International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification of Structural MRI for Detecting Alzheimer's Disease

Ayşe Demirhan*[1]

Accepted 3rd September 2016

Abstract:Alzheimer's Disease (AD) is a pathological form of dementia that degenerates brain structures. AD affects millions of elderly people over the world and the number of people with AD doubles every year. Detecting AD years before the effects of disease using structural magnetic resonance imaging (MRI) of the brain is possible. Neuroimaging features that are extracted from the structural brain MRI can be used to predict AD by revealing disease related patterns. Machine learning techniques can detect AD and predict conversions from mild cognitive impairment (MCI) to AD automatically and successfully by using these neuroimaging features. In this study common structural brain measures such as volumes and thickness of anatomical structures that are obtained from The Open Access Series of Imaging Studies (OASIS) and made publicly available by https://www.nmr.mgh.harvard.edu/lab/mripredict are analysed. State-of-the-art machine learning techniques, namely support vector machines (SVM), k-nearest neighbour (kNN) algorithm and backpropagation neural network (BP-NN) are employed to discriminate AD and mild AD from healthy controls. Training hyperparameters of the classifiers are tuned using classification accuracy which is obtained with 5-fold cross validation. Prediction performance of the techniques are compared using accuracy, sensitivity and specificity. Results of the system revealed that AD can be distinguished from the healthy controls successfully using multivariate morphological features and machine learning tools. According to the performed experiments SVM is the most successful classifier for detecting AD with classification accuracies up to 82%.

Keywords:*Alzheimer's Disease, neuroimaging, structural MRI, multivariate analysis, image classification, machine learning techniques*

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia and affects millions of people around the world. AD is not a curable disease but the progress of the disease can be slowed down if it is detected in an early stage. AD causes pathological changes on the brain. These changes can be detected before clinical symptoms begin. Mild cognitive impairment (MCI) is a stage before AD and healthy aging. MCI is likely to turn into AD. 12% of people with MCI convert to AD in a year and 80% of people with MCI convert to AD after 6 years. [1-4].

Structural magnetic resonance imaging (MRI) is a non-invasive imaging technology that is used successfully for detecting AD. MRI is sensitive to the degenerations that AD caused on the brain such as tissue damage or loss. Hippocampus, enthorhinal cortex and posterior cingulate cortex are the brain regions that are most effected from AD. These brain regions are also predictive of transition of MCI to AD. Tissue loss related to AD correlates well with the scores of the clinical cognitive tests that reveal a cognitive decline. High resolution T1-weighted MRI is the best tool to detect hippocampal atrophy. Structural brain measures such as volumes and thickness of anatomical structures are obtained from the T1-weighted MRI and these measures are used to detect any degenerations on the brain regions. MRI-based estimates help to early diagnose of the AD that can be used to slow down the progress of the disease [5-6]

[1] *Electronics and Computer Education, Technical Education Faculty Gazi University, Ankara, Turkey*
** Corresponding Author:Email: ayseoguz@gazi.edu.tr*
Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.

Computational methods are required to predict subjects with AD and subjects who is under risk to show cognitive decline that can turn into AD in years. Machine learning methods are used for this purpose over years to detect AD and predict conversions from MCI to AD. SVM is one of the most popular method for classification of AD vs normal control (NC) and AD vs MCI. Magnin et al. proposed a method to classify AD patients and NC by using a whole-brain MRI analysis. They have extracted features by using a histogram analysis. They have utilized characteristics of the distribution of the brain tissues such as gray matter (GM), white matter (WM) and cerebrospinal tissue (CSF) that gives information about neurodegenerative disease like AD [7]. Cocosco et al. have used pruning strategy to customize a training set that will not affected by anatomical variability and pathology. They have used prior tissue probability maps in a standard stereotaxic space to generate a set of samples, then they reduced fraction of incorrectly labelled samples in this set and used a supervised kNN classifier for classifying the MRI scans using the corrected set of samples [8]. Amoroso et al. utilized BP-NN for classification of 288 subject to discriminate AD vs NC and AD vs MCI. Their method includes three steps. First, they applied rigid registration and histogram based equalization to the MR images. Then they have calculated important features like hippocampal volume or its thickness from a volume of interest that contains both the left and right hippocampi regions. Finally, they utilized a BP-NN for classification and obtained 0.81 overall accuracy [9].

