ORIGINAL RESEARCH

A simple classification framework for predicting Alzheimer's disease from region-based grey matter volume and APOE genotype status

Reyhaneh Ghoreishiamiri*[1], Graham Little[1,2], Matthew R. G. Brown[1,3], Russell Greiner[1,4]

[1] *Department of Computing Science, University of Alberta, AB, Edmonton, Canada*
[2] *Department of Biomedical Engineering, University of Alberta, AB, Edmonton, Canada*
[3] *Department of Psychiatry, University of Alberta, AB, Edmonton, Canada*
[4] *Alberta Machine Intelligence Institute, University of Alberta, Edmonton, Canada*

Received: September 29, 2019 **Accepted:** November 11, 2019 **Online Published:** January 12, 2020
DOI: 10.5430/air.v8n2p15 **URL:** https://doi.org/10.5430/air.v8n2p15

ABSTRACT

Alzheimer's Disease (AD) is a prevalent neurodegenerative disease currently affecting more than 47 million people in the world. There are now many complex classifiers that can accurately distinguish AD patients from healthy controls, based on the subject's structural magnetic resonance imaging (MRI) brain scan. Most such automated diagnostic systems are blackboxes: While their predictions are accurate, it is difficult for clinicians to interpret those predictions, due to the large number of features used by the classifier, and/or by the complexity of that classifier. This work demonstrates that an automated learning algorithm can produce a simple classifier that can correctly distinguish AD patients from healthy controls (HC) similar to its more-complex counterparts. Here we build this classifier from the data in the Alzheimer's Disease Neuroimaging Initiative database, using a fairly small set of features, including grey matter volumes of 33 regions of interest derived from structural MRI, as well as the APOE genotype. We first considered three simple base-learners that each produce a classifier that is simple and interpretable. Running our overall learner, involving standard feature selection processes and these simple base-learners, on these features, produced a 7-feature elastic net model, EN7, that achieved accuracy of 89.28% on the test set. Next, we ran the same overall learner using two more-complex base-learners over the same initial dataset. The accuracy of the best model here was 90.47%, which was not statistically different from the performance of our much simpler EN_7 model.

Key Words: Alzheimer's disease, Machine learning, Medical informatics, Simplicity, Linear models

1. INTRODUCTION

Alzheimer's Disease (AD) is a highly prevalent neurodegenerative disease affecting an estimated 5.5 million Americans.[1] Patients with AD experience progressive cognitive impairment associated with patterns of structural brain atrophy more severe than the volumetric loss typical of healthy aging populations, but some of these structural changes, which are detected with MRI in a clinical settings, may not be visible to a clinician's eye until the late stages of the disease. The high prevalence of AD combined with downstream progressive impairment has motivated investigations into advanced diagnosis strategies capable of early detection

*Correspondence: Reyhaneh Ghoreishiamiri; Email: ghoreish@ualberta.ca; Address: Department of Computing Science, University of Alberta, AB, Edmonton, Canada.

of the disease. Moreover, recent investigations that apply machine learning techniques to structural brain imaging have shown promise in accurately discriminating AD patients from controls. Multiple studies have used voxel-based morphometry (VBM) to distinguish AD patients from controls. In four consecutive studies, VBM features combined with different feature selection methods showed high prediction accuracies of 89% to 96%.[2–5] VBM combined with texture analysis features has also been successful in classifying AD patients achieving 92.86% accuracy.[6] Another study, using multimodal features – including voxel-wise structural MR and FDG-PET imaging features, CSF biomarkers, cognitive scores and APOE genotype data – to predict conversion of Mild Cognitive Impairment (MCI) patients to Alzheimer's disease, achieved an accuracy of 92.4%.[7] Additionally, applying a 3D convolutional neural network (CNN) on 3D T1-weighted structural MR images from ADNI has shown 99.2% accuracy, 99.5% specificity, and 98.5% sensitivity[8] in classifying AD patients vs controls (this is the current best accuracy result with this dataset on classifying AD patients vs controls to our knowledge). Despite achieving impressive accuracies of 92% to 99% in classifying AD patients, all of these methods use a large number of features (ranging from 100 to 2000) and complicated diagnostic models that are difficult for a human to interpret;[9] see Section 5.1.

Regional brain features, based on cortical/subcortical segmentation, involve many fewer variables than voxel-based features; this is more appropriate for simple classification models. One study – which combined segmentation-based features of cortical thickness, cortical area, cortical curvature, grey matter density, subcortical volumes and hippocampal shape – achieved 0.98 AUC (Area Under the Curve of the Receiver Operating Curve [ROC]).[10] A multimodal study achieved 93% accuracy by combining region-based features from structural MRI, FDG-PET and CSF proteins (189 features total).[11] Such models, which use the features for all brain regions (one feature for each region, at least), involve a total of hundreds of brain features; this means they are necessarily very complex.

A priori selection of the brain regions typically impacted by AD has been a successful strategy for reducing the complexity of classification models. Using 6 features – namely the left and right hippocampus volume, amygdala volume and entorhinal cortical thicknesses – a support vector machine (SVM) classifier with the radial basis function (RBF) kernel, scored 0.89 AUC,[12] suggesting that only few brain features are needed to discriminate AD from healthy controls. Additionally, grey matter volumes and diffusion-based MRI parameters over predetermined brain regions have shown utility in classifying MCI patients from healthy controls, achieving 89.7% accuracy.[13] Taken together, these studies suggest that simple classification models, on a limited number of brain features, are sufficient to discriminate AD patients from controls. However, more work is needed to assess whether simple predictive models involving a limited number of features can achieve classification results comparable to those of more complex diagnostic models.

