Article

Classification of Alzheimer's Disease with and without Imagery Using Gradient Boosted Machines and ResNet-50

Lawrence V. Fulton[1], Diane Dolezel[1], Jordan Harrop[2], Yan Yan[1], Christopher P. Fulton[3]

[1] Texas State University, 601 University Drive, San Marcos, TX 78666
[2] Acushnet Holdings Corporation, Acushnet, MA 02743
[3] United States Air Force Experimental Test Pilot School, Edwards Air Force Base, CA 93524
* Correspondence: lf25@txstate.edu; Tel.: +01-512-245-3453

Abstract: Alzheimer's is a disease for which there is no cure. Diagnosing Alzheimer's Disease (AD) early facilitates family planning and cost control. The purpose of this study is to predict the presence of AD using socio-demographic, clinical, and Magnetic Resonance Imaging (MRI) data. Early detection of AD enables family planning and may reduce costs by delaying long-term care. Accurate, non-imagery methods also reduce patient costs. The Open Access Series of Imaging Studies (OASIS-1) cross-sectional MRI data were analyzed. A gradient boosted machine (GBM) predicted the presence of AD as a function of gender, age, education, socioeconomic status (SES), and Mini-Mental State Exam (MMSE). A Residual Network with 50 layers (ResNet-50) predicted CDR presence and severity from MRI's (multi-class classification). The GBM achieved a mean 91.3% prediction accuracy (10-fold stratified cross validation) for dichotomous CDR using socio-demographic and MMSE variables. MMSE was the most important feature. ResNet-50 using image generation techniques based on an 80% training set resulted in 98.99% three class prediction accuracy on 4,139 images (20% validation set) at Epoch 133 and nearly perfect multi-class predication accuracy on the training set (99.34%). Machine Learning methods classify AD with high accuracy. GBM models may help provide initial detection based on non-imagery analysis, while ResNet-50 network models might help identify AD patients automatically prior to provider review.

Keywords: Alzheimer's Disease; Extreme Gradient Boosting; Deep Residual Learning; conolutional neural networks; machine learning; dementia

1. Introduction

Alzheimer's Disease (AD) is a degenerative brain disease with no cure [1]. AD is the most common type of dementia; it is the sixth leading cause of death in the United States [2]. AD is characterized by progressive cerebral cortex atrophy leading to memory loss, increasing cognitive deficits, and potential loss of motor functions [3]. In 2019, there were 5.8 million people of all ages diagnosed with AD, the prediction is 14 million with AD by 2050 [2]. Specifically, one in ten people over 65 years of age have AD, with females more likely to have AD than males [3].

Due to the lengthy prodromal period for AD, it is commonly diagnosed in seniors [4]. Earlier diagnosis of AD would facilitate treatment interventions and family planning, and it could save up to $7.9 trillion dollars annually [2]. Unfortunately, differential diagnosis of AD is an intense, time consuming, and costly process involving physical and mental exams, laboratory and neurology tests, as well as neurological imaging (MRI, CT and PET) [4]. As an example, the cost for a brain MRI may range from $600 to $1,300 [5], and for the uninsured or underinsured this might be infeasible.

1.1. Background

Dementia is not identical to the normal mental deterioration associated with aging [2]. It is a progressive, chronic disease with symptoms that may include the loss of memory, communication problems, behavioral changes, and personality changes. Individuals with dementia are often not oriented to time or place and may not remember to eat regularly or practice good hygiene. AD is the most common (60-80%) form of dementia [2]. Normal "forgetful" elderly individuals can be assisted to remember; however, those with AD may not even recall reminders. This fact is unsurprising since the etiology of AD (and other dementias) is damage to brain cells visible by imagery. AD is associated with atrophy of the cerebral cortex, which is the outer (grey wrinkled) covering of the cerebrum in the brain. The irreversible atrophy derives from amyloid plaque formation and neurofibrillary tangles. Amyloid (protein) plaques bind to and destroy nerve synapses (where nerves touch) resulting in memory loss, an early symptom of AD. Neurofibrillary tangles, from a twisted protein strand, damage neurons and nerve synapses.

AD has an extremely slow onset.   The prodromal period may be 10-20 years while plaques and tangles continue spreading through the cortex [3]. Definitive diagnosis is complicated because amyloid plaque formation and neurofibrillary tangles are present even in seniors with good cognitive functioning [6,7], although in lesser numbers and without the predictable pattern present in AD brain scans. To that end, diagnostics from brain Magnetic Resonance Images (MRIs), Computed Tomography, and Positron Emission Tomography are useful for monitoring the progression of AD. Brain volume atrophy, especially in the hippocampal areas, also reflects the decline in cognitive functions characteristic of AD pathology [8].

Scores from longitudinal clinical tests, like Mini-Mental State Exam (MMSE), Clinical Dementia Rating (CDR), and Alzheimer's Disease Assessment Scale (ADAS), have shown high correlation with AD disease progression [4,9]. Mapping this association is helpful because it allows clinicians to provide interventions in the asymptomatic preclinical period. To clarify, while there is no cure for AD, the progression may be slowed, and the symptoms treated. Common treatments include memory drugs, behavior mediation, sleep treatments, occupational therapy, and assisted living [2]. Thus, the early identification of AD is vital to slowing the progression of the disease.

1.2. Related Research

Mapping the progression of AD through all six stages requires complex non-linear multifactorial multi-modal modeling. Researchers have used Growth Mixture Modeling (GMM) to model the decay of preclinical MMSE scores of individuals (*n*=528) over 75 years of age [10].   GMM was used to detect both group and individual changes. Bhagwat et al. [4] modeled AD symptom progression classes for one thousand participants over six years with the MMSE scores and the AD AS scale. The authors predicted AD progression using MRI imaging, genetic and clinical characteristics, and by applying a Siamese neural network which is a network with two or more identical equally weighted subnetworks. Zhang et al. [11,12] used a subset of the Open Access Series of Imaging Studies (OASIS) data (*n*=126) and proposed methods to identify binary-coded AD using Eigenbrain imagery. They eliminated individuals under 60 and incomplete observations, modeled AD as binary, limited their analyses to a very small subset of coronal slice images, and used 10-fold cross-validation run 50 times. In this paper, we use 51 slices, all data, random image generation, and *multi-class* classification using ResNet-50 on a 20% blinded validation set. Bansal et al. [13] applied several machine learning (ML) models to the OASIS brains dataset to classify binary-coded AD; however, they used the entirety of the data set which may have resulted in overfitting and prevented the generation of true performance metrics.