In this study, structural MRI is classified for detecting AD using machine learning methods and structural brain measures such as volumes and thickness of anatomical structures. We utilized SVM, kNN and BP-NN for classification task with 5-fold cross validation.

2. Materials and Methods

2.1. Data

T1-weighted cross-sectional structural brain MRI scans of the OASIS database [10] are used to extract structural brain measures of anatomical structures and made publicly available at https://www.nmr.mgh.harvard.edu/lab/mripredict by Sabuncu and Konukoglu [6]. These morphometric brain features including volumes and thickness of the anatomical structures of the brain are analysed in this study to classify AD and mild AD. 190 subjects from the OASIS database is used to obtain brain measures for binary classification. These subjects were the ones that the automatic image processing steps of FreeSurfer (https://freesurfer.nmr.mgh.harvard.edu) is successful. FreeSurfer is a brain MRI analysis software that is used widely and freely available. Table.1 summarizes the demographic features of the data.

Table 1. Features of the Dataset

Variable	*N per group*	*Age (Mean±Std)*	*%Female*	*Number of sites*	*Variable*
		Cases	*Controls*		
AD	25	77.5±6.8	77.5±6.6	72	1
AD mild	70	75.9±7.3	76±7.2	68.6	1

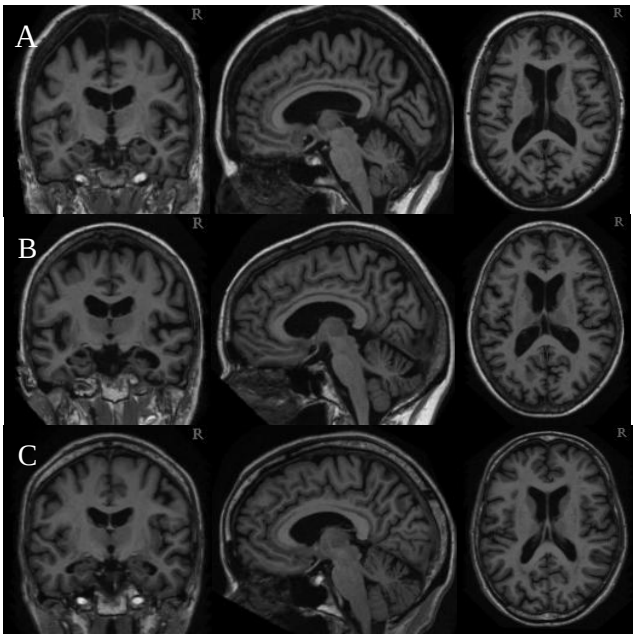

Figure 1. T1-weighted cross-sectional brain MRI scans of subjects with (A) AD;(B) AD mild and (C) NC

Subjects who has clinical dementia rating (CDR) >=1 was defined as AD and CDR>0 as AD mild. AD mild include subjects suffering from MCI and not clinically demented. Figure.1 shows sample MRI scans from the OASIS database belonging to the subjects with AD, AD mild and NC whose subject numbers in the database are OAS1_0003_MR1, OAS1_0035_MR1 and OAS1_0062_MR1, respectively.

2.2. Feature Sets

Four sets of morphological brain features are used for prediction of AD that are obtained using FreeSurfer. Details of feature extraction process can be found at [6]. The brief explanation of the feature sets is given below.

1) Feature Set 1 (aseg): This feature set includes volumes of the 45 anatomical structures. Because of head size variations, these volumes are normalized with each subject's intracranial volume.

2) Feature Set 2 (aparc): This feature set includes 34 average thickness measurements within the cortical parcellations per hemisphere.

3) Feature Set 3 (aseg + aparc): The union of the aseg and aparc feature sets is a 113 dimensional vector.

4) Feature Set 4 (thick): 10,242 vertices per hemisphere is calculated. The cortical thickness values are sampled onto thefsaverage5 template. These values are smoothed on the surface by using an approximate 5 mm Gaussian kernel.

