations: precuneus (PREC), prefrontal cortex (PREF), gyrus rectus (GR), and bilateral temporal (TEMP) [14] [19]. Pearson Correlation Coefficients were calculated among these four variables, and only the variables that had moderate or weak correlations (≤ 0.59) were kept for Machine Learning analyses [22]. As a result, mean cortical SUVR was kept to be merged with the FreeSurfer output for Machine Learning analyses.
- FreeSurfer: the file includes the volume and thickness of various portions in the brain calculated by FreeSurfer from post-processing of MRI images of the patients. For more details on the post-processing, please refer to the FreeSurfer Wiki (https://surfer.nmr.mgh.harvard.edu/) and relevant publications [15] [16] [17] [14] [23]. Similar to Amyloid Burden analyses, only the variables with moderate or weak correlations were kept for Machine Learning analyses.

General Statistics about the patients and visits were calculated. Paired t-tests were performed to identify potential significant differences in age, years of education, amyloid centiloid, and total brain volume among three groups (CN, MCI and AD).

Data Merging

The patient ID and the visit time were used to map the patients. The patients' IDs are in the format of "OAS3xxxx", in which "xxxx" represents a 4-digit number. The time of visit was in the format of the number of days from the first visit. To account for variations in the timing of doctor visits and PET/MRI scans, the visit time was rounded to the nearest year from the first visit, allowing for better mapping of the visits. After that,

the files were merged by patient ID and the year of visit. To prepare for machine learning, missing data were removed from the merged data.

Machine Learning

Overview

Random Forest and Bagging Decision Tree of Recursive Feature Elimination (RFE) Machine learning was performed on the merged data from the Clinical diagnoses, Amyloid Centiloid and FreeSurfer values. RFE has the advantage of mitigating overfitting, especially on features with large dimensions. To avoid the exclusion of important features, the Recursive Feature Elimination (RFE) was conducted with a range of feature numbers including 1-20, 30, 40, 50, 60, 80, and the complete set of all features.

Two types of Machine Learning models were executed:

- Diagnosis ML: used Centiloid measurements from PET scans and FreeSurfer measurements from MRI scans to predict the diagnosis. This will enable machine learning based guidance for diagnosis.
- Progression ML: used Centiloid measurements from PET scans and FreeSurfer measurements from MRI scans to predict the progression from CN to MCI or AD.

For all the machine learning models, the data was divided into a training set (80%) and a test set (20%). The accuracy of the model was evaluated by means of comparison between the prediction and the real results (1 if it is the same and 0 if it is not). Important features for all the RFE models were identified. The most important features were defined to be the features that were identified from all RFE models. The top ten important features determined by each model were summarized and compared.

Diagnosis ML

The purpose of this machine learning model is to use the Mean Cortical Amyloid Centiloid calculated from PET scans as well as the brain volume and thickness calculated from MRI scans to predict the cognitive state. The output of the model is a categorical result of one of the following groups: Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). For Diagnosis ML, each visit that has clinical diagnosis information, and a PET and MRI scan, is counted as one data point.

Progression ML

The purpose of this machine learning model is to use the Mean Cortical Amyloid Centiloid calculated from PET scans as well as the brain volume and thickness calculated from MRI scans to predict if a patient is going to progress from CN to MCI or AD. A new factor variable (Progression Status) was assigned to each patient according to the patients' clinical history: CN at Day 0 with no progression later (CN), AD at Day 0 (AD), and progression from CN (at Day 0) to MCI or AD (PROG). The output of the prediction is a categorical result of one of those three groups.

To simplify the analyses, it was assumed that once a patient progresses from CN to MCI or AD, it does not reverse. If the raw data shows reversal, it would be interpreted as a mistake and the data would be corrected so that no patients have a reverse change from a cognitive worse status to a better status. For example, if the raw data showed a patient has CN, then MCI then back to CN, the patient's Progression Status is assigned to PROG.

As the progression was usually diagnosed at one of the follow-up visits, it is unclear which point is a good time for the machine learning model to predict the progression. Therefore, two experiments were performed:

- Experiment 1: used the PET and MRI scans at Day 0 for all the patients as the input for the machine learning model. The advantage of this approach is that if the Day 0 data gives an accurate prediction, it allows for diagnosis much sooner instead of waiting for more scans while the disease gets worse. The disadvantage is that each patient had only one data point as the data for the rest of visits were not used; therefore, the size of the data is reduced. There is also a risk that the Day 0 data may not be significant enough yet to show the signs of the progression.
- Experiment 2: used the PET and MRI scans of all visits of each patient for machine learning. For patients with stable CN or AD, all visits of the same patient are assigned to the same status, with CN or AD. For the patients that had PROG, only visits that have CN were included and assigned to PROG. The advantage of doing this is the dataset would have much bigger size than that of Experiment 1.

Results

Basic Patients Stats and Demographics

After the data was cleaned, there were 1336 participants in the OASIS-3 dataset, with 941 and 1316 patients receiving PET and MRI scans, respectively. There were 894 patients in total that had clinical diagnoses, PET scans and MRI scans in the merged file excluding the data that was removed. The detailed patient information and associated number of visits were summarized in Table *1*.

Table 1. Basic Patient Information.