A related study focused on predicting individual Alzheimer's Disease Assessment Scale (ADAS) and MMSE scores [14]. The researchers used an anatomically partitioned artificial neural network, cross-sectional baseline data sets, three sets of Alzheimer's Disease Neuroimaging

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2019   doi:10.20944/preprints201907.0345.v1

Peer-reviewed version available at *Brain Sciences* **2019**, *9*, 212; doi:10.3390/brainsci9090212

Initiative cohort data, and two MRI measures. Another study used MRI brain images of tissue perfusion and perfusion scores to develop discrimination maps which successfully distinguished AD, mild cognitive decline, and subjective cognitive decline [15]. Other researchers applied ML to Alzheimer's Disease Neuroimaging Initiative data (*n*=805) to predict AD clinical scores [9]. By using soft-split sparse regression-based Random Forests to estimate missing longitudinal scores, they produced more accurate predictions than traditional Random Forest methods.

This work extends and expands these earlier researchers' efforts. This study has two objectives. The first objective is to build machine learning models, specifically, Gradient Boosted Machines (GBM), that effectively predict both the presence of AD using ML methods *without* imagery. The second objective is to build imagery models using Deep Learning techniques, specifically a Residual Network with 50 layers (ResNet-50), that improves previous classification efforts. The significance of this study is that it is the first of its type to evaluate imagery versus non-imagery methods for classification of Alzheimer's Disease and that it expands the previous work for identification and classification of Alzheimer's Disease by including multi-class classification as well as different architecture. Non-imagery methods may be particularly relevant to the poor and underserved, as methods which adequately forecast AD without imagery reduce the costs of imaging. Imagery methods that are increasingly accurate may help providers with diagnostics.

2. Materials and Methods

2.1. Data and Software

The Open Access Series of Imaging Studies (OASIS-1) is publicly available data that provides researchers access to cross-sectional and longitudinal MRI data [16]. The data derive from the work of Marcus, et al. [17]. The full data set contains cross-sectional data (*N*= 416) derived from MRI studies at Washington University. The OASIS-1 data include MRI imagery in image format as well as patient socio-demographics and clinical variables in data frame format. In this study, we used the atlas-registered, gain-field corrected images provided as well as the associated data in the data frame.

The MRI imagery data includes {x=176, y=208, z=176} images of the {sagittal (x,y), axial (x,z), coronal (y,z)} planes for each of the *N*=416 observations, respectively. The second dimension, y=208, is defined as the number of slices. All images are from right-handed men and women (a control) aged 18 to 96. Both the imagery and non-imagery data sources are used in this study. In this study, the imagery data are restricted to the center 51 slices (78 through 128), as the outer slices contain less information about AD presence and severity. The total number of valid training images used for classification was 16,661 using and 80% training set; however, images were processed through an image generation algorithm to increase the learning of the ResNet model. The remaining 4,139 images were used as a validation set. Images were padded with 3 voxels, so the total image voxels used was over 644 million.

Both R Statistical Software version 3.5 ("Feather Spray") and Anaconda Python version 3.6 were used for this study [18,19]. R was used for processing the labels and non-imagery analysis, while Python was used for constructing and running the ResNet-50 model.

2.2. Dependent Variable

For this study, the single response variable (label) is the Washington University Clinical Dementia Rating (CDR) which was developed to measure dementia severity [17]. The CDR is measured as follows: {0= nondemented; 0.5 = very mild dementia; 1 = mild dementia; 2 = moderate dementia}. This instrument is reliable in estimating clinical dementia, which is largely related to AD [2]. Because only two cases in the N=416 data set involved moderate dementia, the values of

CDR were recoded as follows: {0 = nondemented, 1 = very mild dementia, 2 = mild or moderate dementia}. This recoding still resulted in imbalanced classes. The distribution of the dependent variable values was {316, 70, 30} for {no dementia, very mild dementia, mild or moderate dementia}, respectively.

There are also three sets of predictor variables available in the OASIS-1 dataset: demographic variables, clinical predictor variables, and imagery predictor variables. Each of these sets of variables is defined in the following sections.

2.2.1. Socio-demographic Predictor Variables, Non-Imagery Data

Several independent variables from the OASIS-1 dataset provided patient sociodemographic characteristics. A listing of these variables with their associated operational definitions follows.

- Gender: {0=Female, 1=Male}, 100% complete (416 of 416)

- Age: [18, 96] years of age, 100% complete (416 of 416)

- Education: {1 < high school (HS), 2=HS Graduate, 3=Some College, 4=College Graduate, 5=Beyond College Graduate}, 56% complete (235 of 416)

- Socioeconomic Status (SES): {1=lower, 2=lower middle, 3=middle, 4=upper middle, 5=upper}, 52% complete (216 of 416)

While much of the data were complete, education and SES were not. Imputation using the median was used to facilitate non-imagery analysis. While a large percentage of observations were missing, analyses were run omitting the variables, including non-response as a factor level, and including imputation using the median. The median was selected based on performance.

2.2.2. Clinical predictor variables, Non-Imagery Data

Clinical predictor variables were also available. These variables include the Mini-Mental State Exam (MMSE) score, the Atlas Scaling Factor, the Estimated Total Intracranial Volume, and the Normalized Whole Brain Volume. Three of the four variables require imagery (all except the MMSE), so they were not used in the non-imagery model. The definitions of these variables follow.

- Mini-Mental State Exam (MMSE): [0,30]. The MMSE is a 30-point questionnaire that has been shown to be valid and reliable in identifying dementia [17,20]. The variable was 56% complete (235 of 416). Missing values were imputed with the median. The MMSE was included in the non-imagery model. [20,21].

- Atlas Scaling Factor (ASF): [.88, 1.56] (observed). The ASF is a one-parameter scaling factor that allows for comparison of eTIV based on differences in human anatomy [17]. This variable was 100% complete (416 of 416). [10].

- Estimated Total Intracranial Volume (eTIV): [1132, 1992] mm^3 [22]. The eTIV variable estimates intracranial brain volume. This variable was 100% complete (416 of 416).

- Normalized Whole Brain Volume (nWBV): [.64, .90] mg (observed). These variable measures the volume of the whole brain. This variable was 100% complete (416 of 416).