This study explores the challenge of learning a simple classifier that can accurately distinguish patients with AD from healthy controls. We also considered learning a model that involved MCI patients – e.g., distinguished MCI from AD, and MCI from Control. However, we realized that this was problematic as there is tremendous variability among human clinicians, who supply the labels. We therefore decided to focus on AD vs control, as these labels are much more consistent. Section 2 describes the dataset we used, that includes a small set of 33 brain volumes along with the APOE genotype status. Section 3 describes how we use that database to produce an effective classifier. This involves pre-processing and feature extraction step, baselearners, feature selection methods (all performed "in-fold"), evaluation method, and overall learners. Section 4 presents our training, test, and feature selection results. Using these 34 features, we first compare the value of applying our overall learner, involving standard feature selection processes, with 3 base-learners selected for their simplicity – decision tree, elastic net and linear SVM – versus 2 relatively complicated base-learners – SVM with RBF kernel and extreme gradient boosting learner. Section 5 then compares the result of our simple classifier, produced by our overall learner with the simple base-learners, to other studies that also classify Alzheimer's patients versus healthy controls.

2. PARTICIPANTS / IMAGING DATA

This analysis used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, with the primary goal of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).[14, 15] For up-to-date information, see www.adni-info.org.

Out of initially 793 subjects who had baseline scans, 35 subjects were excluded due to segmentation problems and 6 due to missing APOE genotype data. Our analysis considers (n = 752) individuals from various ADNI projects, described using their baseline MR imaging and genetic sequencing (including the APOE genotype status, indicating

the number of the APOE4 alleles at the APOE gene locus) – including 337 with a diagnosis of AD (age 75.26 ± 7.81, 44.8% male) and 415 controls (age 74.79 ± 5.72, 49.6% male). Mini-mental state examination (MMSE)[16] scores as well as clinical dementia rating sum of boxes (CDRSB)[17] scores were collected for all subjects. MMSE scores ranged from 18-28 for AD patients and from 24-30 for controls, and CDRSB scores ranged from 1-10 for AD patients and from 0-1 for controls.

The ADNI data was acquired from 60 different sites across US and Canada. To demonstrate our methodology, we merged data from 48 sites to become the training set, ADNI_TRAIN (with n = 584 subjects) and merged the remaining 12 sites into the held-out test set, ADNI_HO (with n = 168 subjects). Note these two sets are disjoint. Table 1 provides demographic information, showing the sex, age, diagnostic distributions and other information, for the training and test sets.

Table 1. Demographic information for the participants included in the training and test datasets (divided based on the data acquisition sites). Numbers for age, MMSE, CDRSB, and the number of APOE4 alleles are each shown as mean ± STD

	Training Set (ADNI_TRAIN)		Test Set (ADNI_HO)	
	AD	HC	AD	HC
Number	149 ♂/ 113 ♀	160 ♂/162 ♀	37 ♂/ 38 ♀	49 ♂/44 ♀
Age	75.04 ± 7.87	75.03 ± 5.66	74.54 ± 7.40	73.92 ± 5.90
MMSE	23.16 ± 2.08	29.07 ± 1.13	23.42 ± 2.07	29.06 ± 1.09
CDRSB	4.45 ± 1.64	0.03 ± 0.13	4.25 ± 1.71	0.03 ± 0.11
Number of APOE4	0.86 ± 0.70	0.31 ± 0.53	0.84 ± 0.75	0.25 ± 0.43

3. MATERIALS AND METHODS

3.1 Image acquisition and segmentation

As part of the ADNI data collection, standardized structural MR imaging data was acquired for all participants using a sagittal magnetization-prepared rapid gradient echo sequence, with 1 × 1 mm^2 in-plane resolution, 1.2 mm slice thickness, and a field of view of 192 × 192 mm^2.[18] The scanner field strength varied from 1.5 to 3.0 T, depending on the site.[18] Here, we focused on the participants' baseline imaging data. We used the Freesurfer (version 5.3) segmentation pipeline to extract regional cortical and subcortical volumetric measurements from each subject's MRI scan.[19] This process generated 68 regional cortical volumes (34 in each hemisphere), as well as 43 subcortical volumes for each subject. To compensate for the possible existing variance in the brain sizes of individuals, we used the normalized percent volumes of each region, which is 100 times the ratio of the volume of that region, divided by the total intracranial volume.

From the segmentation output, we selected the 33 normalized regional brain volumes that have shown robust group-level differences in previous imaging studies of AD: 10 subcortical regions (left and right thalamus,[20] putamen,[20] amygdala,[21] hippocampus[22] and lateral ventricles[22]), 10 medial temporal regions[21] (right and left parahippocampal gyrus, entorhinal cortex, inferior temporal gyrus, middle temporal gyrus and superior temporal gyrus regions), 8 parietal regions[23] (left and right posterior cingulate gyrus, isthmus of cingulate gyrus, inferior parietal lobule and precuneus), 3 callosal regions[24] (posterior, central and anterior corpus callosum), and bilateral cerebellar cortex.[25] In addition to these brain imaging results, Corder et al.[26] showed that the number of the APOE4 alleles at the APOE gene locus is widely associated with late onset Alzheimer's disease. Our dataset therefore described each patient using this one genotype, as well as the normalized grey matter volumes of these 33 regions.