Variables	All	CN	MCI	AD
Number of Visits	7653	5799	1369	485
Male	0.449	0.699	0.081	0.22
Female	0.551	0.806	0.049	0.145
Age (Years)	74.6 +/- 8.3	73.5 +/- 8.1	77.3 +/- 7.7	78.6 +/- 7.9
Education (Years)	15.8 +/- 2.6	16.0 +/- 2.5	15.6 +/- 2.7	15.2 +/- 3.0
Left-Handed	0.096	0.784	0.058	0.158
Right-Handed	0.904	0.758	0.064	0.178
Duration of Follow Up (Years)	3.8 +/- 4.5	4 +/- 4.5	4.2 +/- 5.2	2.1 +/- 3.6
Number of PET	1585	1358	70	157
Number of MRI	2164	1742	127	295

The t-test results for age and years of Education among groups of CN, MCI and AD are shown in Table 2.

Table 2. The p values of the t-tests.

Variables	CN and MCI	CN and AD	MCI and AD
Age	2.2E-16	2.2E-16	1.6E-03
Education	8.6E-04	2.2E-16	3.7E-03

The following observations were made:

In the participants of OASIS-3, females had fewer MCI and AD than that of male patients

The t-test showed significant differences in age and years of education of the patients with cognitive state of CN, MCI and AD, suggesting that the risk of cognitive impairment increases with aging and decreases with more years of education.

No relationship was found between the handedness and cognitive state.

Correlation Among Variables

Correlations among all the variables for Centiloid and FreeSurfer files were determined for dimension reduction. For the amyloid centiloid, the mean cortical binding potential (BP) or mean cortical standard uptake ratio (SUVR), all showed high correlation with each other (Figure 2). Therefore, only one amyloid centiloid, fSURV_rsf was used for machine learning. There were originally 203 variables in the FreeSurfer file that included the volume and thickness. After removing the variables that have strong correlations, 85 features were kept.

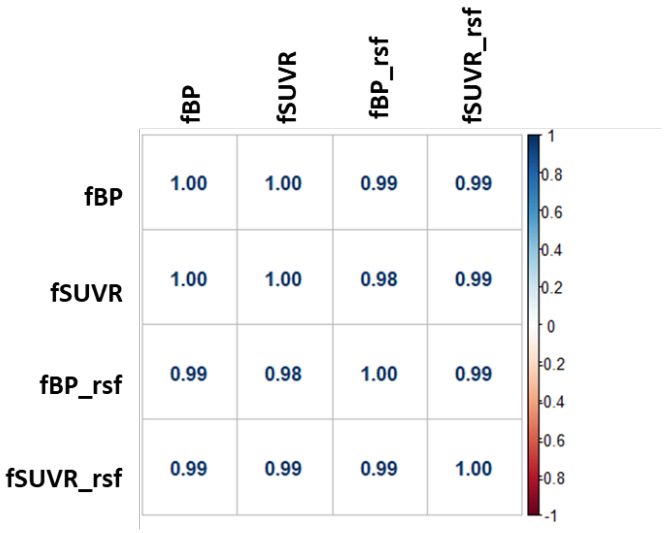


Figure 2. Correlation among mean total cortical amyloid centiloids, where fBP: FreeSurfer calculated Binding Potential; fBP_rsf: FreeSurfer calculated, partial volume corrected Binding Potential; fSUVR: FreeSurfer calculated SUVR; and fSUVR_rsf: FreeSurfer calculated, partial volume corrected SUVR, the gold standard *[24]*.

Progression Statistics

The time taken to progress is demonstrated in Figure 3. As shown, there is no significant difference between the time it takes to progress from CN to MCI and the time it takes to progress from CN to AD. Significant differences existed between progression types of CN to AD and MCI to AD, and between CN to MCI and MCI to AD.

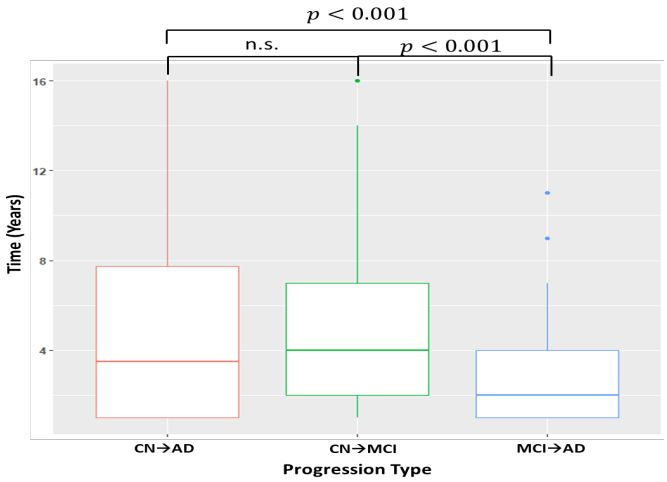


Figure 3. The boxplot showing the time takes to progress (Years), where "n.s" : not significant; CN□AD: Progression from CN to AD; CN□MCI: Progression from CN to MCI; and MCI□AD: Progression from MCI to AD.

Diagnosis RFE Machine Learning

The accuracy of Random Forest and Bagging Decision Tree vs number of features was shown in Figure 4, indicating that:

Using 17 and 18 variables yielded the best predictions for Ransom Forest and Bagging Decision Tree, respectively. Both observations suggested that using all 85 variables would result in overfitting.

