Title: Ensemble based on static classifier selection for automated diagnosis of Mild Cognitive Impairment

Authors: Loris Nanni, Alessandra Lumini, Nicolò Zaffonato

PII: S0165-0270(17)30382-5
DOI: https://doi.org/10.1016/j.jneumeth.2017.11.002
Reference: NSM 7887

To appear in: *Journal of Neuroscience Methods*

Received date: 26-6-2017
Revised date: 31-10-2017
Accepted date: 1-11-2017

Please cite this article as: Nanni Loris, Lumini Alessandra, Zaffonato Nicolò.Ensemble based on static classifier selection for automated diagnosis of Mild Cognitive Impairment.*Journal of Neuroscience Methods* https://doi.org/10.1016/j.jneumeth.2017.11.002

Ensemble based on static classifier selection for automated diagnosis of Mild Cognitive Impairment

Loris Nanni[1], Alessandra Lumini[2], Nicolò Zaffonato[1]

[1]DEI, University of Padua, viale Gradenigo 6, Padua, Italy;

[2]DISI, Università di Bologna, Via Sacchi 3, 47521 Cesena, Italy. E-Mail: alessandra.lumini@unibo.it;

Highlights

- *Early diagnosis of Alzheimer's Disease by machine learning.*
- *Classification of AD from pre-processed sets of T1-weighted MRI*
- *Comparison of different techniques for feature selection*
- *Integration of machine learning methods: SVM, GPC, AdaBoost*
- *Static Classifier Selection using the best features.*

Abstract

Background

Alzheimer's disease (AD) is the most common cause of neurodegenerative dementia in the elderly population. Scientific research is very active in the challenge of designing automated approaches to achieve an early and certain diagnosis. Recently an international competition among AD predictors has been organized: "A Machine learning neuroimaging challenge for automated diagnosis of Mild Cognitive Impairment" (MLNeCh). This competition is based on pre-processed sets of T1-weighted Magnetic Resonance Images (MRI) to be classified in four categories: stable AD, individuals with MCI who converted to AD, individuals with MCI who did not convert to AD and healthy controls.

New Method

In this work, we propose a method to perform early diagnosis of AD, which is evaluated on MLNeCh dataset. Since the automatic classification of AD is based on the use of feature vectors of high dimensionality, different techniques of feature selection/reduction are compared in order to avoid the

curse-of-dimensionality problem, then the classification method is obtained as the combination of Support Vector Machines trained using different clusters of data extracted from the whole training set.

Results

The multi-classifier approach proposed in this work outperforms all the stand-alone method tested in our experiments. The final ensemble is based on a set of classifiers, each trained on a different cluster of the training data. The proposed ensemble has the great advantage of performing well using a very reduced version of the data (the reduction factor is more than 90%).

The MATLAB code for the ensemble of classifiers will be publicly available[1] to other researchers for future comparisons.

***Index Terms*—** Alzheimer's disease; Magnetic Resonance Images; Static Classifier Selection; Ensemble of Classifiers, Feature Selection.

1. Introduction

Alzheimer's disease (AD) is the most common cause of neurodegenerative dementia in the elderly population. The number of people over 60 years of age people suffering from AD is estimated at nearly 44 million worldwide[2] and this number is expected to almost double every twenty years. Because of the dramatic increase in the occurrence of AD, the identification of effective biomarkers for the early diagnosis and treatment of AD in individuals at high risk to develop the disease is crucial. Early and accurate diagnosis has great potential to gives patients access to effective treatments able to slow the clinical progression of AD and can help them maintain their independence for a longer period.

The goal of achieving such an early and accurate diagnosis of AD makes necessary the study of the symptomatic pre-dementia stage of disease, including memory impairment, whereas the underlying

[1] https://www.dropbox.com/s/bguw035yrqz0pwp/ElencoCode.docx?dl=0
[2] http://www.alzheimers.net/resources/alzheimers-statistics/

pathological processes appear to begin and covertly progress several years before. This stage is named Mild Cognitive Impairment (MCI) and consists in a transitional stage between age-related cognitive decline and AD, characterized by memory disturbance in the absence of dementia, followed by widespread cognitive deficits in multiple domains.