2.2.3. Imagery Variables

The imagery used in this study are the masked version of the gain-field corrected, atlas-registered image to the Talairach atlas space. These images are preprocessed with spatial normalization of the 3-D brain images. The intensity inhomogeneity (IIH) or gain field are intensity variations (i.e. noise) not related to the patient's anatomical image which include patient movement, nearby static fields, radio-frequency variation, or other non-patient source [23]. These gain field variations are corrected to form an averaged 3D image. Next, the identifying 3D facial and non-brain values are masked. The Talairach space is, then, a 3-D coordinate system (called an atlas) used for mapping locations of brain features [24,25]. The averaged 3D images are spatially normalized into the Talairach space with a transform that maps the subject's brain image to a reference brain and smooths out individual brain anatomy size and shape differences [23, 24]. Figure 1 illustrates the sagittal, axial, and coronal viewpoints as well as the resolution used in this study.
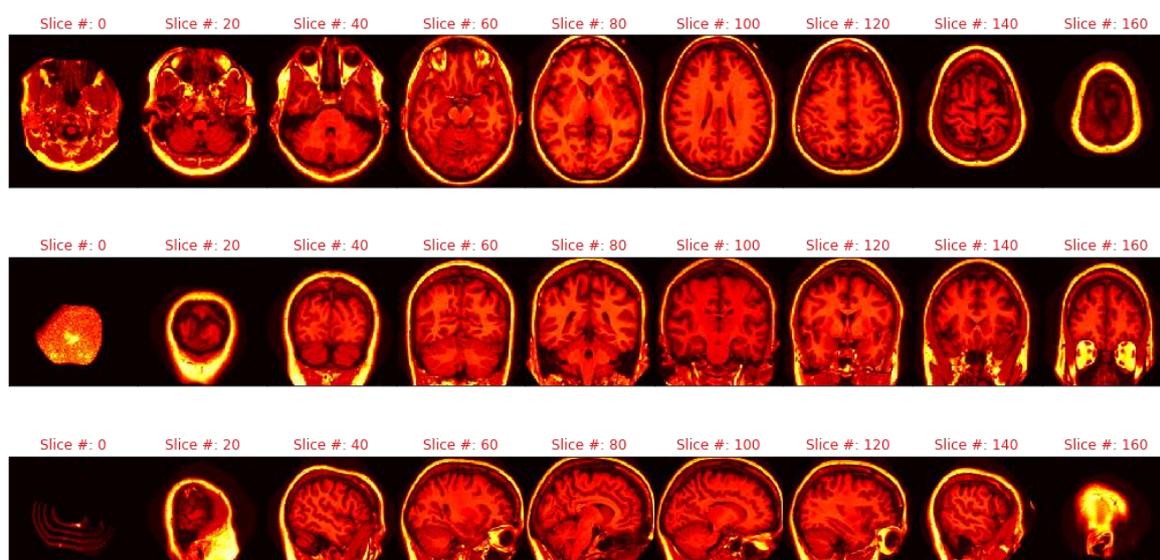
Figure 1.   Axial, sagittal, and coronal slices (respectively)

2.3.   Non-Imagery Models

2.3.1. Imputation, Cross Validation, and Pseudo-Random Seeds

For the non-imagery data file, imputation was used to complete missing data fields. Education level and socio-economic status (SES) had high percentages of missing data (44% and 48%, respectively).   Medians were imputed to investigate whether the variables might still be useful in forecasting.   All other variables were 100% complete, including all imagery.

K-fold cross validation was used to assess the non-imagery model, where k=10.   This technique subdivides the data into five separate folds or sections.   Nine folds are used for training the models, and one-fold is used for evaluating the performance after training.   Each fold serves as a validation set one time, and performance metrics are calculated based on the results from all 10 training runs.

Pseudo-random seeds were used for replication of the results. The use of random seeds ensures that the results are not due simply to the selection of random numbers via pseudo-random number generation. Further, random seeds allow for replication of exact results.
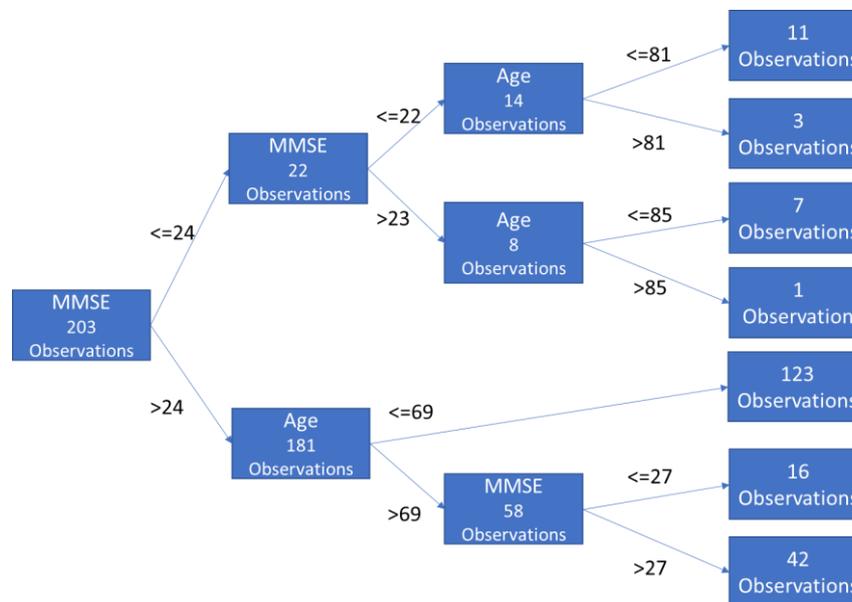
2.3.2. Gradient Boosted Tree Ensembles (Gradient Boosting)

Gradient boosted tree ensembles were used to classify AD using the socio-demographic variables and MMSE. Gradient boosted tree ensembles, known often as gradient boosted machines (GBM), is a model that optimizes prediction accuracy based on iterations of weaker decision/classification tree models.   A classification tree model seeks to split the independent variables at various points in an effort to create a decision tree that provides a classification vote. Often, the Gini impurity or cross-entropy formulae are used to determine those splits (Equation 1 and Equation 2) [26,27].   In these equations, k is the number of classes and p_i is the proportion of cases belong to case i.

Gini impurity: $\sum_{i=1}^{k} p_i(1 - p_i)$ [1]

Cross-entropy: $-\sum_{i=1}^{k} p_i \log(p_i)$ [2]

With these equations, in a greedy fashion the splitting algorithm is trying to achieve homogeneity within each subset of the tree. To avoid overfitting, the trees are pruned by limiting the number of internal splits/branches. In this implementation, individual trees grew no more than three branches to avoid overfitting. Each tree recommends a classification status (e.g., no AD or AD).   Figure 2 illustrates a sample classification tree.

Figure 2.   A sample tree classification with a maximum of three branches

At a high level, a gradient boosted tree ensemble is a collection of tree models that is derived by sequentially incorporating, in an additive manner, direct residual errors (i.e. boosting). At each iteration, a differentiable loss function is optimized (e.g., area under the curve) [28].   A good discussion of trees, forests, and boosting is from Chen [29].

The R Statistical Software Package, XGBoost [30], was used to implement gradient boosting. The XGBoost implementation of gradient boosting is ideal for this use case. First, this approach has seen considerable success in prediction accuracy across a variety of datasets. Second, XGBoost takes advantage of unique algorithms that minimizes computation complexity which allows gradient boosting to scale fairly easily in parallel fashion. A hyperparameter grid search provided parameters for use in building the trees and a ten-fold cross-validation provided accuracy metrics.