For both methods, using the first 5 -10 variables could lead to an accuracy close to the most accurate ones.

For both methods, using one variable could achieve around 84.5% accuracy. Adding a second variable and/or third one significantly improved the accuracy.

Both methods resulted in about the same average accuracy for the training set.

Random Forest yielded a slightly higher prediction accuracy than Bagging Decision Tree did.

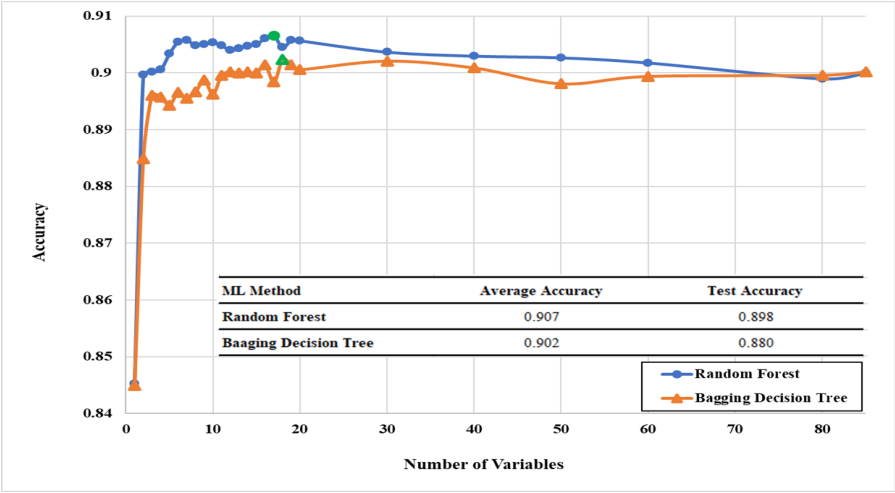


Figure 4. RFE Machine Learning accuracy using Random Forest and Bagging Decision Tree for diagnosing the cognitive state. The curve shows the accuracy of the prediction as a function of the number of variables.

The larger green marker on each curve indicates the best accuracy for each method. The table in the figure shows the comparison of the average accuracy and test accuracy between Random Forest and Bagging Decision Tree method.

The important features identified by both methods is shown in Figure 5, suggesting that the mean total cortical Amyloid Centiloid is the most important feature in predicting cognitive status. The volumes of Total Hippocampus left Amygdala, right Amygdala, left Accumbens Area, and Right Inferior Lateral Ventricle as well as the thicknesses of Left Hemispherical middle and inferior Temporal are among the top ten important features.

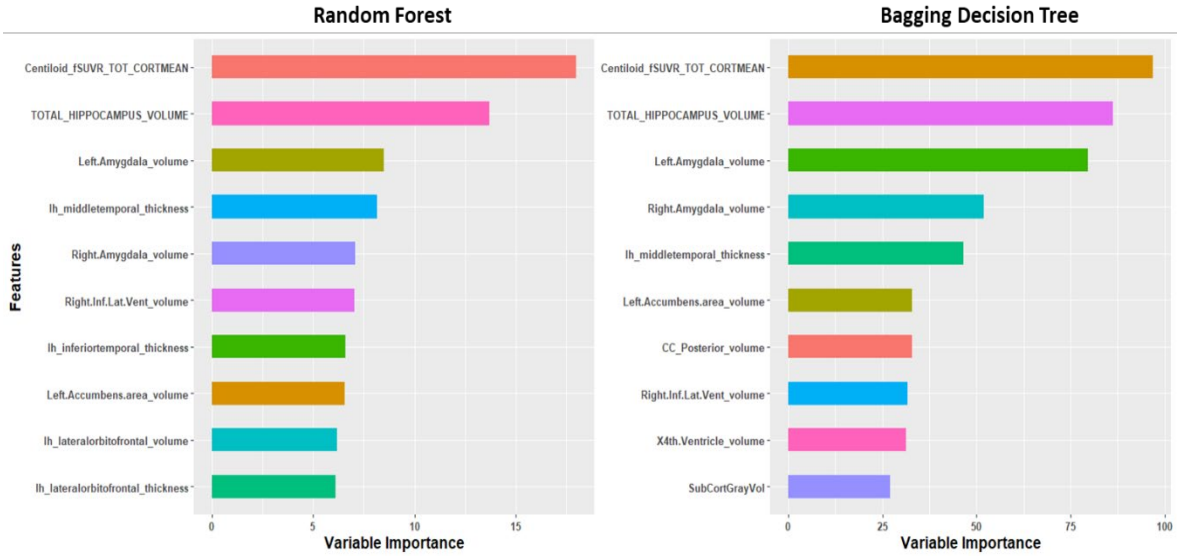


Figure 5. The top 10 important features identified by RFE of Random Forest and Bagging Decision Tree methods for diagnosing the cognitive state.

To confirm the important features are different among different cognitive states, boxplots of Right Inferior Lateral Ventricle Volume and Right Amygdala Volume of CN, MCI and AD groups were plotted and shown in Figure 6. MCI and AD are associated with increased volume of Right Inferior Lateral Ventricle Volume and decreased volume of Right Amygdala. The t-tests showed there are significant differences among all three groups with $p < 0.05$.