Even if prediction of MCI conversion or not-conversion to AD is currently an open issue [1], several studies [2] have revealed that MCI patients convert to AD at an annual rate of 10–15% per year, and among MCI patients who do not convert to AD, most remain stable or develop other forms of dementia and only very few revert to normal status.

In recent years, numerous laboratories and multicenter consortia worldwide [3] have put a considerable effort into the development of advanced neuroimaging processing techniques and other biomarkers to evaluate Alzheimer disease (AD) risk factors and improve the diagnostic confidence of clinical diagnosis.

The National Institute on Aging-Alzheimer's Association (NIA-AA) has recently revised the guidelines for AD [4], including the five most commonly investigated biomarkers in the new diagnostic criteria for AD. These biomarkers can be divided into two categories: 1) measures of brain amyloid, which include cerebrospinal fluid (CSF) measures of Aβ42 and amyloid positron emission tomography (PET) imaging, and 2) measures of neuronal injury and degeneration, which include CSF tau measurement, fluoro deoxyglucose (FDG) PET and structural Magnetic Resonance Imaging (MRI). Of these biomarkers, structural MRI, which allows for accurate measurement of the 3-dimensional (3D) volume of brain structures, especially the size of the hippocampus and related regions, is very important as it is widely available, noninvasive, offers good diagnostic accuracy with moderate costs and it is proven to be able to detect cerebral atrophic regions even before dementia is apparent [5].

MRI processing techniques have been widely studied in the literature with the aim of improving diagnostic accuracy, and in some recent surveys [3][6] several artificial intelligence approaches have been reviewed which can classify pathological and normal subjects from their MRI brain studies,

without requiring a priori hypotheses of where this information may be coded in the images. Unfortunately MRI-based prediction often fails to predict MCI-to-AD conversion [7], therefore it is essential to couple MRI-predictor with other biomarkers. Two important factors for the MRI-based prediction of MCI-to-AD conversion are the use of preprocessing techniques (i.e. spatial registration) to reduce the large intersubject variability, and feature selection.

Several MRI measures have been extracted in the literature [8] for predicting AD and different approaches do not concord in determining the most accurate one; anyway the combination of different MRI-based features has proven to be a feasible way for improving prediction accuracy [9]. In [10] an automated method for measurement of hippocampal subfields and extrahippocampal medial temporal lobe (MTL) cortical regions is presented and the relative involvement of different MTL subregions across a range of cognitive impairment associated with AD is studied. In [11] a novel grading biomarker is proposed based MR images registration, aging effects removal and feature selection, finally, the grading biomarker is combined with age and cognitive measures to improve accuracy.

Diffusion-weighted MR imaging (DWI) examines micro-structural properties of white matter in order to find capable biomarkers that provide sufficient power to predict conversion to AD. Recent studies in time-series analysis of DWI data used a two-level strategy based on "tract-based statistics" to extract image metrics in region-of-interests (ROIs) and a further statistical analysis [12][13].

A fair comparison among the classification studies proposed in the literature requires the availability of benchmark datasets where measuring performance. Many approaches uses different databases [14][15][16] for the prediction of the early diagnosis of AD, the differential diagnosis of MCI and the prediction of its conversion into AD, based on structural MRI features, therefore they are hardly comparable each other. In 2015 Bron and colleagues [17] proposed the CADDementia challenge, where different algorithms were evaluated on a dataset of structural MRI scans from 384 patients (related to a multiclass problem: 112 AD, 131 MCI and 141 CN, with no differentiation among individuals with MCI who converted to AD, and individuals with MCI who did not convert to AD). The training set provided by the challenge was not so large, therefore participants were allowed to

use independent datasets to train their methods. The winner of challenge [18] proposed a method which combines a range of volumetric measurements, cortical thickness measurements, hippocampal texture, and hippocampal shape as features in a linear discriminant analysis (LDA) classifier.

Recently another international competition among AD predictors have been organized, named "A Machine learning neuroimaging challenge for automated diagnosis of Mild Cognitive Impairment" (MLNeCh) and hosted on the web platform Kaggle[3]. The competition is based on pre-processed sets of T1-weighted Magnetic Resonance Images (MRI) consisting of four categories: stable AD, individuals with MCI who converted to AD (cMCI), individuals with MCI who did not convert to AD (MCI) and healthy controls. This classification problem is harder than that proposed in the CADDementia challenge, since the discrimination among cMCI and MCI, not required by CADDementia challenge, is very challenging: for this reason, the expected performance is lower than the 73% AUC obtained in CADDementia challenge by the winner method.