2.4. Imagery Data

2.4.1. Pre-processing: Min-Max Scaling, PCA Investigation, Pseudo-Random Seeds

While gradient-boosted forests are scale invariant, ResNet-50 (an adaptation of a CNN) is not. The data for the ResNet were min-max scaled based on the data set minimum and maximum values, {0, 5089}. The minimum value for any voxel is known to be zero; however, extracting the maximum value ensured that no values in either the training or validation set would be below zero. The leakage of information from the validation set was at most *one* value (assuming 5,089 was randomly assigned to the validation set).

After splitting the data into training and validation sets, Principal Component Analysis (PCA) was used to extract significant Eigenbrain imagery *separately* to avoid bias [31,32]. Eigenbrains are reconstructed images based on Eigenvectors (orthogonal). Analysis of these structures, rather than original imagery, failed to improve the performance of the ResNet. The *nibabel* library in Anaconda Python 3.6 provided the MRI import utility, while *scikit learn* provided the PCA algorithm. Once again, pseudo-random seeds were used for replication in all analysis.

2.4.2. Extraction and Manipulation of Individual Images

The 416 observations contained 208 slices of 176 x 176 voxels. Fifty-one of these slices (range 78-128) were used in the analysis. These slices were extracted and saved as *.png* single channel (grayscale) files from the original format in a separate directory. Matching classification labels were replicated for each individual picture (i.e., CDR=0 for patient 1 was replicated 51 times for each associated slice.) As part of the modeling process, the images were padded with 3 voxels on all sides and subjected to random image transformation to improve the ML learning. Not all images were valid. Instead of 51 x 416 = 21,216 images, there were only 20,800 contained within all MRIs. Without considering the unique (and infinite) transformations, the total images include more than 640 million voxels.

*2.4.3. Flow from Dataframe, Training and Validation Set, Training Image Generation*

Due to the large size of the dataset, file names and associated labels for each slice were saved in a data frame. This data frame was randomly divided into two separate data frames, an 80% training set and a 20% validation set. A *flow_from_dataframe* function was used to load image in batches of 128 (investigated through tuning) by randomly selecting file names in the training data frame. A training image generator then rescaled the data to be between 0 and 1 (dividing by 255 for grayscale). That same generator randomly rotated the images up to 20 degrees, zoomed up to 10%, shifted the height and width up to 10%, sheared up to 10%, and horizontally flipped some images. By editing the training images in this fashion, the ML model learns to adapt to deviations in expected format. The validation data were left intact. No processing of those images was done to avoid the infusion of bias from manipulation. Those images were unknown.

2.4.4. Deep Learning with Residual Networks (ResNet-50), Imagery Data

ResNet-50 was used to classify CDR based solely on MRI imagery data. The model relied on *Keras* (*TensorFlow* backend) [30,33]. ResNet-50 is a residual deep learning network (with 50 layers) which attempts to address the problem of vanishing gradients that occur during back-propagation of convolutional neural networks (CNN). The ResNet-50 model was developed by He, Zhang, and Sun, and an ensemble of these ResNet models of various depths won imagery classification awards [34]. ResNet-50 is a residual deep learning network (with 50 layers) which attempts to address the problem of vanishing gradients and degradation problem. Increasing the depth of the network should increase the accuracy of the network, as long as over-fitting is considered. However, the problem with increased depth is that the signal required to change the weights, which arises from the end of the network by comparing ground-truth and prediction (observed versus predicted),

becomes very small at the earlier layers because of the increased depth of the network. It essentially means that earlier layers remain almost unlearned.   This is called the "vanishing gradient" problem, as the matrix of second order derivatives (the gradient) in the nonlinear optimization that attempts to tune weights becomes near zero. The second problem with training the deeper networks is, performing the optimization on huge parameter space and therefore naively adding the layers leading to higher training error. Residual networks allow training of such deep networks by constructing the network through modules called residual models. This is called degradation problem. The ResNet-50 architecture is documented in [35]. The loss-metric optimized was categorical cross-entropy (formula 2).

Convolutional blocks are a major component of the ResNet-50 model. These networks apply multiple filters (e.g., 3 x 3-pixel size filter) to images to classify.   These filters are moved by strides across the original image.   The values in the filters (which are learned) are multiplied against the values in the images.   The results of those filters are pooled (e.g., maximum values extracted after the application of filters), which is effectively down sampling while retaining the most important features.   Figure 3 is a symbolic picture of CNN.
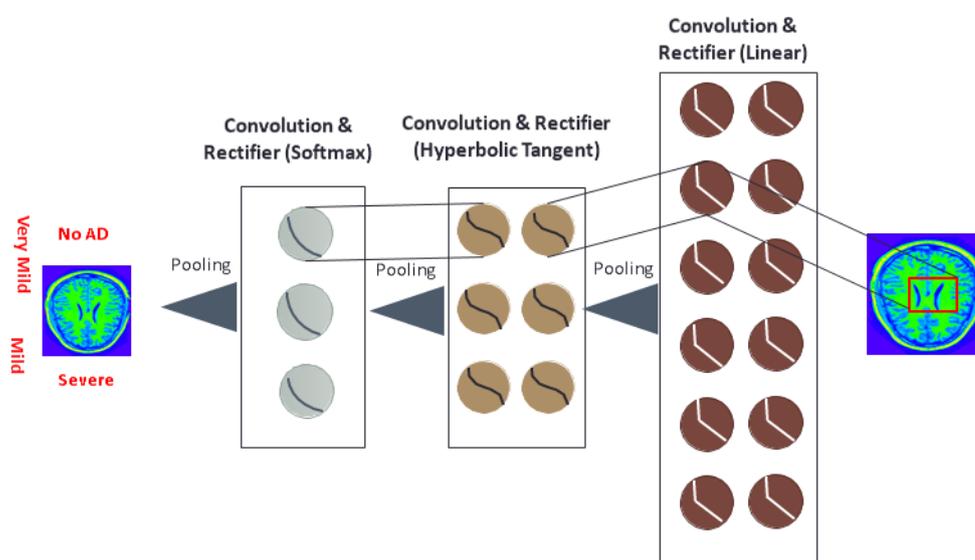
Figure 3.   Example convolutional neural network (read from right to left)

3. Results

3.1. Descriptive Statistics

The data from the OASIS-1 study include 416 patient diagnostic files with 100 of those files confirming dementia. No patients under the age of 60 were diagnosed with dementia, as this is a rare event.   One might assume that in GBM models, then, age would be the most important predictor. (It is not.) The table of dementia vs. non-dementia patients and associated statistics and frequencies adopted from the OASIS (2018) data is shown in Table 1. Because of the apparently dichotomous split at age 60, imagery and non-imagery models were run with the entirety of the dataset as well as with a subset of those 60 and above. ResNet-50 is capable of handling imbalanced data without techniques such as synthetic minority oversampling when trained sufficiently. Oversampling techniques often help GBM models.