We considered several base-learners. We z-scored each feature to have a mean value of 0 and a standard deviation of 1. Our training dataset describes each subject

$$x = [x_1, \cdots, x_{34}]$$

based on normalized brain volumes from 33 brain regions (x_i for $i \leq 33$), and the APOE genotype status x_{34}.

3.2 Based learning algorithms

We input this data to various base learning algorithms, all implemented in the Python software packages scikit-learn (See the scikit website, http://scikit-learn.org/) and XGBoost (See the XGBoost website, https://xgboost.readthedocs.io). Binary decision trees (DT) are one of the most visually simple classifiers; they are also similar to clinical algorithms used for sequential diagnosis in medicine. We limited the depth of our decision tree to 10 and the number of leaf nodes to 20 for further simplicity and also to reduce the chance of overfitting. We also consider two linear models, each learning a weight vector $W = [w_0, w_1, ..., w_{34}] \in R^{34}$, used in the function

$$y_W(x) = \sum_{i=1}^{34} w_i \times x_i + w_0 \quad (1)$$

Here, the model predicts the subject x has AD if $y_W(x)$ is larger than 0. One linear model, logistic regression with elastic net penalty (EN), was our second simple classifier, combining L1 (Lasso) and L2 (Ridge) regularizers with a ratio (L1 ratio) that weighs the two penalties and an α parameter that weights the penalty term.[27] Support Vector Machines (SVM) is one of the most commonly used classifiers in Alzheimer's prediction studies.[2–5, 11] In this study, we consider SVM classifiers with two different kernels: the

linear and radial basis function (RBF) kernel.[28] The linear SVM (lSVM) is an example of a simple classifier while SVM with RBF kernel (rSVM) is used as an example of a more complex non-linear method. Extreme Gradient Boosting (XGB)[29] is a gradient boosted decision method, which is also used as as an example of a more complex non-linear base-learner in this study.

3.3 Evaluation

Our "overall learner" OL is the system that invokes the pre-processing and feature selection steps, before running the base-learners. It also does the grid-searches to find the best learning algorithm and feature selection parameter settings, and also selects the best of these learners, based on the performance over the training data. We ran OL twice, over different sets of base-learners: once for the simple base-learners – lSVM, EN, DT – and once for the more-complex base-learners – rSVM and XGB; for further clarity, we name the first one S.OL and the second one C.OL. Each of these runs produced a classifier: one simple and one complex (see Figure 1).

Recall we used a training set ADNI_TRAIN, of 584 subjects, and a disjoint test set ADNI_HO, of 168 subjects. In both cases, OL partitioned ADNI_TRAIN into 5 folds and and used the same folds for all 5 base-learners. Each base-learner had to determine the best values for a set of hyper-parameters (described below), and use these when learning its classifier. The learning algorithm identified the best settings for these hyperparameters based on average 5-fold cross-validation (CV) accuracies using grid search.

OL then used the 5-fold cross-validation accuracies of these base-learners (accompanied by various feature selection methods, described in Section 3.4) with these selected hyper-parameters, to identify the best-learners. It then ran that best base-learner on the entire training set to produce classifiers; we then tested the learned classifiers (simple and complex) on the test set, ADNI_HO. We evaluated our final models using specificity, sensitivity, balanced accuracy (average of specificity and sensitivity), and the Receiver Operating Curve (ROC), as well as accuracy. For all three measures, for each model (simple and complex), we reported the performance of that model on the test set, as well as the mean and standard deviation, over the 5 folds of the training set. For SVM methods, the C hyper-parameter was chosen from 1E-5, 1E-4, ..., 1E3, 1E4, and γ was chosen from 1E-6, 1E-5, ..., 1E1, 1E2. For elastic net, the α hyper-parameter was chosen from 1E-4, 1E-3, 1E-2, 1E-1, 2E-1, ..., 9E-1 values and L1 ratio from 5E- 2, 1E-1, ..., 9E-1, 9.5E-1 values. Note that setting the $L1\_ratio$ to 1 means the learner only applies L1 regularization (aka Lasso classification) and setting it to 0 only applies L2 regularization (aka Ridge classification). For the decision tree base-learner, we set the maximum depth of the tree to 10 for further simplicity of our tree model and then used internal cross-validation to find the best values of three hyper-parameters: *minimum samples split* is the minimum percentage of training set instances required to split an internal node, chosen from a range of values between 0.005 to 0.480 (of total number of samples); *minimum samples leaf* is the minimum percentage of instances required to be at a leaf node, chosen from a range of values between 0.005 to 0.480 (of total number of samples); and *maximum number of leaf nodes* controls the width of the tree at its leaf level, chosen from range of 2 to 20. For the extreme gradient boosting learner, the number of tree estimators was chosen from 50, 100, 150, 200, the *maximum depth of the trees* from 2, 4, 6, 8; the learning rate from 0.0001, 0.001, 0.01; the *minimum child weight* (which is the minimum sum of instance weight that is needed in a child) from 1, 3, 5; the *subsample ratio* (which is the subsample ratio of the training instance) from {0.6, 0.7, 0.8, 0.9}; and *column sample by tree* is the subsample ratio of the columns when constructing each of the trees, chosen from {0.6, 0.7, 0.8, 0.9} (See Elastic Net's API, `https://scikit-learn.org/stable/modules/generated/sklearn.linearmodel.ElasticNet.html`).