This paper reports a method designed for this machine learning competition, which is a classification system consisting of an ensemble of Support Vector Machine (SVM): the ensemble is built training each SVM with a different set of patterns/features. In order to maximize the diversity of the ensemble, different approaches are tested for improving the performance obtained by a stand-alone SVM:

- Different feature selection approaches are compared.
- The training set is divided in clusters and then a different SVM is trained using the data of each cluster. A given test pattern is assigned to a cluster and then it is classified by the SVM trained with the patterns of that cluster.

2. Material and methods

2.1 Study population

[3] https://inclass.kaggle.com/c/mci-prediction

The dataset used in the experiments is composed by 400 samples equally distributed among the 4 classes healthy controls (HC), Alzheimer's patients (AD), Mild Cognitive Impairment (MCI) who did not convert their diagnosis, and Mild Cognitive Impairment who converted to Alzheimer's disease (cMCI). MRI data were selected and processed from the Alzheimer's disease Neuroimaging Initiative, an international project that collects and validates neurological data. MRIs were analyzed using FreeSurfer v.5.3 [19][4] to extract several features useful for the diagnosis of AD: cortical thickness and subcortical volumes, hippocampal subfields included.

Each sample includes numerical data containing unique id, age, gender, subcortical volumes, cortical thickness, cortical volume, cortical surface area, cortical thickness standard deviation, cortical curvature, hippocampal subfields volumes (lh=left hemisphere, rh=right) and Mini-Mental State Examination score (MMSE). The complete set is split into training (240, unique id TRAIN_XXX) and test (160, unique id TEST_XXX). The training set also includes the Diagnosis field, i.e. the class of a subject: HC (healthy control), AD (Alzheimer's), MCI (Mild Cognitive Impairment without diagnosis conversion), cMCI (MCI converted to AD). The validation of the best approach proposed in this work has been performed according to a 4-fold cross validation protocol on the training set. Notice that all the parameters are optimized using only the training data, since the test set is blind. The demographic characteristics of the 240 study participants taken from the training set are listed in Table 1.

2.2 Methods

In this subsection, we briefly describe the classification system proposed in this work, which is composed by the following steps: the feature selection, feature transform and classification.

2.2.1 Feature selection

The aim of this work is the classification of unknown data represented by a numerical descriptor into four classes. This is done by first training a stand-alone or a multi classifier by using a training

[4] http://freesurfer.net/

set containing labeled samples from the four classes, and then using that classifier to predict the labels of new samples. This procedure is complicated by the high dimensionality of the data: while the descriptors include more than 400 features, the training set includes no more than 240 samples. Standard classifiers do not work well in these situations where the number of features exceeds the number of samples. A commonly used approach to deal with this problem consists in selecting only the features that are most relevant for discriminating between the classes. Therefore, the first step of the proposed classification system is feature selection, which has the objective of retaining the features with higher prediction performance, providing faster and more cost-effective predictors, reducing the curse of dimensionality problem and the possibility of overfitting during the training phase. In this work we test four among the most used feature selection methods available in literature [20]:

- **KernelPLS (KPLS)** [21]: a kernel method based on the partial least squares (abbreviated PLS) that discovers inherent nonlinear correlations among features as well as between features and labels;
- **Fisher score (FS)** [22]: a widely used approach based on discriminative methods;
- **Lagrange Multipliers (LM)** [23]: a method that selects features according to the rank of the corresponding optimal Lagrange multipliers;
- **Mutual Information (MI)** [24]: an approach that estimates the mutual information between features and associated class labels using a quantized feature space.