Table 1.   Descriptive Statistics for Dementia by Age and Gender (Adopted from OASIS , 2018)

| | | | Non-Demented | | | | Demented | | | |
Age	N	n	Mean	male	female	n	mean	Male	female	CDR 0.5/1/2
<20	19	19	18.53	10	9	0	0	0	0	0/ 0/ 0
[20,30)	119	119	22.82	51	68	0	0	0	0	0/ 0/ 0
[30, 40)	16	16	33.38	11	5	0	0	0	0	0/ 0/ 0
[40, 50)	31	31	45.58	10	21	0	0	0	0	0/ 0/ 0
[50, 60)	33	33	54.36	11	22	0	0	0	0	0/ 0/ 0
[60, 70)	40	25	64.88	7	18	15	66.13	6	9	12/ 3/ 0
[70, 80)	83	35	73.37	10	25	48	74.42	20	28	32/15/ 1
[80, 90)	62	30	84.07	8	22	32	82.88	13	19	22/ 9/ 1
[90, 100)	13	8	91.00	1	7	5	92.00	2	3	4/ 1/ 0
Total	416	316	n/a	119	197	100	n/a	41	59	70/28/ 2

Most of the patients in the dataset exhibited no dementia (*n*=316).   Only two demonstrated moderate dementia.   Table 2 provides the frequency distribution for CDR by gender.   Because of only two individuals experienced moderate dementia, CDR 1 and CDR 2 were combined for the imagery analysis.

Table 2. *CDR Frequency Distribution by Gender*

	CDR=0, No Dementia	CDR=0.5, Very Mild Dementia	CDR=1, Mild Dementia	CDR=2, Moderate Dementia
Male	119	31	9	1
Female	197	39	19	1
Total	316	70	28	2

The average patient in the dataset was 52.7 years old with a slightly less than perfect mental state evaluation (27.5 out of 30), an estimated brain volume of 1480.53 mm^3, and 79% of the intracranial cavity occupied by the brain (nWBV).   Table 3 displays statistics for the quantitaive variable statistics.

Table 3. Descriptive Statistics for the Quantitative Variables (n = 416)

| Variable	Mean	SD	Median	Min	Max
Age	52.70	25.08	56	18	96
Mini-Mental State Exam	27.50	3.13	29	14	30
eTIV (Intracranial Volume)	1480.53	158.34	1475	1123	1992
nWBV (Brain Volume)	0.79	0.06	0.8	0.64	0.89
ASF (Atlas Scaling Factor)	1.2	0.13	1.19	0.88	1.56

Qualitative Variables Modal responses follow: 62% women, 28% college graduates, 30% middle class. Spearman's rank correlation between SES and education indicated a negative correlation ($r = -.715$, $p<.01$). For the 121 individuals without a high school education, 119 were in the lower middle and lower classes.   Figure 4 displays barplots for the variables of gender, SES, and education.
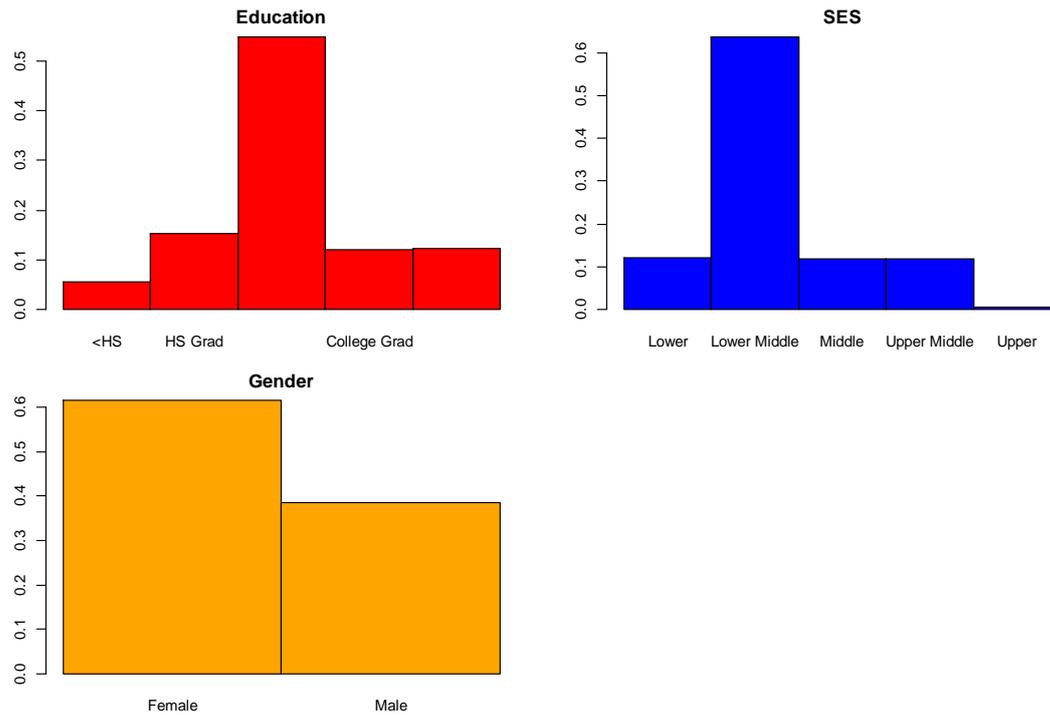
Figure 4. Bar charts of qualitative variables

3.2. Correlations and Variable Transformations

Correlation between ASF and eTIV is negative and significant (t_{414}=-88.7, *p*<.001) as is the correlation between age and nWBV (t_{414}=-36.4, p<.001), an expected finding. (Older individuals tend to lose brain volume.)   While some machine learning methods require transformations, tree methods are location/scale invariant, and ResNet (like all neural-net based models) benefits primarily from scaling. Thus, transformations (other than regularization through min-max scaling for imagery models) were not pursued.   The scatterplot matrix of all variables is shown in Figure 5.
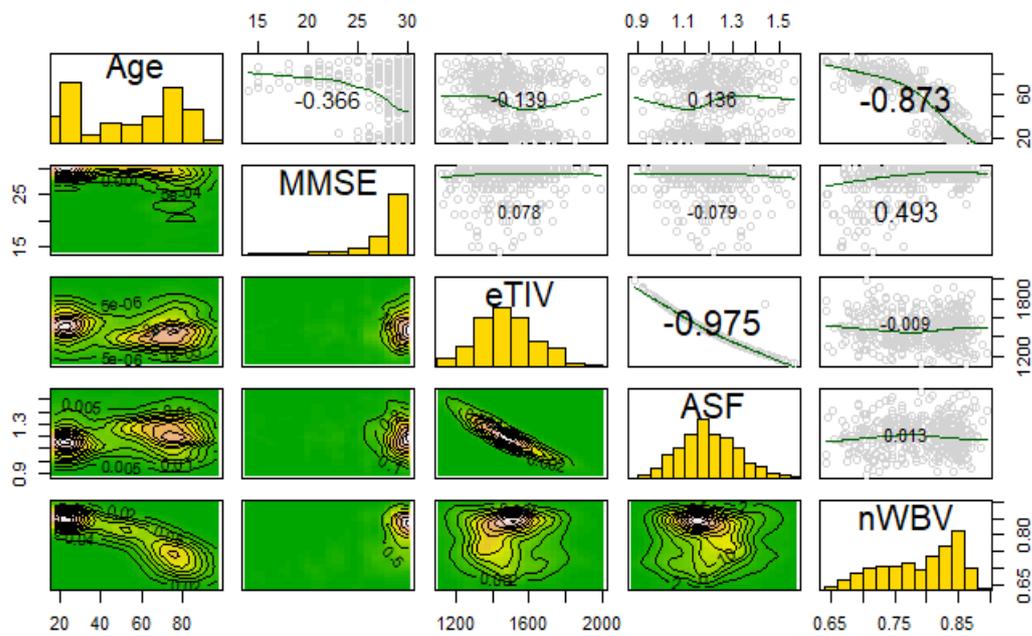
Figure 5. Correlation plots of all quantitative variables