To statistically compare the accuracy of our classifiers (on ADNI_HO) against each other and see if their classification rates are significantly different, we used the mid-p-value: McNemar test[30] and reported the null hypothesis test result at β = 0.05 significance level, as well as the p-values. Any p-value smaller than β suggests rejection of the null hypothesis.

3.4 Feature selection

First, note that the decision tree learner (similarly, extreme gradient boosting learner, which is a tree-based learner) has its own inherent way of choosing the best subset of features. At each internal node, this learner splits the available training instances based on the feature that best separates the class labels in terms of reducing the Gini impurity criterion.[31] This process stops when the current node is sufficiently pure; this means the resulting decision tree will typically only use a small subset of the features.

The OL system also considered several approaches to learn yet simpler models, which involved fewer features. Here, it explored two filtering feature selection methods, each as a pre-processing step to reduce the number of features that are given to the base-learner (and hence the learned classifier): a simple univariate feature selector (UFS), and minimum redundancy maximum relevance (mRMR).[32] Univariate feature selection method selects the top k features based on

ANOVA (Analysis of Variance) F-values.[33] This method first computes the F-value for each individual feature, and then selects those with top k values. Such univariate feature selection methods, however, do not consider the correlation between the features.[33] The mRMR method addresses this by sequentially seeking a set of features that maximizes the mutual information between each feature and the target classification variable while minimizing the mutual information between the currently selected features.

For the linear models (SVM-linear and elastic net), S.OL also considered the "recursive feature elimination (RFE)" algorithm:[34] a wrapper feature selection method that sequentially removes the least important features, based on the value of learned linear weights – i.e., initially the feature indexed by $i^* = \arg min_{i \in 1,..,34} |w_i|$. (There are fewer feature weights to consider in successive iterations.)

All of our feature selection methods – $i.e.$, UFS, mRMR and RFE – take as input, both the initial dataset and a number k^*, which is the number of features to use. To determine k^*, OL first computes the average cross-validation accuracy $acc(k)$ for each number of features $k \in 1, 2, ..., 34$; then sets $a^* = max_k\{acc(k)\}$ to be the most accurate, and $k^* = \arg max_k\{acc(k)\}$ to be the associated value. To find this k^*, OL of course ran the feature selection methods "in-fold" – $i.e.$, determining the best size-k subset of features for each fold during training. Note that setting k = 34 in each of the of feature selection methods is equivalent to applying no feature selection.

Figure 1 summarizes our method, showing the combinations of base-learners and feature selection methods, within each version of OL.

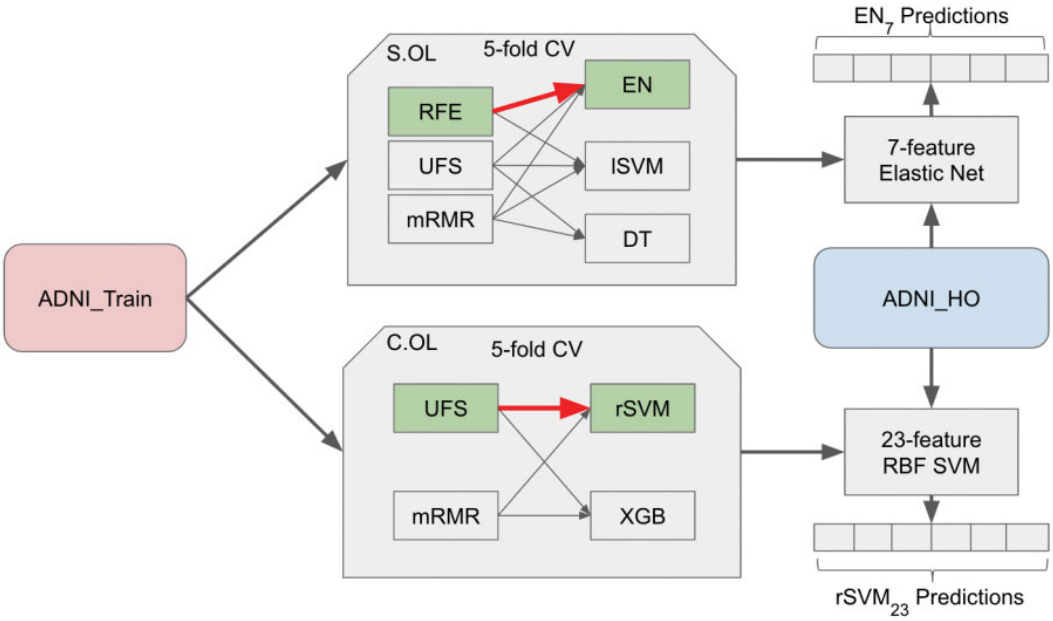

Figure 1. Our overall framework: Left-to-right is the training component, to produce both a simple classifier (here EN_7, on top, using S.OL) and a complicated classifier (here $rSVM_{23}$, on bottom, using C.OL). Each of these OLs considered a set of feature selection methods, from { RFE, UFS, mRMR }, and a given set of possible base-learners – S.OL considered EN, lSVM and DT, while C.OL: rSVM and XGB. (Note that DT did not use RFE). The RED arrow, in each, is the combination of feature selection method and base-learner with the best 5-fold cross-validation accuracy. We then evaluated each of these classifiers, by running each on the (data from) ADNI_HO (vertical, on right)