2.3. Support Vector Machines

SVM is the most popular binary classification algorithm used for the prediction of diseases from the structural MRI [4, 11]. SVM learns class differences in a supervised manner by using labelled training samples. SVM classifies a binary labelled data by mapping the data to a very high-dimensional feature space. A separating hyperplane is constructed in the feature space that is used as a decision surface to separate the training data. This hyperplane is defined as wT x+b=0, where b is the bias for the input vector x and w is the weight vector. Informative subsets of the training data are used as support vectors to determine the decision surface. The margin between support vectors are minimized by maximizing‖w‖. Classes are determined by two subspaces that are obtained after training. Figure.2 illustrates support vectors and decision surfaces of a SVM [4, 12].

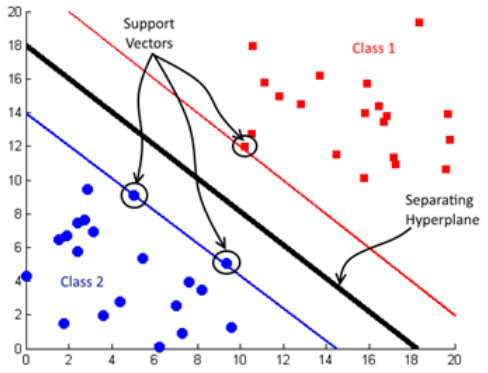

Figure 2. Support vectors and decision surface.

There are three different kernels namely polynomial, radial basis function (RBF) and sigmoid, that are used for nonlinear feature mapping of the SVM. In this study, sequential minimal optimization (SMO) learning algorithm with a Gaussian RBF kernel is used to train the SVM. Equation.1 defines the Gaussian RBF kernel [12, 13]. Here, σ is the width of the RBF function. There is also a regularization parameter C for the soft-margin SVM. Small C values tend to produce a large margin, while C=∞ leads to a hard margin. σ and C parameters determines the prediction accuracy of the SVM classifier with RBF kernel [4].

$$K(x, x') = exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) (1)$$

2.4. K-nearest Neighbour Algorithm

kNN is a supervised classifier that computes closest k training samples of the data point that will be classified in the feature space. The data is classified with the label of the most representative neighbour among the closest ones that are detected by kNN. kNN which is a non-parametric classifier learns from the training data. Large size of the training data increases the prediction performance of the kNN by estimating the true class distributions in feature space [8]. The value of k that is the number of nearest neighbours to the classified data is the key parameter for the algorithm.

2.5. Neural Network

Neural networks are used widely in medical image classification task since there is no need any information related to the probability distribution of the data and a priori probabilities of different classes [14, 15]. Backpropagation (BP) uses a feed-forward and supervised learning algorithm. Feed-forward NN has three main layers. First layer is the input layer. Then there are hidden layers. Each hidden layer has a connection from the previous layer. The final layer calculates the output of the NN. Each layer consists a number of neurons that will map input to the output by updating their weights using gradient descent learning rule. BP algorithm adjusts the weights of the neurons in the steepest descent direction that the performance function decreases most rapidly. Gradient of the error function is computed relative to the hidden units by back propagation of the error.

3. Results and Discussion

Subjects that has AD (N=25) and mild AD (N=70) are discriminated from NC subjects (N=95) by using four different sets of structural brain measures [6] that reveals the degeneration of the AD on the brain which leads to classify normal and abnormal brain scans. SVM, kNN and BP-NN are used for the classification. 5-fold cross validation is applied for all classification tasks to assess the generalizability of the performance. The number of cases and controls were the same for all folds. Results of the classification are evaluated in terms of their accuracy, sensitivity and specificity.