3.3. Transformation of MRIs and Eigenbrain Development

Figure 6 shows the Eigenbrain associated with Figure 1. The Eigenbrain images are derived via Principal Components Analysis (PCA). In PCA, images are restructured by building linear combinations of the original imagery where the first reconstructed image captures the most variability, the second captures the second most, etc. PCA maximizes the likelihood function in Equation 3 where Vector a and scalar λ form the Eigenspace (eigenvector and eigenvalue), while matrix R is the correlation matrix [36]. For each eigenvector indexed i, variance capture of the remaining variance is maximized. There are k total eigenvectors, where k is the number of columns in the dataset.

$$Max_{a(i)}\ L_i = a^T_i\,R\,a_i - \lambda\,(a^T_i\,a_i),\ I = \{1,\ 2\ldots k\} \tag{3}$$
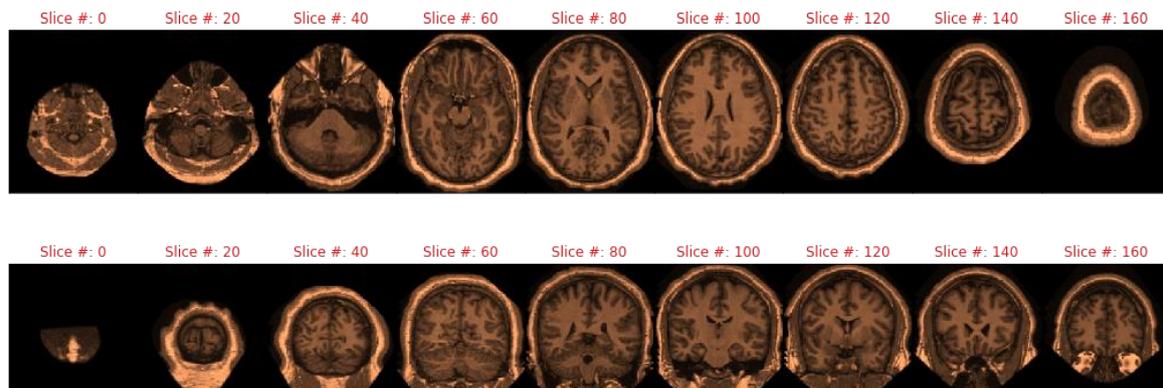
Figure 6. Eigenbrain imagery associated with Figure 1

3.4. Gradient Boosting with XGBoost [29], No Imagery

Gradient boosted random forests were used to model dichotomous CDR (presence or absence) as a function of socio-demographics and MMSE scores (no imagery). Hyperparameter grid search was used for tuning parameters. Ten-fold stratified cross-validation provided accuracy metrics. (Stratification addresses the class imbalance problem.) To avoid overfitting, the tree depth was constrained to three, a size investigated during grid search. The GBM achieved a mean of 91.3% classification accuracy with socio-demographic and MMSE variables. The gain, relative importance of the variable based on appearance in the trees, is depicted in Figure 7, while Figure 8 is Receiver Operating Characteristic (ROC) curve. From Figure 7, the MMSE test would appear to be the most valuable classifier with age following. From Figure 8, the area under the curve is shown to be solidly in the 90th percentile for all of the 10 folds.
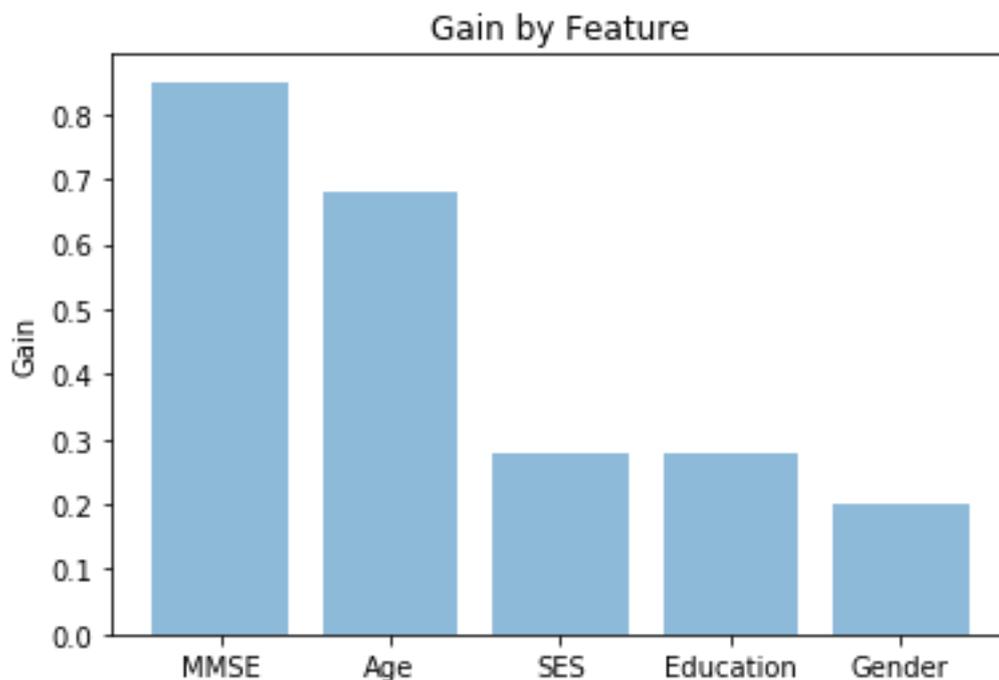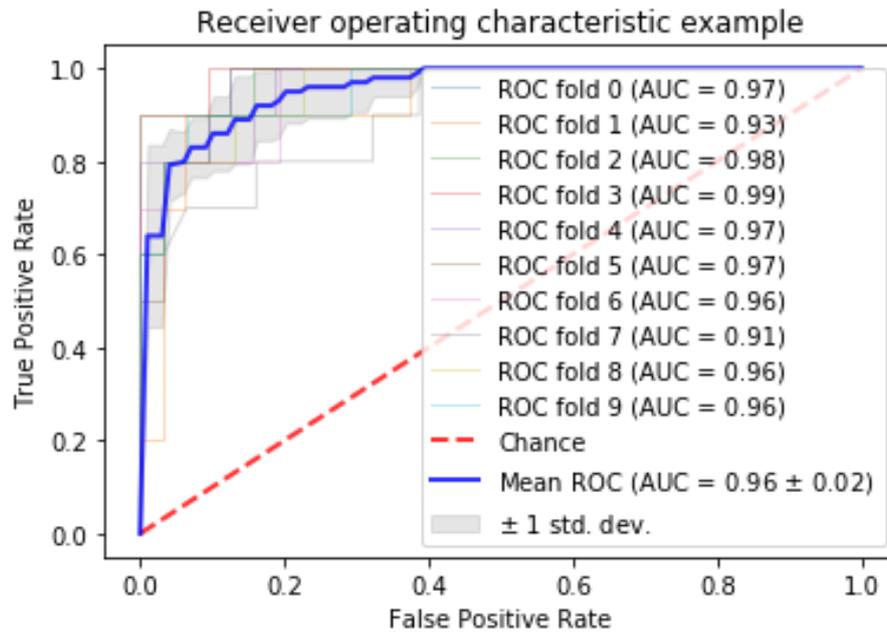