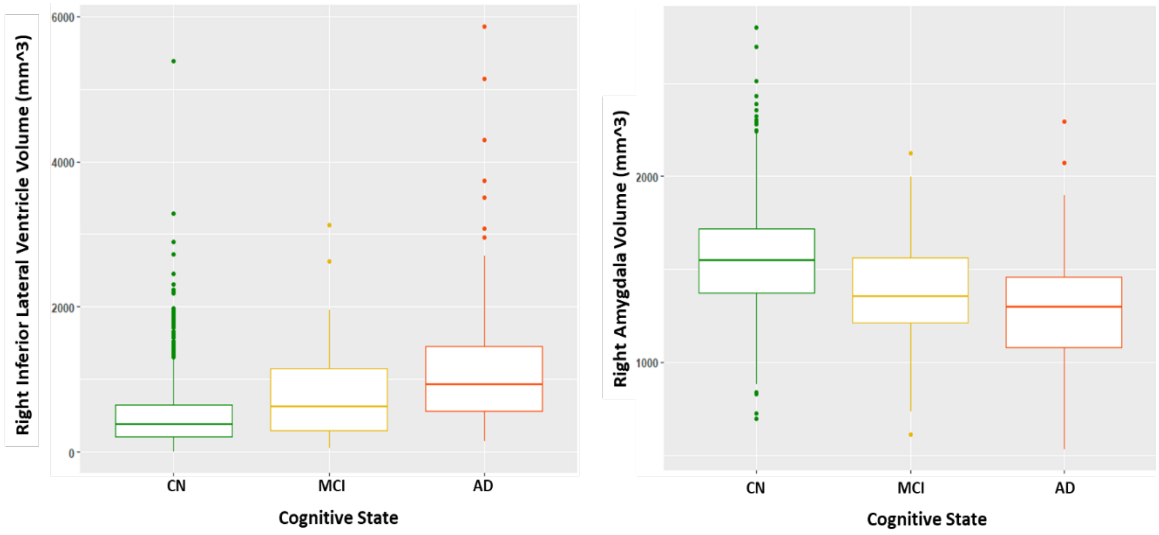
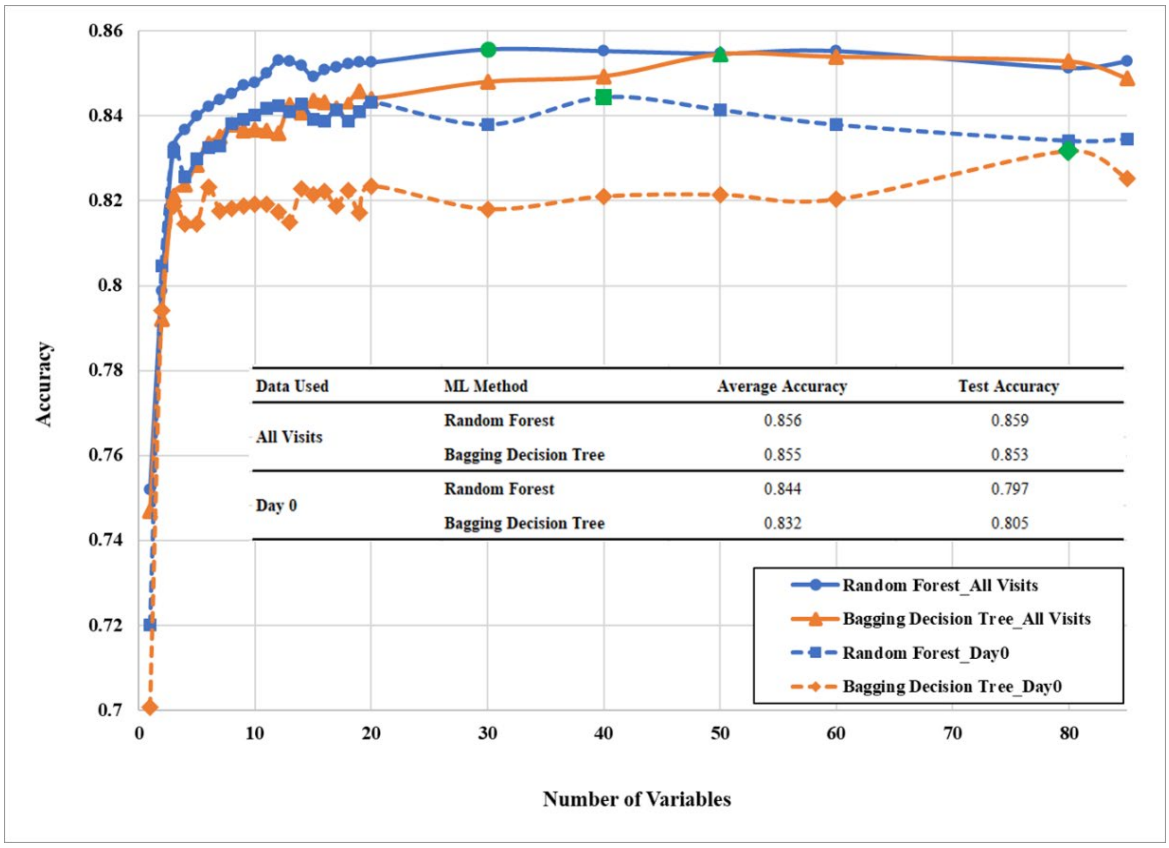


Figure 6. Boxplots of volume of Right Inferior Lateral Ventricle and Right Amygdala.