SMO learning function and RBF kernel is employed for SVM. C and σ, the two parameters of the RBF kernel should be determined carefully for a successful classification of SVM. A grid search is employed for selection of the best parameter set with 5-fold cross validation to reduce the selection-related bias. A coarse grid is generated by growing values of C = [2-9,2-8, …, 215] and σ = [2-5,2-4, …, 215]. A finer grid search is performed after identifying a better region on the coarse grid [16]. The best σ and C values obtained from grid search were applied to the whole training data [4]. Average classification accuracies obtained from the SVM are given in Table.2. The most successful feature set was aseg and the most unsuccessful feature set was aparc for the SVM.

kNN classifier is tuned using classification accuracies which is obtained with 5-fold cross validation. Number of nearest neighbors is determined empirically as 10 for classifying each point when predicting. Table.3 gives the performance of the kNN algorithm. The most successful feature set was aseg that includes volumes of the anatomical structures and the most unsuccessful feature set was thick that includes cortical thickness values for the kNN. Including aparc features besides aseg decreased the classification accuracies for both cases.

Table 2. Classification Accuracies of the SVM

Feature Set	*Accuracy*		*Sensitivity*		*Specificity*	
	AD mild	*AD*	*AD mild*	*AD*	*AD mild*	*AD*
aseg	0,821	0,800	0,843	0,800	0,800	0,800
aparc	0,657	0,760	0,729	0,840	0,586	0,680
aseg+aparc	0,743	0,800	0,743	0,840	0,743	0,760
thick	0,671	0,760	0,643	0,800	0,700	0,720

Table 3. Classification Accuracies of the kNN

Feature Set	*Accuracy*		*Sensitivity*		*Specificity*	
	AD mild	*AD*	*AD mild*	*AD*	*AD mild*	*AD*
aseg	0,657	0,800	0,471	0,760	0,843	0,840
aparc	0,643	0,740	0,471	0,680	0,814	0,800
aseg+aparc	0,643	0,740	0,471	0,680	0,814	0,800
thick	0,614	0,680	0,329	0,480	0,900	0,880

Gradient descent learning algorithm is used for training of the BP-NN. One hidden layer with 5 neurons are trained with 5-fold cross validation with 1000 epochs at each fold. These parameters are determined empirically according to their classification performance. Results that are obtained using BP-NN are given in Table.4. OoM means out of memory that BP-NN could not converge up to 1000000 epochs. Different number of hidden layer neurons and learning algorithms did not help BP-NN to converge using *thick* feature set. The most successful feature set was *aseg* for the BP-NN for both AD and AD mild cases.

Table 4. Classification Accuracies of the BP-NN

Feature Set	*Accuracy*		*Sensitivity*		*Specificity*	
	AD mild	*AD*	*AD mild*	*AD*	*AD mild*	*AD*
aseg	0,750	0,800	0,743	0,880	0,757	0,720
aparc	0,621	0,740	0,543	0,720	0,729	0,760
aseg+aparc	0,693	0,700	0,686	0,680	0,700	0,720
thick	OoM	OoM	OoM	OoM	OoM	OoM

Performance comparison of the classifiers and the feature sets is given in Fig. 3. SVM was the most successful classifier independent from the feature set and the aseg was the most successful feature set independent from the classifier type. The highest classification accuracy is achieved with the combination of the SVM and the aseg feature set. Classifying AD mild was more difficult then classifying AD since the disease related changes on the brain is more evident in the AD case.