Figure 7. Variable importance

Figure 8. ROC Curve

3.4. *ResNet-50 (Deep Learning) with Keras [30]*

ResNet-50 was run to classify MRI imagery.   The batch size for analysis was 64. Each epoch took between 213 and 711 seconds to run on an Acer Predator G-9 with an Intel I7-6700, 64GB of fast random access memory, an NVidia GTX 980M graphical processing unit (GPU), running Windows 10. (The maximum run time was due to additional processing activities.)   A total of 152 epochs were run, and the total run time for this analysis alone was 10.2 hours.   The ResNet-50 was able to run on the computer's GPU, although other processing functions were off-loaded to the CPU.   Initial runs on the GPU were at least four times as long

As depicted in Figure 8, ResNet-50 performed on the training set correctly (and surprisingly) classified 98.99% of the validation set at 133 epochs.   The training set and test set nearly converged in accuracy.   The loss metrics dropped accordingly to near zero on both sets.   ResNet-50 proves to be a good automatic classifier.
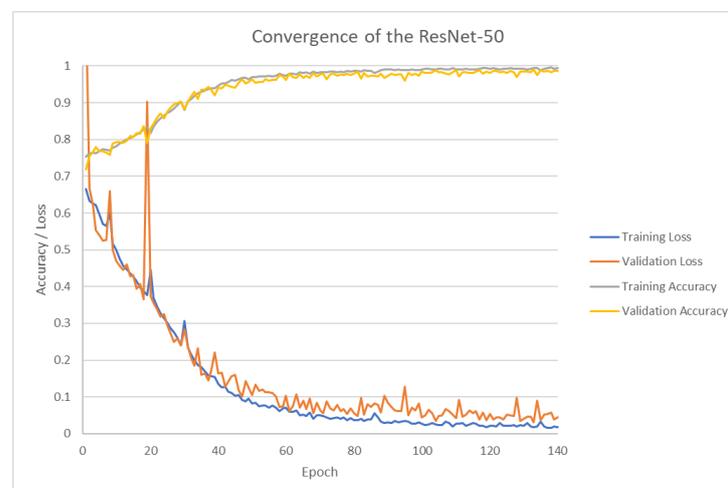

Figure 8.   Classification accuracy versus epoch

4. Discussion

Extreme gradient-boosted tree models provided 91.3% accuracy with only socio-demographics data.   The two most important variables in predicting AD were MMSE and (of course) age.   What is interesting is that MMSE had higher feature importance when no patients below 60 in the dataset had AD.

With imagery data, ResNet-50 correctly classified 98.99% of the validation set and 99.34% of the training set AD by severity after 152 epochs.   ResNet-50 is a powerful predictor, and such results support the notion that ML methods might be used as the first screening for diseases such as AD.

Another interesting finding is that while image methods may support radiologists' processes, non-imagery methods are reasonable at classifying the present of AD during screening.   The area under the curve in 8 of the 10 folds was 96% or better.   MMSE and age are powerful screening tools for AD.   Coupling these tools with ML imagery analysis to support radiologists and physicians may help diagnose earlier.   An MRI generated for any reason might be automatically screened as part of a best-practice policy.

The ResNet-50 multi-class classification itself is an improvement over previous studies such as [12, 13] that used cross-validation (which exposes the validation set to the same pre-processing as the training set thus inducing bias) and binary classification.   Further, it demonstrates that ResNet-50 can provide excellent predictions when trained and applied to Alzheimer's MRI imagery. The study team saved the trained weights for use in additional medical image classification and is providing the code for replication.

There are several limitations of this study.   First, only 51 slices of each patient's MRI were used in the imagery analysis.   Doing so might have eliminated slices that were helpful in diagnosing AD. Second, the actual number of individuals in the study, 416, is small (although the number of images is certainly sufficient). This limits the generalizability of the findings for both the GBM and ResNet models, as they would need re-training on an additional dataset.   This retraining is not problematic, but slightly different classification accuracies would be likely.

Future work will investigate the classification capability of the pre-trained ResNet-50 on some additional datasets.   Further, some modifications of ResNet, implementation of other network schemes, and early identification based on longitudinal analysis will be explored.

What is noteworthy here is that analysis with and without imagery result in reasonable classification accuracy. The implication here is that AD might be detected by the MMSE and the associated socio-demographics of the patient early on without imagery evaluation, potentially reducing costs. Further, feeding information already collected into ML algorithms might help identify patients at risk. In addition, application of ML image recognition to MRIs will successfully code and classify AD for confirmation by providers with high accuracy, even if a patient's intended visit was not for AD.   ML algorithms may flag these images automatically for review.

References

1. Mayeux, R., & Sano, M. Treatment of Alzheimer's disease. *The New England Journal of Medicine* **1999**, *341*, 1670–1679, doi:10.1056/NEJM199911253412207.

2. Alzheimer's Association. Facts and Figures. Availabe online: https://www.alz.org/alzheimers-dementia/facts-figures (accessed on October 15,2018).

3. VanMeter, K.; Hubert, R.J. *Gould's Pathophysiology for the Health Professions (6th ed.)*; Elsevier: St. Louis, Missouri, 2017.