4. RESULTS

This section first describes our cross-validation results on the training set ADNI_TRAIN (composed of subjects' data from 48 acquisition sites), $i.e.$, the cross-validation accuracies of the best classifiers – one from S.OL and one from C.OL (Section 4.1). This analysis identified the best learners; we then ran just these two resulting classifiers on the independent held-out test set, ADNI_HO; those results appear in Section 4.2. Section 4.3 describes the features selected by the simple classifier, EN_7.

4.1 Cross-validation accuracy on the training set, ADNI_TRAIN

Table 2 and Figure 2 show the mean and standard deviation of the cross-validation performance of our best simple and complex learners, EN_7 and $rSVM_{23}$. Note that the mean cross-validation accuracy, specificity, and sensitivity of the two models are close to each other – i.e., within the boundaries of each other's error bars.

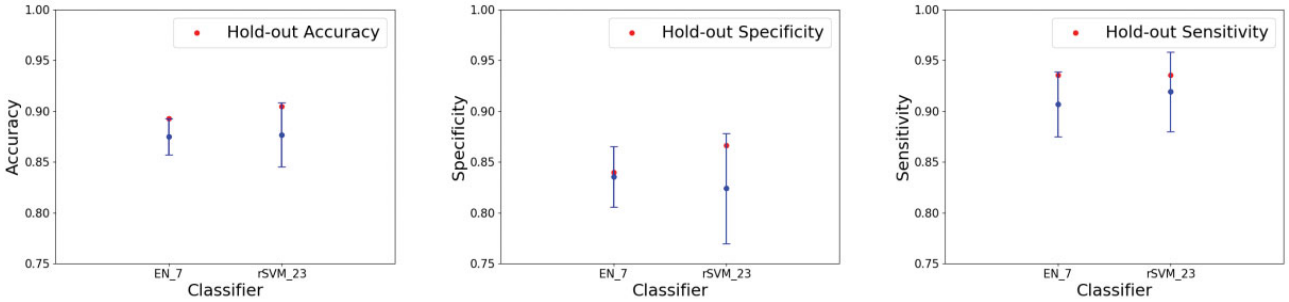

Figure 2. Mean and standard deviation (STD) of the 5-fold cross-validation (CV) performance of EN_7 and $rSVM_{23}$ models, on ADNI_TRAIN. The red dots show the hold-out performance of these models, on ADNI_HO.

Table 2. Mean and standard deviation (STD) of the 5-fold cross-validation (CV) performance of EN_7 and $rSVM_{23}$ models

	EN_7	$rSVM_{23}$
Accuracy	87.50 ± 1.76	87.67 ± 3.16
Specificity	83.56 ± 2.98	82.40 ± 5.44
Sensitivity	90.70 ± 3.19	91.93 ± 3.92

4.2 Results on the Held-Out Test Set, ADNI_HO

As described in Section 3.3, we twice ran our overall learner OL over the training set ADNI_TRAIN (composed of 48 sites, with a total of n = 584 patients) to produce two classifiers – here, the elastic net model with 7-features, EN_7, and the RBF SVM model with 23-features, $rSVM_{23}$. Then, to evaluate and compare the effectiveness of these classifiers, we ran those classifiers on the held-out test set, ADNI_HO (composed of 12 sites, with a total of n = 168 patients). Table 3 shows the test accuracies of these two produced classifiers, along with the result of their statistical comparison, based on the McNemar test. Since the p-value of the McNemar test was above .05 (0.4531), no statistical difference was found between the accuracy of the simple and complex models on the held-out dataset (ADNI_HO). Additionally, Figure 3 shows the ROC curves of the two produced classifiers.

Table 3. Test (hold-out) results using EN_7 and rSVM23 models and the p-value of the statistical comparison of their accuracy based on McNemar test

	Model		p-value
	EN_7	$rSVM_{23}$	
Accuracy	89.28	90.47	0.4531
Specificity	84.00	86.66	
Sensitivity	93.54	93.54	
Balanced accuracy	88.77	90.10	