Machine Learning to Forecast Progression

The RFE results with Random Forest and Bagging Decision Tree results for progression from CN to MCI or AD are shown in Figure 7, which demonstrates that using the data from all visits resulted in a better accuracy than using the data from Day 0 only. It is also observed that it takes more features to get the best prediction using just Day 0 data, as expected. There was no significant difference between Random Forest and Bagging Decision Tree method, although Random Forest gave a slightly higher accuracy.



Data Used	ML Method	Average Accuracy	Test Accuracy
All Visits	Random Forest	0.856	0.859
	Bagging Decision Tree	0.855	0.853
Day 0	Random Forest	0.844	0.797
	Bagging Decision Tree	0.832	0.805

Figure 7. Results of RFE Machine Learning for progression from CN to AD. "_All Visits" means using the data from all visits of the patients; "Day 0" means only the data from the first day of visit is used. The larger green marker in each curve indicated the best accuracy for each method.

The important features identified by all the methods are quite similar, especially the top 5 ones, as shown in Figure 8, which again suggested a very similar pattern to what is seen in diagnosis prediction. In order to verify whether the crucial features for disease progression differed across the CN, AD, and PROG groups, boxplots were generated for the Mean Cortical Amyloid Centiloid and Volume of Hippocampus, as shown in Figure 9. Compared to the CN group with no progression, the group with progression (PROG) shows a reduced volume of Hippocampus and increased Mean Cortical Amyloid Centiloid. As expected, the AD group showed the highest amyloid deposition and the smallest volume of Hippocampus. The t-tests showed there were significant differences among all three groups with $p < 0.05$. Table 3 presents a detailed comparison between CN and PROG, along with the corresponding t-test results.

Figure 8. Top 10 important features identified by RFE machine learning to predict the progression from CN to MCI or AD.

Summarized Important Features

Among all the 85 features, the following features have been shown to be top features by all 6 models: Mean total cortical Amyloid Centiloid, Volume of total Hippocampus, Volume of Left Amygdala, Volume of Right Amygdala, Volume of Left Accumbens Area, Volume of Right Inferior Lateral Ventricle, Thicknesses of Left Hemispherical Middle Temporal, and Thicknesses of Left Hemispherical Superior Temporal.

The following two variables have been suggested by 5 models: Thicknesses of Left Hemispherical Inferior Temporal and Volume of Subcortical gray Matter.

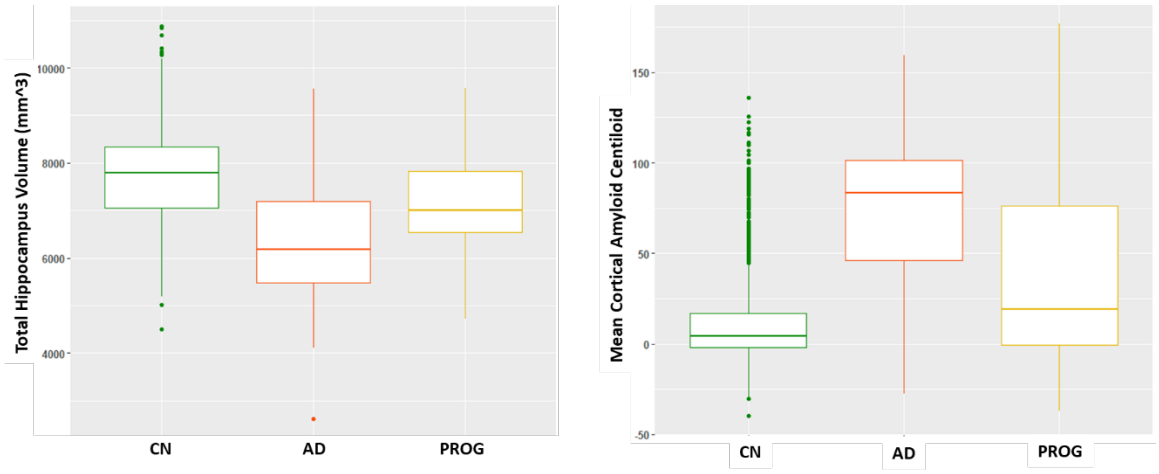


Figure 9. Differences in Total Hippocampus volume and Amyloid Centiloid amount. "CN", "AD" and "PROG" represent the patient group that stays on CN, AD and progresses from CN to AD.

Table 3. CN that does not progress to AD vs CN that progresses to AD. The unit for volume and thickness is mm^3 and mm, respectively. The * suggests no significant difference with $p < 0.05$.