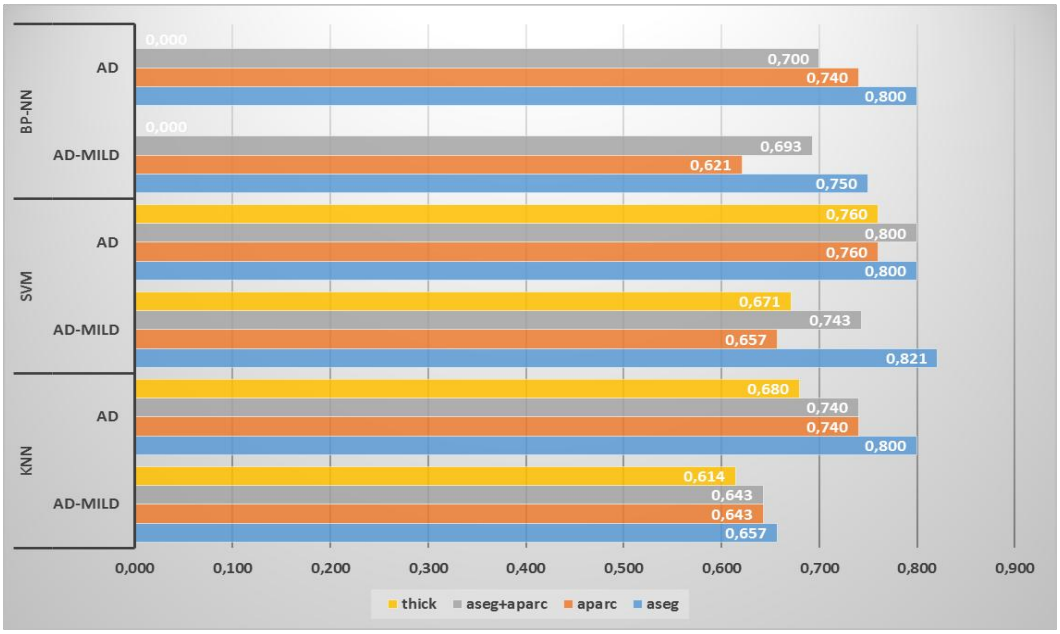

Figure3. Performance comparison of the classifiers and the feature sets

References

[1] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," Journal of Magnetic Resonance Imaging, vol. 27, no. 4, pp. 685-691, 2008.

[2] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," Journal of Internal Medicine, vol. 256, no. 3, pp. 183-194, 2004.

[3] P. M. Thompson, K. M. Hayashi, G. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell, A. W. Toga, "Dynamics of gray matter loss in Alzheimer's disease," Journal of Neuroscience, vol. 23, no. 3, pp. 994-1005, 2003.

[4] A. Demirhan, T. M. Nir, A. Zavaliangos-Petropulu, C. R. Jack Jr., W. M. Weiner, M. A. Bernstein, P. M. Thompson, N. Jahanshad, "Feature selection improves the accuracy of classifying Alzheimer disease using diffusion tensor images," in Proceedings - International Symposium on Biomedical Imaging, 2015, art. no. 7163832, pp. 126-130.

[5] G. B. Frisoni, N. C. Fox, C. R. Jack Jr., P. Scheltens, P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," Nature Reviews Neurology, vol. 6, no. 2, pp. 67-77, 2010.

[6] M. R. Sabuncu, E. Konukoglu, "Clinical Prediction from Structural Brain MRI Scans: A Large-Scale Empirical Study," Neuroinformatics, 13 (1), pp. 31-46, 2015.

[7] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," Neuroradiology, vol. 51, no. 2, pp. 73-83, 2009.

[8] C. A. Cocosco, A. P. Zijdenbos, A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," Medical Image Analysis, vol. 7, no. 4, pp. 513-527, Dec. 2003.

[9] N. Amoroso, R. Errico, R. Bellotti, "PRISMA-CAD: Fully automated method for Computer-Aided Diagnosis of Dementia based on structural MRI data," in Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data, 2014, pp. 16–23.

[10] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, & R. L. Buckner, "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," Journal of Cognitive Neuroscience, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.

[11] R. Casanova, F.-C. Hsu, & M. A. Espeland, Alzheimer's Disease Neuroimaging Initiative, "Classification of Structural MRI Images in Alzheimer's Disease from the Perspective of ill-Posed Problems," PLoS One, vol. 7, no. 10, e44877, 2012.

[12] J. P. Vert, K. Tsuda, B. Schölkopf, "A primer on kernel methods," in Kernel Methods in Computational Biology, 2nd ed., Cambridge, MA: MIT Press, 2004.

[13] J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, M. Gómez-Río, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," Information Sciences, vol. 237, no. 10, pp. 59-72, Jul. 2013.

[14] A. Demirhan, Y. A. Kılıç, İ. Güler, "Artificial Intelligence Applications in Medicine," Turkish Journal of Intensive Care Medicine, vol. 9, no. 1, pp. 31-41, 2010.

[15] Y. Zhang, Z. Dong, L. Wu, S. Wang, "A hybrid method for MRI brain image classification," Expert Systems with Applications, vol. 38, no. 8, pp. 10049-10053, Aug. 2011.

[16] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Department of Computer Science, National Taiwan University, Taipei, Taiwan, Tech. Rep., 2003.