4. Bhagwat, N.; Viviano, J.D.; Voineskos, A.N.; Chakravart, M.M. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLOS Computational Biology* **2018**, *14*, 1-25, doi:10.1371/journal.pcbi.1006376.

5. Amino. Brain MRI Costs. Availabe online: https://amino.com/blog/brain-mri-cost (accessed on 15 October 2018).

6. Driscoll, I.; Troncoso, J. Asymptomatic Alzheimer's Disease: A Prodrome or a State of Resilience? *Current Alzheimer Research* **2011**, *8*, 330-335.

7. Rubin, E.H.; Storandt, M.; Miller, J.P.; Kinscherf, D.A.; Grant, E.A.; Morris, J.C.; Berg, L. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives Of Neurology* **1998**, *55*, 395-401.

8. Aljondi, R.; Szoeke, C.; Steward, C. A decade of changes in brain volume and cognition. *Brain Imaging and Behavior* **2018**, *Springer Nature 2018*, doi:0.1007/s11682-018-9887-z.

9. Huang, L.; Jin, Y.; Gao, Y.; Thung, K.-H.; Shen, D. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of Aging* **2016**, *46*, 180-191, doi:10.1016/j.neurobiolaging.2016.07.005.

10. Small, B.; Backman, L. Longitudinal trajectories of cognitive change in preclinical Alzheimer's disease: A growth mixture modeling analysis. *Cortex* **2007**, *43*, 826-834.

11. Zhang, Y., Dong, Z., Phillips, P. , Wang, S., , Ji, G., Yang, J., Yuan,T. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Frontiers in Computational Neuroscience* **2015**, *9*, 1-13, doi:10.3389/fncom.2015.00066.

12. Zhang, Y.; Wang, S.; Phillips, P.; Yang, J.; Yuan, T. Three-Dimensional Eigenbrain for the Detection of Subjects and Brain Regions Related with Alzheimer's Disease. *Journal of Alzheimer's Disease* **2016**, *50*, 1163-1179, doi:10.3233/JAD-150988.

13. Bansal, D.; Chhikara, R.; Khanna, K.; Gupta, P. Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia. *Procedia Computer Science* **2018**, *132*, 1497-1502, doi:10.1016/j.procs.2018.05.102.

14. Bhagwat, N.; Pipitone, J.; Voineskos, A.; Chakravarty, M. An artificial neural network model for clinical score prediction in Alzheimer disease using structural neuroimaging measures. *J Psychiatry Neurosci* **2019**, *5*, 1-15, doi:10.1503/jpn.180016.

15. Collij, L.E.; Heeman, F.; Kuijer, J.P.A.; Sanz-Arigita, E.J.; Van Berckel, B.N.M.; Barkhof, F.; Wink, A.M.; Ossenkoppele, R.; Benedictus, M.R.; Möller, C., et al. Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. *Radiology* **2016**, *281*, 865-875, doi:10.1148/radiol.2016152703.

16.   Oasis. OASIS Open Access Series of Imaging Studies. Availabe online: http://www.oasis-brains.org/ (accessed on April 25 2019).

17.   Marcus, D.S.; Wang, T.H.; Parker, J.; Csernansky, J.G.; Morris, J.C.; Buckner, R.L. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal Of Cognitive Neuroscience* **2007**, *19*, 1498-1507, doi:10.1162/jocn.2007.19.9.1498.

18.   R Core Team. *R: A Language and Environment for Statistical Computing*; Vienna, Austria, 2016.

19.   Python Software Foundation. Python Language Reference, version 2.7.

20.   Rubin, E.H., Storandt, M., Miller, J. P., Grant, E. A., Morris, J. C., Berg, L. . A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Arch Neurol* **1998**, *55*, 395-401, doi:10.1001/archneur.55.3.395.

21.   Buckner, R.L.; Head, D.; Parker, J.; Fotenos, A.F.; Marcus, D.; Morris, J.C.; Snyder, A.Z. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **2004**, *23*, 724-738, doi:10.1016/j.neuroimage.2004.06.018.

22.   Morris, J.C. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **1993**, *43*, 2412-2414, doi:10.1212/WNL.43.11.2412-a.

23.   Zujun, H. A Review on MR Image Intensity Inhomogeneity Correction. *International Journal of Biomedical Imaging* **2006**, *2006*, 1-11, doi:10.1155/IJBI/2006/49515.

24.   Talairach J; Tournoux, P. *Co-planar stereotaxic atlas of the human brain*; Thieme: New York: 1988.

25.   The Talairach Project. Talairach Coordinate Space. Availabe online: http://www.talairach.org/about.html (accessed on

26.   Cao, J.; Su, Z.; Yu, L.; Chang, D.; Li, X.; Ma, Z. Softmax Cross Entropy Loss with Unbiased Decision Boundary for Image Classification. IEEE: 2018; 10.1109/CAC.2018.8623242p 2028.

27.   Zhang, Y.; Yao, J. Gini objective functions for three-way classifications. *International Journal of Approximate Reasoning* **2017**, *81*, 103-114, doi:10.1016/j.ijar.2016.11.005.

28.   Sakata, R.; Ohama, I.; Taniguchi, T. An Extension of Gradient Boosted Decision Tree Incorporating Statistical Tests. IEEE: 2018; 10.1109/ICDMW.2018.00139p 964.

29.   Chen, L. Basic Ensemble Learning (Random Forest, AdaBoost, Gradient Boosting)- Step by Step Explaineds. **2019**.

30.   Tianqi, C.; Tong, H.; Michael, B.; Vadim, K.; Yuan, T.; Hyunsu, C.; Kailong, C.; Rory, M.; Ignacio, C.; Tianyi, Z., et al. xgboost: Extreme Gradient Boosting. R package version 0.82.1. Availabe online: https://CRAN.R-project.org/package=xgboost (accessed on 15 October 2018).

31.   Álvarez, I.; Górriz, J.M.; Ramírez, J.; Salas-Gonzalez, D.; López, M.; Segovia, F.; Puntonet, C.G. Alzheimer's diagnosis using eigenbrains and support vector machines. *Electronics Letters* **2009**, *45*, 342-343, doi:10.1049/el.2009.3415.

32.   Poli, R.; Citi, L.; Salvaris, M.; Cinel, C.; Sepulveda, F. Eigenbrains: The free vibrational modes of the brain as a new representation for EEG. *32nd Annual International Conference of the IEEE EMBS* **2010**, 6011-6014.

33.   Francois, C. Keras Tensorflow. *GitHub Repository* **2015**.

34.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. IEEE: 2016; 10.1109/CVPR.2016.90p 770.

35.   Fung, V. An Overview of ResNet and its Variants. **2017**.

36.   Qiu, A.; Younes, L.; Miller, M.I. Principal Component Based Diffeomorphic Surface Mapping. *IEEE Transactions on Medical Imaging* **2012**, *31*, 302.