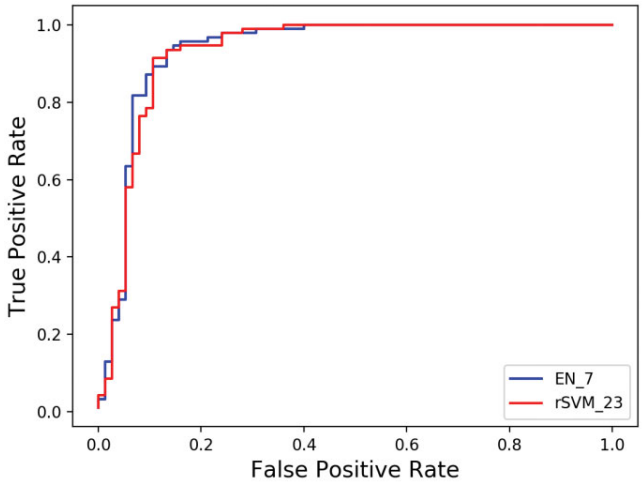

Figure 3. The receiver operation curve (ROC) for our EN_7 and $rSVM_{23}$ models

4.3 Feature importance, based on EN_7

EN_7 selected APOE and 6 brain regions; Figure 4 shows the locations of the 6 regions, and Table 4 shows their associated weights, which corresponds to their "importance". Appendix A shows the hold-out (on ADNI_HO) classification results of each of the base-learners, but using different feature sets: (1) left and right (bilateral) hippocampus regions, (2) only APOE genotype, (3) the 6 regions, without the APOE genotype, as well as (4) the results based on all 7 features.

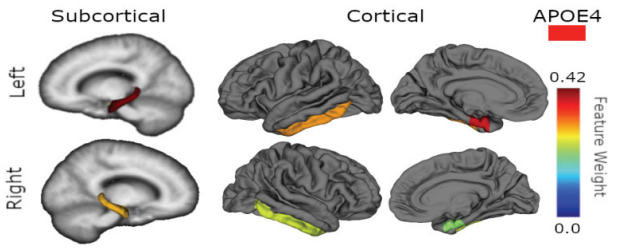

Figure 4. Locations of the features used by the EN_7 models. Color is based on the absolute value of EN_7's weight of the feature

Table 4. EN_7's weights for the features. (2nd column refers to the location in Figure 4.)

Name	Location	Weight
Left hippocampus	top left	0.4221
Right hippocampus	bottom left	0.2896
Left inferiortemporal volume	top middle	0.3036
Right inferiortemporal volume	bottom middle	0.2535
Left entorhinal volume	top right	0.3802
Right entorhinal volume	bottom right	0.2181
APOE		-0.3621

5. Discussion and analysis

5.1 Performance of simple Alzheimer's disease classification (EN_7)

In this study, we applied various machine learning algorithms to APOE genotype status, and regional grey matter volumes from 33 brain regions (that previous clinical studies have shown to be influenced by progression of AD) to learn a model that can predict Alzheimer's disease. We considered five base learners (including three simple models within S.OL). We also considered the effect of feature selection.

As noted in Section 1, there are many previous studies on prediction of AD using structural MRI. Most of these previous studies concentrated on either achieving high prediction accuracy or mere simplicity (by using the grey matter volumes for a very small number of recognized brain regions), but this study is an attempt to create a balance between prediction accuracy and simplicity of the prediction framework. An earlier study demonstrated that a 3D convolutional neural network could achieve high accuracy (99%) using a large number of voxel-based features.[8] There were other region-based studies using structural MRI data, either combining a variety of measures with regional grey matter volumes, including cortical thickness, surface area and cortical curvature,[10] or combining regional data from different imaging modalities,[11] that can achieve high classification performances of 0.98% AUC and 93% accuracy, respectively. The problem with these approaches is that a system that involves too many features might not be used in a clinical environment. This means that even though they have high accuracies, clinicians may be uncomfortable using them because clinicians might find it difficult to understand processes that involve a large number of features. In another study, Jongkreangkrai et al.[12] learned an RBF-kernel SVM over bilateral hippocampus and amygdala grey matter volumes and entorhinal cortical thickness features. This achieved an AUC of 0.89, which is especially impressive as it used only 6 features. However, the resulting "SVM with RBF kernel" classifier involves a complex combination of the features, which prevents users from reasoning about the influence of each feature. By contrast, it is easy to reason about linear classifiers (Equation 1) as the sign of the coefficient w_i tells whether that feature's value x_i increases the risk of AD, or decreases i_t; see Table 4.

5.2 Explaining EN_7's feature selection results

Table 4 shows the weights for the 7 features that appeared in EN_7. Jongkreangkrai et al.[12]'s 6-feature RBF SVM model, mentioned in Section 5.1, also used the cortical thickness for 4 of these brain regions: left and right hippocampus and entorhinal cortex. However, there is no easy way to read off the influence of a variable, nor even the directionality, in non-linear models, like RBF SVM or decision trees, in general. This is possible in linear models, such as EN: here finding a feature whose associated weight is positive, means the chance of AD increases as that variable's value increases, mutatis mutandis.

Previous studies on dynamics of grey matter loss in Alzheimer's disease suggest that bilateral hippocampus regions are areas of the brain that are most strongly affected by AD, which makes them appear as the most discriminating features in the classification task.[35] We also saw that the feature ranks of bilateral regions are not similar to each other. This is consistent with the findings of clinical studies that claim grey matter loss in AD is asymmetric.[36] Studies claim that entorhinal cortex, which is the gateway to hippocampus, is one of the first areas that AD begins to affect, which suggests that grey matter volume for this area may help identify patients at early stages of AD.[37] All of these 6 marked areas were located in the temporal lobe, which is consistent with the previous literature on diagnosis of AD.[38] Genetic studies show that the number of the APOE4 alleles at the APOE gene locus is strongly associated with late onset Alzheimer's disease,[26] explaining why APOE genotype appears among discriminating features in our AD diagnosis prediction framework.