Variables	CN	Prog	p value
Mean Cortical Amyloid Centiloid	13.42	36.86	1.97E-15
Left Accumbens Area Volume	510.40	452.85	2.22E-13
Left Amygdala Volume	1493.96	1377.98	2.32E-12
Left Hemisphere Middle Temporal Thickness	2.72	2.67	7.92E-08
Left Hemisphere Superior Temporal Thickness	2.61	2.54	1.03E-11
Right Amygdala Volume	1561.05	1481.45	2.60E-06
Right Inferior Lateral Ventricles Volume	443.35	736.41	1.76E-15
Total Hippocampal Volume	7732.71	7101.62	2.20E-16
Right Hemisphere Inferior Temporal Thickness	2.73	2.69	2.77E-04
SubCortical Gray Volume	54023.31	55116.13	0.4456*

Discussion

In this study, Random Forest and Bagging Decision Tree RFE was used to (1) diagnose the cognitive state, i.e., CN, MCI, and AD; (2) forecast the progression from CN to MCI or AD. The results from RFE Machine Learning were effective in diagnosis and predicting progression with an accuracy of about 89.8% and 85.9% (Figure 4 and Figure 7), respectively. This is significantly better than the previously reported average, of 79% and 76% for diagnosing CN, MCI, and AD, and 76% for predicting progression from MCI to AD [8] [25].

The present study, for the first time, identified a complete list of important biomarkers for diagnosing cognitive state and progression from CN to MCI or AD. All the features identified by the machine learning were consistent with previous research, which independently suggested that CN patients with a high risk of

progression to AD showed increased mean cortical amyloid plaques [6], atrophy on the hippocampus and Nucleus Accumbens [26] [27] [28] [29], left and right amygdala [30] [31] [27] [28], thinning of Left Hemispherical Middle and superior Temporal [29], and enlarged Right Inferior Lateral Ventricle [32]. All the previous studies have been focused on a specific brain portion, either using statistical methods to compare CN, MCI, and AD groups to determine the difference, or using Machine Learning that uses a certain amount of the features [12]. The present study uses all the available features calculated from MRI and PET scans to identify the important features without making assumptions about the importance of any portion of the brain. Such an approach allows a comprehensive evaluation of all possible MRI-based biomarkers, and therefore, enables the identification of a complete list of important features for diagnosis and assessment of progression to AD at an early stage.

It is concluded that aging and lower education increases the risk of AD, but gender and handedness do not play significant roles in the development of AD. Based on the demographical data from OASIS3, the risk of cognitive impairment increases slightly with age and decreases slightly with longer years of education, which is consistent with previous studies [6] [33]. Within the OASIS3 data, there were more females than males involved, however, more males were reported to have MCI and AD (Table 1). However, there is no statistical evidence showing the effect of gender on the risk of MCI and AD, which is similar to the conclusions from the literature [33], although Mielke [34] suggested that due to longer lifespan, women have a higher lifetime risk of AD. The present study also suggested that handedness does not play a significant role in the risk of AD. This aligns with a previous study that concluded that left-handed patients with AD do not differ from right-handed patients in the severity or pattern of neuropsychological deficits [35]. However, left-handedness may contribute to the early appearance of cognitive deficits during the development of Alzheimer disease, but such effects taper off with progression of the disease.

It was found that it takes much longer to progress from CN to MCI or from CN to AD, but progression from MCI to AD is much shorter. This suggests that the capability of predicting the progression from CN to AD would provide a much better opportunity for doctors to develop treatment and therapy to prevent progression. Once a sign of MCI shows, progression could happen fast, leaving little to no time for early treatment and intervention. Hence, studies that aim to identify the risk of progression to MCI and AD or determine the features that could be used to predict the onset of AD in clinically cognitively normal individuals are essential for early prevention.

This study demonstrated that the machine learning method employed significantly enhanced the accuracy of predicting the progression from CN to AD, achieving an accuracy of approximately 86% when using all visit data and 80% when using only Day 0 data (Figure 7) [8]. It was expected that using Day 0 data only would yield a lower accuracy than using the data from all visits because (1) the sample size from Day 0 was much smaller than that from all visits; (2) the symptom at Day 0 was not as pronounced as that at the later visits. These two reasons also caused the result that 80 features were needed for best accuracy if only Day 0 data were sued, while 40 features were needed if data of all visits were used. Despite this, the accuracy only improves from 82% using the first 10 features from 83% by using 80 features, suggesting that focusing only the top 10 features would still yield a relatively accurate prediction for progression (Figure 7).

The machine learning method implemented in this study was able to avoid overfitting successfully by using an ensemble algorithm. There were originally 203 features derived from post-processed MRI and global amyloid centiloid. With dimension reduction, features with strong correlations were removed, resulting in 85 features with weak correlations in the final dataset, which is still a dataset with high dimensions that could result in overfitting. To mitigate that, classification methods with ensemble algorithms, Random Forest or Bagging Decision Tree, were used. Both methods are more accurate than a single decision tree because it minimizes overfitting. Bagging is an ensemble algorithm that fits multiple models on different subsets of a training dataset, then combines the predictions from all models to enhance the accuracy of the predictions. Random Forests is an extension of bagging that uses an enhanced version of the bootstrap sampling model to build multiple decision trees and aggregate them to get an accurate result [36].

To prevent the risk of missing important features while avoiding overfitting, different feature selection methods were employed, ranging from 1 to all 85 features, using (1:20, 30, 40, 50, 60, 80) in RFE. As shown in Figure 4 and Figure 7, using all 85 features did not give the best accuracy, suggesting overfitting. Another important observation made from both figures was that Mean Cortical Amyloid Centiloid was the most important feature. However, the addition of the volume of Hippocampus significantly improved the accuracy. This indicates that increased amyloid centiloid and atrophy of the hippocampus are two of the most important characteristics of AD or high risk of progression from CN to MCI or AD. It also suggests the importance of having both MRI and PET scans for accurate diagnosis as well as predicting the progression of the disease [25].