2.2.2 Classifiers

A factor that strongly influences the accuracy of a classification system is represented by the stand-alone classifier used to make the prediction. Specifically, the performance can be highly variable depending on the type of classification applied, and above all, even in presence of similar prediction accuracy, the predictions of different classifiers can be highly complementary. In this study, we test and compare several classifiers in order design a multiclassifier that exploits the

diversity of single classifiers. The stand-alone methods tested in this work are the following:

- **Support Vector Machine (SVM)** [25] is a binary classifier which performs classification by cutting the n-dimensional feature space into two regions (one per class) by means of an n-dimensional hyperplane that has the largest possible distance from the training vectors of the two classes. SVM maps the data into a higher dimensional input space using kernel functions and constructs an optimal separating hyperplane in this space in order to separate samples of the two classes. This basically involves solving an optimization problem based on the structural risk minimization principle (see [25] for details). Three kernels are tested in our experiments: linear, polynomial and radial basis function. For each kernel, a dataset driven fine-tuning of parameters is performed. The multi-class extension of SVM is obtained according to the one-vs-all approach.
- **Gaussian Process Classifier (GPC)** GPC [26] is a nonparametric classification method, based on a Bayesian methodology, which assumes some prior distribution on the underlying probability densities of the data. This probabilistic approach involves two procedures for approximating inference for binary classification: 1) Laplace's approximation, which is based on an expansion around the mode of the posterior distribution; this used to obtain a prediction for a new point, by estimating its distribution given that we know the posterior of the parameters; 2) the Expectation Propagation algorithm, which is based on matching moments approximations to the marginal of the posterior. The multi-class extension of GPC is performed according to the one-vs-all approach.
- **Random Subspace of Adaboost (RS_AB)** [27] is not a stand-alone approach but an ensemble obtained by the combination of the well-known Random Subspace approach [28] for building ensembles and the AdaBoost.M2 learners. The given dataset suffer from small sample size while its dimensionality is large, therefore to avoid overfitting random subspace ensemble is used. In random subspace ensemble method, the variability is reached by a random selection of features from the whole set of selected features: each single classifier is trained on one subset

of features which is less dimensional than the original feature space. In this work Random Subspace is coupled with AdaBoost.M2. AdaBoost stands for Adaptive Boosting, an algorithm designed for multiclass problems with weak base classifiers, where base classifiers have to minimize the pseudo-loss instead of the error rate [29]. Each weak learner is initially provided with equally weighted training data set, and after training the learner is tested for its pseudo-loss error rate on the training samples in previous iteration. At each iteration the distribution weights are updated in order to force the learners to concentrate more on hard-to-discriminate samples. In our version of RS_AB the output of N=50 weak learners are combined by the sum rule. The AdaBoost.M2 learners are trained on N different subspaces that include 50% of the original features (randomly selected).

- **Random Subspace of Rotation Boosting (RS_RB)** is another ensemble obtained as the Random Subspace version of rotation boosting [30]. The original RotBoost algorithm [14] is a simple integration of Rotation Forest and AdaBoost: the main difference between AdaBoost and Rotation Forest is the method used to perturb the training set: reweighting for AdaBoost, random splitting for Rotation Forest. RotBoost first reduces the dimensionality of the training set and maps original data into a new feature space by a Rotation Matrix (as in RotationForest), then builds weak classifiers according to the AdaBoost technique. In this work according to result reported in [31] we use the Neighborhood Preserving Embedding (NPE) [32] for dimensionality reduction, instead of PCA. The number of learners is, as above, N=50.

2.2.3 Ensemble of Classifiers

In order to improve the performance and reduce the curse of dimensionality problem we propose for this classification problem an ad-hoc strategy to design an ensemble of classifiers based on a variation of the Static Classifier Selection (SCS) [33]: the feature space is divided into selected regions during the training phase, prior to classify the unknown samples. Then, during the testing phase, the region of the test sample is first found and the associated classifier is used for computing the final decision.

In this work the feature space is divided into 4 regions by thresholding the two "best" features (selected on the basis of their discriminative power): MMSE_bl, Left-Hippocampus.

3. Results

The performance indicators used in this work are the area under the ROC curve (AUC) [22] and the accuracy. Since this problem is multiclass, AUC is calculated using the one-versus-all approach.

The first experiment is aimed at comparing the feature selection approaches; in table 2, the performance obtained using different feature selection algorithms are reported using a SVM with radial basis function kernel as classification technique. Results are reported both for the stand-alone approach and the ensemble obtained using SCS. The ensemble versions obtain very valuable results with a drastically dimensionality reduction. It is clear that the methods based on SCS greatly improve the performance with the further advantage of requiring a very low number of features.

The list of the 11 anatomical features obtained with the selection process by FS and MI and used in the second experiment to train different classifiers is reported in table 3. We can note that 8 features over 11 are the same.