6. Conclusion

In this study, we attempted to build a classification framework for learning a simple model that can accurately distinguish patients with Alzheimer's disease from healthy controls. The performance results, on the 168 subjects in our test set, show that a learned simple linear classifier using only a small set of features – grey matter volume for 6 brain regions and a single genotype datum – can accurately distinguish Alzheimer's patients from controls. We found that the APOE genotype status had one of the highest feature importance in our linear classifiers; its inclusion in the set of imaging features (grey matter volumes) improved the performance of our models.

Although we started from 34 features that were already identified as relevant to AD, we provide a learned linear classifier using just 7 of these features that is statistically as accurate

as its more complex counterparts. The best accuracy on this task and dataset in the literature (99%)[8] was achieved using a much more complicated, non-interpretable model (convolutional deep neural network). Our simple method, achieving 89% accuracy, approaches clinical relevance, which justifies future research into simple systems whose decision process would be accessible to clinicians and could help improve clinical diagnosis. As a line of future work, it would be also valuable to explore the idea of decision fusion[39] to find if combining the decisions made by our simple base-learners would further improve the performance of our framework.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding from NSERC (Natural Sciences and Engineering Research Council of Canada), AMII (Alberta Machine Intelligence Institute), and IBM CAS (IBM Centre for Advanced Studies)(Edmonton). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] Alzheimer's Association, 2017 alzheimer's disease facts and figures. Alzheimer's & Dementia. 2017; 13(4): 325–373. https://doi.org/10.1016/j.jalz.2017.02.001

[2] Beheshti I, Demirel H, Initiative ADN, et al. Probability distribution function-based classification of structural mri for the detection of alzheimer's disease. Computers in Biology and Medicine. 2015; 64: 208–216. PMid:26226415. https://doi.org/10.1016/j.compbiomed.2015.07.006

[3] Beheshti I, Demirel H, Farokhian F, et al. Structural mribased detection of alzheimer's disease using feature ranking and classification error. Computer Methods and programs in Biomedicine. 2016; 137: 177–193. PMid:28110723. https://doi.org/10.1016/j.cmpb.2016.09.019

[4] Beheshti I, Demirel H, Initiative ADN, et al. Feature-ranking-based alzheimer's disease classification from structural MRI. Magnetic Resonance Imaging. 2016; 34(3): 252–263. PMid:26657976. https://doi.org/10.1016/j.mri.2015.11.009

[5] Beheshti I, Demirel H, Matsuda H, et al. Classification of alzheimer's disease and prediction of mild cognitive impairment-to-alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. Computers in Biology and Medicine. 2017; 83: 109–119. PMid:28260614. https://doi.org/10.1016/j.compbiomed.2017.02.011

[6] Ding Y, Zhang C, Lan T, et al. Classification of alzheimer's disease based on the combination of morphometric feature and texture feature, in: Bioinformatics and Biomedicine (BIBM). 2015 IEEE International Conference on, IEEE. 2015, pp. 409–412. https://doi.org/10.1109/BIBM.2015.7359716

[7] Hinrichs C, Singh V, Xu G, et al. Predictive markers for ad in a multi-modality framework: an analysis of mci progression in the adni population, Neuroimage. 2011; 55(2): 574–589. PMid:21146621. https://doi.org/10.1016/j.neuroimage.2010.10.081

[8] Basaia S, Agosta F, Wagner L, et al. Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks. NeuroImage: Clinical. 2018: 101645. PMid:30584016. https://doi.org/10.1016/j.nicl.2018.101645

[9] Ribeiro MT, Singh S, Guestrin C. Why should I trust you?: Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. 2016. p. 1135–1144. https://doi.org/10.1145/2939672.2939778

[10] de Vos F, Schouten TM, Hafkemeijer A, et al. Combining multiple anatomical mri measures improves alzheimer's disease classification, Human brain mapping. 2016; 37(5): 1920–1929. PMid:26915458. https://doi.org/10.1002/hbm.23147

[11] Zhang D, Wang Y, Zhou L, et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. Neuroimage. 2011; 55(3): 856–867. PMid:21236349. https://doi.org/10.1016/j.neuroimage.2011.01.008

[12] Jongkreangkrai C, Vichianin Y, Tocharoenchai C, et al. Computer-aided classification of alzheimer's disease based on support vector machine with combination of cerebral image features in mri. Journal

of Physics: Conference Series. IOP Publishing. 2016; 694; 012036. https://doi.org/10.1088/1742-6596/694/1/012036

[13] Zhang Y, Schuff N, Camacho M, et al. Mri markers for mild cognitive impairment: comparisons between white matter integrity and gray matter volume measurements. PloS one. 2013; 8(6): e66367. PMid:23762488. https://doi.org/10.1371/journal.pone.0066367

[14] Weiner MW, Aisen PS, Jack CR, et al. The alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimer's & Dementia. 2010; 6(3): 202–211. PMid:20451868. https://doi.org/10.1016/j.jalz.2010.03.007

[15] Jones-Davis DM, Buckholtz N. The impact of adni: What role do public-private partnerships have in pushing the boundaries of clinical and basic science research on alzheimer's disease? Alzheimer's & dementia: the journal of the Alzheimer's Association. 2015; 11(7): 860. PMid:26194319. https://doi.org/10.1016/j.jalz.2015.05.006

[16] Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research.