Conclusion

In conclusion, RFE machine learning with Random Forest and Bagged Decision Tree was effective in determining the cognitive state and forecasting the progression from CN to MCI or AD. The combination of MRI and PET scans provided an effective, non-invasive way for diagnosis. Important features from the scans were identified, providing important guidance on reading the MRI and PET scan images for doctors for diagnosis as well as evaluating the risk of progression. Future work will be focused on understanding the details of the amyloid distribution and identifying the regions with amyloid plaque deposition that have the most impact on cognitive health as well as progression.

Limitations

Two limitations in the study were identified. The first limitation was that all the dementia in the database were assumed to be AD. Although most of the dementia in the OASIS-3 database was AD, other types of dementia did exist, for example, Lewy Body Dementia. The data was not removed to ensure the sample size of dementia positive was not too small. The second limitation came from the data used to characterize the amyloid plaque. Only the global amyloid amount, i.e., mean cortical amyloid centiloid, was involved in the machine learning; therefore, regions that have the most impact on cognitive state were not identified, which will be the focus of future work.

Acknowledgments

I would like to thank the Open Access Series of Imaging Studies (OASIS) for providing free access to all the data.

References

[1] M. Prince, R. Bryce, E. Albanese, W. A. E and W. Ribeiro, "The global prevalence of dementia: a systematic review and metaanalysis," *Alzheimers Dementia,* vol. 9, pp. 63-75, 2013.

[2] K. Yiannopoulou and S. Papageorgiou, "Current and Future Treatments in Alzheimer Disease: An Update," *J. Cent. Nerv. Syst. Dis.,* vol. 12, 2020.

[3] J. Neugroschl and S. Wang, "Alzheimer's disease: diagnosis and treatment across the spectrum of disease severity," *Mt Sinai J Med,* vol. 78, p. 596–612, 2011.

[4] Editorial, "The need for early detection and treatment in Alzheimer's disease," *EBioMedicine,* vol. 9, pp. 1-2, 2016.

[5] U. F. a. D. Administration, "Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industy," February 2018. [Online]. [Accessed 2023].

[6] Z. Breijyeh and R. Karaman, "Comprehensive Review on Alzheimer's Disease: Causes and Treatment," *Molecules,* vol. 25, no. 24, p. 5789, 2020.

[7] C. Marcus, E. Mena and R. M. Subramaniam, "Brain PET in the Diagnosis of Alzheimer's Disease," *Clin Nucl Med.,* vol. 39, no. 10, p. e413–e426, October 2014.

[8] A. Ebrahimighahnavieh, S. Luo and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Computer Methods and Programs in Biomedicine,* vol. 187, p. 105242, 2020.

[9] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang and X. Zhao, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Neurocomputing,* vol. 333, p. 145–156, 2019.

[10] S. Esmaeilzadeh, D. I. Belivanis, K. M. Pohl and E. Adeli, "End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification," *Mach Learn Med Imaging,* vol. 11046, pp. 337-345, 2018.

[11] X. Long, L. Chen, C. Jiang and L. Zhang, "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLoS One,* vol. 12, pp. 1-19, 2017.

[12] J. Albright, "Forecasting the progression of Alzheimer's disease using neural network and a novel preprocessing algorithm" *Alzheimer's & Dementia: Translational Research & Clinical Interventions,* vol. 5, pp. 483-491, 2019.

[13] P. M. Granitto, C. Furlanello, F. Biasioli and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometric and Intelligent Laboratory Systmes,* vol. 83, no. 2, pp. 83-90, 2006.

[14] P. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant , J. Hassentab, K. Moulder, A. Vlassenko, M. E. Raichle, C. Cruchaga and D. Marcus, "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer's Dissease," 2019.

[15] M. Reuter, N. Schmansky, H. Rosas and B. Fischl, "Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis," *Neuroimage,* pp. 1402-1418, 2012.

[16] A. F. B. S. M. Dale, "Cortical surface-based analysis. I. Segmentation and surface reconstruction," *Neuroimage,* vol. 9, pp. 179-194, 1999.

[17] B. S. M. D. A. Fischl, "Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system," *Neuroimage,* vol. 9, pp. 195-207, 1999.

[18] Y. Su, T. Blazey, A. Snyder, M. Raichle, D. Marcus, B. Ances, R. Bateman, N. Cairns, P. Aldea, L. Cash, J. Christensen, K. Friedrichsen, R. Hornbeck, A. Farrar, C. Owen, R. Mayeux, A. Brickman, W. Klunk, J. Price, P. Thompson, B. Ghetti, A. Saykin and R. Sperling, "Su, Y. et al. Partial volume correction in quantitative amyloid imaging. NeuroImage," *Neuroimage,* vol. 107, pp. 55-64, 2015.

[19] Y. Su, G. M. D'Angelo, A. G. Vlassenko, G. Zhou, A. Z. Snyder,, D. S. Marcus, T. M. Blazey, J. . J. Christensen, S. Vora, J. C. Morris, M. A. Mintun and T. L. S. Benzinger, "Quantitative Analysis of PiB-PET with FreeSurfer ROIs," *Plos One,* vol. 8, p. e73377, 2013.