A comparison with [19], where a univariate analysis of several anatomical features is reported, shows agreement in considering the total Left-hippocampus and Right-hippocampus volume as the best features to differentiate patients who will converted to AD from those who will not. Left-hippocampus and Right-hippocampus volumes and other volumetric measurements and cortical thickness measurements have also been used by the winner [18] of challenge CADDementia.

The second experiment is aimed at comparing the performance obtained using different classifiers. Moreover the results of some fusion approaches are evaluated:

- Fus8, sum rule among all the first 8 methods in table 4;
- Fus6, sum rule among the 6 methods in table 4 excluding SVM.

Both the ensembles Fus8 and Fus6 permit to boost the AUC but the accuracy is lower to that obtained by SVM coupled with FS and SCS. Anyway, Fus6 outperforms all the stand-alone classifiers that build it.

In table 5 the confusion matrix obtained by the method Fus6 is reported. Most errors are due to the confusion between HC and MCI classes. The AD class, instead, is quite easy to be recognized. In the final ranking of the international competition MLNeCh the method Fus6 presented in this paper is denoted as "Loris Nanni": the final "selected submission accuracy" is comparable with most of the approaches participating to the competition.

Conclusions

The aim of this work was to propose an ensemble of classifiers for the early diagnosis of AD, which participated to the international competition among AD predictors "A Machine learning neuroimaging challenge for automated diagnosis of Mild Cognitive Impairment".

The present study contains a detailed description of the method and the obtained results. We perform several experiments to investigate the best feature selection procedure and the best classifier for this problem and we propose a static classifier selection approach for design a multi-classifier which outperforms all the stand-alone method tested in this work.

The final ensemble is based on a set of classifiers, each trained on a different cluster of the training data. The proposed ensemble has the great advantage of performing well using a very reduced version of the data (a feature selection approach is applied which reduces the dimensionality of original data of more than 90%). Our selection procedures determined a very small set of just 14 features useful for classification including MMSE, left and right hippocampus volumes, left and right entorhinal thicknesses and volumes. A comparison with [19] where a univariate analysis of several anatomical features is reported, shows agreement in some features, in particular the left and Right-hippocampus volumes, as the best features to differentiate patients who will converted to AD from those who will not. A performance based comparison with other participants to the competition shows that the proposed approach works very similar to most of the other competitors (about 55% accuracy),

excepted for the winner which has performance higher than all others (>60% accuracy). The reported results, and above all, the final results of all the participants to the competition obtained on the private dataset are considerably lower than the state-of-the-art in AD classification (e.g. >75% accuracy in [11]). However a consideration is due: the dataset contains a high percentage of fake subjects, which may affect the learning of classifiers, moreover the set of features extracted from the MRI images are quite correlated to each other and many of them are discarded during the learning process. Maybe a different preprocessing and feature extraction from the MRI images could be useful to improve the performance.

References

[1] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup, Alzheimer's Dement. 7 (2011) 270–279. doi:10.1016/j.jalz.2011.03.008.

[2] R.C. Petersen, R.O. Roberts, D.S. Knopman, B.F. Boeve, Y.E. Geda, R.J. Ivnik, et al., Mild cognitive impairment: ten years later., Arch. Neurol. 66 (2009) 1447–55. doi:10.1001/archneurol.2009.266.

[3] R.A.A. Al-falluji, MRI based Techniques for Detection of Alzheimer: A Survey, Int. J. Comput. Appl. 159 (2017) 20–24. doi:10.5120/ijca2017912929.

[4] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack, C.H. Kawas, et al., The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement. 7 (2011) 263–269. doi:10.1016/j.jalz.2011.03.005.

[5] R.A. Sperling, P.S. Aisen, L.A. Beckett, D.A. Bennett, S. Craft, A.M. Fagan, et al., Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement. 7 (2011) 280–292. doi:10.1016/j.jalz.2011.03.003.

[6] C. Salvatore, P. Battista, I. Castiglioni, Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines., Curr. Alzheimer Res. 13 (2016) 509–533.

[7] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.O. Habert, et al., Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database, Neuroimage. 56 (2011) 766–781. doi:10.1016/j.neuroimage.2010.06.013.