[17] Hughes CP, Berg L, Danziger WL, et al. A new clinical scale for the staging of dementia. The British Journal of Psychiatry. 1982; 140(6): 566–572. PMid:7104545. https://doi.org/10.1192/bjp.140.6.566

[18] Jack CR, Bernstein MA, Fox NC, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging. 2008; 27(4): 685–691. PMid:18302232. https://doi.org/10.1002/jmri.21049

[19] Fischl B. Freesurfer. Neuroimage. 2012; 62(2): 774–781. PMid:22248573. https://doi.org/10.1016/j.neuroimage.2012.01.021

[20] De Jong L, Van der Hiele K, Veer I, et al. Strongly reduced volumes of putamen and thalamus in alzheimer's disease: an mri study. Brain. 2008; 131(12): 3277–3285. PMid:19022861. https://doi.org/10.1093/brain/awn278

[21] St JT, Pruessner JC, Faltraco F, et al, Comprehensive dissection of the medial temporal lobe in ad: measurement of hippocampus, amygdala, entorhinal, perirhinal and parahippocampal cortices using mri. Journal of Neurology. 2006; 253(6): 794–800. PMid:16511646. https://doi.org/10.1007/s00415-006-0120-4

[22] Thompson PM, Hayashi KM, De Zubicaray GI, et al. Mapping hippocampal and ventricular change in alzheimer disease. Neuroimage. 2004; 22(4); 1754–1766. PMid:15275931. https://doi.org/10.1016/j.neuroimage.2004.03.040

[23] Jacobs HI, Van Boxtel MP, Jolles J, et al. Parietal cortex matters in alzheimer's disease: an overview of structural, functional and metabolic findings. Neuroscience & Biobehavioral Reviews. 2012; 36(1): 297–309. PMid:21741401. https://doi.org/10.1016/j.neubiorev.2011.06.009

[24] Di Paola M, Di Iulio F, Cherubini A, et al. When, where, and how the corpus callosum changes in mci and ad a multimodal mri study. Neurology. 2010; 74(14): 1136–1142. PMid:20368633. https://doi.org/10.1212/WNL.0b013e3181d7d8cb

[25] Canu E, Frisoni GB, Agosta F, et al. Early and late onset alzheimer's disease patients have distinct patterns of white matter damage. Neurobiology of Aging. 2012; 33(6): 1023–1033. PMid:21074899. https://doi.org/10.1016/j.neurobiolaging.2010.09.021

[26] Corder E, Saunders A, Strittmatter W, et al. Pericak-Vance, Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. Science. 1993; 261(5123): 921–923. PMid:8346443. https://doi.org/10.1126/science.8346443

[27] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005; 67(2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[28] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995; 20(3): 273–297. https://doi.org/10.1007/BF00994018

[29] Chen T, Guestrin C. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM. 2016. p. 785–794.

[30] Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural computation. 1998; 10(7): 1895–1923. PMid:9744903. https://doi.org/10.1162/089976698300017197

[31] Raileanu LE, Stoffel K. Theoretical comparison between the gini index and information gain criteria. Annals of Mathematics and Artificial Intelligence. 2004; 41(1): 77–93. https://doi.org/10.1023/B:AMAI.0000018580.96245.c6

[32] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence. 2005; 27(8): 1226–1238. PMid:16119262. https://doi.org/10.1109/TPAMI.2005.159

[33] Chen YW, Lin CJ. Combining svms with various feature selection strategies. Feature extraction. Springer. 2006: 315–324.

[34] Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. Machine learning. 2002; 46(1-3): 389–422. https://doi.org/10.1023/A:1012487302797

[35] Jack CR, Petersen RC, O'brien PC, et al. Mr-based hippocampal volumetry in the diagnosis of alzheimer's disease. Neurology. 1992; 42(1): 183–183. PMid:1734300. https://doi.org/10.1212/WNL.42.1.183

[36] Derflinger S, Sorg C, Gaser C, et al. Grey-matter atrophy in alzheimer's disease is asymmetric but not lateralized. Journal of Alzheimer's Disease. 2011; 25(2): 347–357. PMid:21422522. https://doi.org/10.3233/JAD-2011-110041

[37] Van Hoesen GW, Hyman BT, Damasio AR. Entorhinal cortex pathology in alzheimer's disease. Hippocampus. 1991; 1(1): 1–8. PMid:1669339. https://doi.org/10.1002/hipo.450010102

[38] Erkinjuntti T, Lee DH, Gao F, et al. Temporal lobe atrophy on magnetic resonance imaging in the diagnosis of early alzheimer's disease. Archives of Neurology. 1993; 50(3): 305–310. PMid:8442711. https://doi.org/10.1001/archneur.1993.00540030069017

[39] Soriano A, Vergara L, Ahmed B, et al. Fusion of scores in a detection context based on alpha integration. Neural Computation. 2015; 27(9): 1983–2010. PMid:26161815. https://doi.org/10.1162/NECO_a_00766