[20] W. . E. Klunk, R. A. Koeppe, J. C. Price, T. L. Benzinger, M. . D. Devous Sr , W. J. Jagust, K. A. Johnson, C. A. Mathis, D. Minhas, M. J. Pontecorvo, C. C. Rowe, D. M. Skovronsky and M. . A. Mintun, "The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET," *Alzheimers Dement,* vol. 11, no. 1, pp. 1-15, 2015.

[21] Y. Su, S. Flores, R. Hornbeck, B. Speidel, A. Vlassenko, B. Gordon, R. Koeppe, W. Klunk, C. Xiong, J. Morris and T. Benzinger, "Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies," *NeuroImage: Clinical,* vol. 19, pp. 406-412, 2018.

[22] J. D. Evans , Straightforward Statistics for the Behavioral Sciences, Pacific Grove, Calif: Brooks/Cole Publishing, 1996.

[23] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc Natl Acad Sci U S A,* vol. 97, pp. 11050-11055, 2000.

[24] *OASIS-3: IMAGING METHODS AND DATA DICTIONARY Version 2,* 2022.

[25] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review," *Alzheimer's Research & Therapy,* vol. 13, no. 162, 2021.

[26] X. Nie, Y. Sun, S. Wan, H. Zhao, R. Liu, X. Li, S. Wu, Z. Nedelska, J. Hort, Z. Qing, Y. Xu and B. Zhang, "Subregional Structural Alterations in Hippocampus and Nucleus Accumbens Correlate with the Clinical Impairment in Patients with Alzheimer's Disease Clinical Spectrum: Parallel Combining Volume and Vertex-Based Approach," *Front Neurol.,* vol. 8, no. 399, 2017.

[27] B. Patenaude, S. M. Smith, D. N. Kennedy and M. Jenkinsona, "A Bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage.,* vol. 56, no. 3, pp. 907-922, 2012.

[28] A. Qiu, C. Fennema-Notestine, A. M. Dale and M. I. Miller, "Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease," *Neuroimage,* vol. 45, no. 3, pp. 656-661, 2009.

[29] A. M. Butts, M. M. Machulda, P. Martin, S. A. Przybelski, J. R. Duffy, J. Graff-Radford, D. S. Knopman, R. C. Petersen, C. R. Jack, V. J. Lowe, K. A. Josephs and J. L. Whitwellf, "Temporal Cortical Thickness and Cognitive Associations among Typical and Atypical Phenotypes of Alzheimer's Disease," *J Alzheimers Dis Rep.,* vol. 6, no. 1, p. 479–491, 2022.

[30] S. Knafo, "Amygdala in Alzheimer's Disease," in *The Amygdala - A Discrete Multitasking Manager*, B. Ferry, Ed., 2012.

[31] M. Zarei, B. Patenaude, J. Damoiseaux, C. Morgese, S. Smith, P. M. Matthews, F. Barkhof, S. A. R. B. Rombouts, E. Sanz-Arigita and M. Jenkinson, "Combining shape and connectivity analysis: an MRI study of thalamic degeneration in Alzheimer's disease," *Neuroimage,* vol. 49, no. 1, pp. 1-8, 2010.

[32] J. Min, W.-J. Moon, J. Y. Jeon, J. W. Choi, Y.-S. Moon and S.-H. Han, "Diagnostic Efficacy of Structural MRI in Patients With Mild-to-Moderate Alzheimer Disease: Automated Volumetric Assessment Versus Visual Assessment," *AJR Am J Roentgenol,* vol. 208, no. 3, pp. 617-623, 2017.

[33] E. S. Sharp and M. Gatz, "The Relationship between Education and Dementia An Updated Systematic Review," *Alzheimer Dis Assoc Disord.,* vol. 25, no. 4, pp. 289-304, 2011.

[34] M. M. Mielke, "Sex and Gender Differences in Alzheimer's Disease Dementia," *Psychiatr Times.,* vol. 35, no. 11, pp. 14-17, 2018.

[35] R. S. Doody, J. L. M. Vacca, P. Paul J. Massman and e. al, "The Influence of Handedness on the Clinical Presentation and Neuropsychology of Alzheimer Disease," *Arch Neurol. ;56(9):. doi:10.1001/archneur.56.9.1133,* vol. 56, no. 9, pp. 1133-1137, 1999.

[36] P. Sarkar, "Bagging and Random Forest in Machine Learning," KnowledgeHut, 17 April 2022. [Online]. Available: https://www.knowledgehut.com/blog/data-science/bagging-and-random-forest-in-machine-learning. [Accessed 2023].

[37] W. J. Henneman, J. B. J. v. d. F. W. M. Sluimer, I. C. Sluimer, N. C. Fox, P. V. H. Scheltens and F. Barkhof, "Hippocampal atrophy rates in Alzheimer disease," *Neurology.,* vol. 72, no. 11, p. 999–1007, 2009.

[38] B. Hussain, K. Huh, H. Wing, E. Chan and S. Patanwala, "Surviving in a Random Forest with Imbalanced Datasets," SFU Professional Computer Science, [Online]. Available: https://medium.com/sfu-cspmp/surviving-in-a-random-forest-with-imbalanced-datasets-b98b963d52eb. [Accessed November 2022].