[8] X. Ouyang, K. Chen, L. Yao, X. Wu, J. Zhang, K. Li, et al., Independent component analysis-based identification of covariance patterns of microstructural white matter damage in alzheimer's disease, PLoS One. 10 (2015). doi:10.1371/journal.pone.0119714.

[9] T.M. Schouten, M. Koini, F. De Vos, S. Seiler, J. Van Der Grond, A. Lechner, et al., Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease, NeuroImage Clin. 11 (2016) 46–51. doi:10.1016/j.nicl.2016.01.002.

[10] D.A. Wolk, S.R. Das, S.G. Mueller, M.W. Weiner, P.A. Yushkevich, Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's Disease, Neurobiol. Aging. 1 (2016) 1–10. doi:10.1016/j.neurobiolaging.2016.09.011.

[11] T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, D. Rueckert, A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease, IEEE Trans. Biomed. Eng. 64 (2017) 155–165. doi:10.1109/TBME.2016.2549363.

[12] T.M. Nir, J.E. Villalon-Reina, G. Prasad, N. Jahanshad, S.H. Joshi, A.W. Toga, et al., Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease, Neurobiol. Aging. 36 (2015) S132–S140. doi:10.1016/j.neurobiolaging.2014.05.037.

[13] F. Kruggel, F. Masaki, A. Solodkin, Analysis of longitudinal diffusion-weighted images in healthy and pathological aging: An ADNI study, J. Neurosci. Methods. 278 (2017) 101–115. doi:10.1016/j.jneumeth.2016.12.020.

[14] C. Salvatore, A. Cerasa, P. Battista, M.C. Gilardi, A. Quattrone, I. Castiglioni, Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach, Front. Neurosci. 9 (2015). doi:10.3389/fnins.2015.00307.

[15] I. Beheshti, H. Demirel, Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease, Comput. Biol. Med. 64 (2015) 208–216. doi:10.1016/j.compbiomed.2015.07.006.

[16] S. Ray, M. Britschgi, C. Herbert, Y. Takeda-Uchimura, A. Boxer, K. Blennow, et al., Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins, Nat. Med. 13 (2007) 1359–1362. doi:10.1038/nm1653.

[17] E.E. Bron, M. Smits, W.M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, et al., Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge, Neuroimage. 111 (2015) 562–579. doi:10.1016/j.neuroimage.2015.01.048.

[18] L. Sørensen, C. Igel, A. Pai, I. Balas, C. Anker, M. Lillholm, et al., Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry, NeuroImage Clin. 13 (2017) 470–482. doi:10.1016/j.nicl.2016.11.025.

[19] R. Vasta, A. Augimeri, A. Cerasa, S. Nigro, V. Gramigna, M. Nonnis, et al., Hippocampal subfield atrophies in converted and not-converted mild cognitive impairments patients by a Markov random fields algorithm, Curr. Alzheimer Res. 13 (2016) 566–574.

[20] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (2014) 16–28. doi:10.1016/j.compeleceng.2013.11.024.

[21] S. Sun, Q. Peng, A. Shakoor, M. Vannucci, B. Mallick, A Kernel-Based Multivariate Feature Selection Method for Microarray Data Classification, PLoS One. 9 (2014) e102541. doi:10.1371/journal.pone.0102541.

[22] R.O. Duda, P.E. Hart, D.G. Stork, Pattern classification, John Wiley & Sons, 2012.

[23] S. Sun, Q. Peng, X. Zhang, Global feature selection from microarray data using Lagrange multipliers, Knowledge-Based Syst. 110 (2016) 267–274. doi:10.1016/j.knosys.2016.07.035.

[24] J. Pohjalainen, O. Rasanen, S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Comput. Speech Lang. 29 (2015) 145–171. doi:10.1016/j.csl.2013.11.004.

[25] C.C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Min. Knowl. Discov. 2 (1998) 121–167. doi:10.1023/A:1009715923555.

[26] C.E. Rasmussen, C.K.I. Williams, Gaussian processes for machine learning., 2004. doi:10.1142/S0129065704001899.

[27] R.E. Schapire, A Short Introduction to Boosting, Society. 14 (2009) 771–780. doi:10.1.1.112.5912.

[28] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844. doi:10.1109/34.709601.

[29] G.Ü. Eibl, K. Pfeiffer, Multiclass boosting for weak classifiers, J. Mach. Learn. Res. 6 (2005) 189–210. http://jmlr.csail.mit.edu/papers/volume6/eibl05a/eibl05a.pdf.

[30] C.X. Zhang, J.S. Zhang, RotBoost: A technique for combining Rotation Forest and AdaBoost, Pattern Recognit. Lett. 29 (2008) 1524–1536. doi:10.1016/j.patrec.2008.03.006.

[31] L. Nanni, S. Brahnam, S. Ghidoni, A. Lumini, Toward a General-Purpose Heterogeneous Ensemble for Pattern Classification, Comput. Intell. Neurosci. 2015 (2015). doi:10.1155/2015/909123.

[32] Xiaofei He, Deng Cai, Shuicheng Yan, Hong-Jiang Zhang, Neighborhood preserving embedding, Tenth IEEE Int. Conf. Comput. Vis. Vol. 1. 2 (2005) 1208–1213 Vol. 2. doi:10.1109/ICCV.2005.167.

[33] L. Nanni, S. Brahnam, A. Lumini, Classifier Ensemble Methods, in: Wiley Encycl. Electr. Electron. Eng. Encycl. Electr. Electron. Eng., Wiley Online Library, 2015: pp. 1–12.

Figures and tables

	HC	AD	MCI	nMCI
Total	60	60	60	60
Female, n(%)	25(41.7)	31(51.7)	30(50)	32(53.3)
Male, n(%)	35(58.3)	29(48.3)	30(50)	28(53.3)
Age, years (mean±SD)	72.9±7.3	74.7±7.3	72.3±5.7	72.2±7.5
MMSE (mean±SD)	27.1±1.9	23.4±2.1	29.1±1.1	28.3±1.6

Table 1. Demographic characteristics of the study subjects (training set)

Ensemble mode	Feature selection	#Features	AUC	Accuracy
None	None	429	63.8	40.4
None	KPLS	100	70.3	37.9
None	FS	100	68.5	40.8
None	LM	100	72.3	44.6
None	MI	100	72.2	41.2
SCS	None	429	65.6	45.2
SCS	KPLS	11	76.5	54.5
SCS	FS	11	78.8	57.1
SCS	LM	11	74.0	52.1
SCS	MI	11	74.5	54.5

Table 2. Performance obtained by different methods based on SVM

Rank	Feature selected by FS	Feature selected by MI
1	**MMSE_bl**	**MMSE_bl**
2	**Left-Hippocampus**	**Right-Hippocampus**
3	**Right-Hippocampus**	**Left-Hippocampus**
4	**lh_parahippocampal_volume**	**Right-Amygdala**
5	**Right-Inf-Lat-Vent**	**Left-Inf-Lat-Vent**
6	**Left-Inf-Lat-Vent**	**Right-Inf-Lat-Vent**
7	rh_entorhinal_thickness	**lh_entorhinal_volume**
8	**lh_entorhinal_volume**	Left-Amygdala
9	lh_middletemporal_thickness	**lh_parahippocampal_volume**
10	**Right-Amygdala**	Right-Accumbens-area
11	lh_entorhinal_thickness	rh_entorhinal_volume

Table 3. Rank of the anatomical features selected by FS and MI. Features selected by both approaches are bolded.

Classifier	Ensemble mode	Feature selection	#Features	AUC	Accuracy
SVM	SCS	FS	11	78.8	**57.1**
GPC	SCS	FS	11	77.57	50.42
RS_AB	SCS	FS	11	74.23	52.08
RS_RB	SCS	FS	11	75.75	50.83
SVM	SCS	MI	11	74.5	54.5
GPC	SCS	MI	11	77.89	50.00
RS_AB	SCS	MI	11	79.44	49.58
RS_RB	SCS	MI	11	75.52	52.08
Fus8	Sum rule	FS&MI	14	79.30	52.92
Fus6	Sum rule	FS&MI	14	**79.60**	52.92

Table 4. Performance obtained by different classifiers.

		Predicted class			
		HC	MCI	AD	cMCI
Real class	HC	32	5	1	2
	MCI	20	6	9	5
	AD	0	1	39	0
	cMCI	10	8	11	11

Table 5. Confusion matrix of Fus6