Texture descriptors, voxels, and cognitive tests for the early diagnosis of Alzheimer’s disease

Loris Nannia, Sheryl Brahnamb, Christian Salvatore c, [[1]](#footnote-1) Isabella Castiglionic\*, and the Alzheimer’s Disease Neuroimaging Initiative +

a Department of Information Engineering, University of Padua, Via Gradenigo, 6/A, 35131 Padua, Italy

b Department of Management and Computer Information Systems, Glass Hall, Room 387, Missouri State University, Springfield, MO 65804, USA. E-mail: sbrahnam@missouristate.edu

c Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Via F.lli Cervi, 93, 20090 Segrate, Milano, Italy

Abstract

A method for performing early diagnosis of Alzheimer's Disease (AD) is proposed in this paper that combines separate SVMs trained on different texture descriptors extracted from slices of MRI that represent the Magnetic Resonance Image as well as separate SVMs trained on markers built from the voxels of MRIs. To reduce computation time, both texture features and voxel-based features are selected using different feature selection algorithms. To enhance performance, a combined biomarker that includes the age of each subject and several cognitive measures are used to train another SVM. These SVMs are then combined by weighted-sum rule for a final decision. Moreover, we have tested the performance of five different convolutional neural networks fine-tuned for MRI classification. The evaluation of our system on the ADNI dataset shows the efficacy of the proposed ensemble and demonstrates a significant contribution in accurate prediction of AD. The code for reproducing all experiments will be available at <https://www.dropbox.com/s/bguw035yrqz0pwp/ElencoCode.docx?dl=0>

Keywords: Alzheimer’s Disease, ensemble of classifiers, pattern recognition, feature selection.

1. Introduction

Today over forty-seven million people around the world are affected by Alzheimer’s Disease (AD) (Alzheimer's Disease International, 2015a, 2015b). An early and accurate diagnosis of AD is currently highly demanded, as it represents a way for treatments to effectively slow the progression of the disease. Research that investigates the symptomatic pre-dementia stage of AD most commonly referred to as Mild Cognitive Impairment (MCI) is essential if we are to develop better methods for predicting whether MCI will convert to AD (MCIc) or not (MCInc).

Although a definite diagnosis of AD can only be obtained through a post-mortem analysis, currently the clinical diagnosis relies mostly on a neuropsychological assessment of the patient, with the aim of ascertaining the presence of a cognitive impairment (G. McKhann et al., 1984). In a recent revision of the diagnostic criteria by the National Institute on Aging-Alzheimer’s Association workgroup (G. M. McKhann et al., 2011), new supportive features were considered for the diagnosis of AD, including neurogenetic testing, measurements of cerebrospinal fluid (CSF), amyloid and tau, and neuronal injury biomarkers as measured through neuroimaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). MRI and PET neuroimaging techniques are indeed able to provide measurements of atrophy and metabolism/amyloid markers, respectively. Changes in these features are detectable even before dementia is evident (Albert et al., 2011; Sperling, Aisen, et al., 2011).

MRI is less expensive than PET and is noninvasive and more widespread in both Western and non-Western regions. These features make MRI a suitable technique for the early detection of AD neuronal degeneration and for monitoring the progression of the disease in clinical trials (Sperling, Jack, et al., 2011).

Because of this, a considerable effort has recently been spent in the research for the implementation and development of advanced MRI processing techniques able to exploit the power of Machine Learning (ML) for enhancing diagnostic accuracy in the automatic early diagnosis of AD. Such systems may be capable of detecting pathologies through the automatic analysis of brain MRI volumes without apriori hypotheses on the location of the relevant information. This capability could not only move the detection of AD forward, but also improve our understanding of the disease.

One of the biggest issues in the classification of MRI for the automatic diagnosis of AD is the so-called curse-of-dimensionality problem, which is a consequence of the high dimensionality of MRI feature vectors and the low number of training patterns available in most MRI datasets relevant to AD. Therefore, the implementation of ML approaches able to select a subset of the original MRI feature set that are also powerful enough to achieve high classification performance will play a key role in the early detection of AD.

To date several texture approaches, such as the Gray-Level Co-occurrence Matrix (GLCM), Wavelet Transformation, the Statistical Approach, and the Local Binary Pattern (LBP), have been applied to the problem of AD diagnosis (Bustamam, Sarwinda, & Ardenaswari, 2018). In (Zhou, Liu, Zhou, & Xia, 2010) texture was found to be highly correlated with the hippocampal volume changes associated with AD. In (Zhang, Chunsui , & Lian, 2012) the GLCM approach using different sized Regions of Interest (ROI) in 3D MRI texture features succeeded in discriminating MRI images of patients with AD from images of normal patients. Zhang et al. found that GLCM and Run Length Matrix (RLM) could also be used to analyze the texture of the hippocampus. In  Simoes, Slump, and Marie (2012), a local statistical method based on a co-occurrence matrix texture map is introduced to diagnose the onset of AD by discriminating images of patients with MCI from Cognitively Normal (CN) subjects.

In this work, we propose a new classification system based on multi-domain sets of features. Separate SVMs are trained on these data and combined, and different feature-extraction and selection approaches are compared for training the SVMs that select different subsets of the whole set of the original data features.

Specifically, we test the following feature-extraction and feature-selection techniques on different MRI studies:

* Two different feature methods based on voxels (Nanni, Salvatore, Cerasa, & Castiglioni, 2016);
* A set of texture descriptors extracted from each slice of an MRI; for each descriptor, a different SVM is trained and the set of SVMs is then combined by sum rule. Since the descriptors are extracted from each slice of the MRI, a huge feature vector is finally obtained; a subset of the feature set is retained using different feature selectors before feeding into the SVMs;

The labels predicted by the classifier trained on MRI data are then combined with those predicted by an SVM trained on age and cognitive measures. These cognitive measures are Mini-mental state examination (MMSE), Clinical Dementia Rating-Sum of Boxes (CDR-SB), Rey's Auditory Verbal Learning Test (RAVLT), Functional activities questionnaire (FAQ), and Alzheimer's disease assessment scale-cognitive subtest (ADAS-cog).

1. Proposed system

The proposed system relies on the analysis and combination of different feature-selection methods and texture-descriptor approaches, each of which are briefly introduced below. SVM (Duda, Hart, & Stork, 2000) is the main classifier (tested using both histogram and the radial basis function kernel) and is implemented using LibSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/).

* 1. Feature selection

In the experimental section, we test the following feature-selection (FS) approaches:

* Fisher score (Fi) (Duda et al., 2000), an approach based on discriminative methods;
* Kernel PS (KPS) (Gutkin, Shamir, & Dror, 2009), an FS technique that discovers nonlinear correlation among the features by computing an approximation between a given matrix and a given vector of labels;
* Aggregate selection (AS) (Nguyen, Khosravi, Creighton, & Nahavandi, 2015), an FS approach that combines the feature ranking obtained using Fi (Duda et al., 2000), T-test (Duda et al., 2000), and the Sparse Multinomial Logistic Regression via Bayesian L1 Regularization (Cawley, Talbot, & Girolami, 2007).
  1. Texture Descriptors

The final feature vector is provided by concatenating the feature vector extracted from each slice of an MRI. The size of this feature vector is huge and needs to be further reduced via an FS algorithm. In the experimental section, we test the following FS approaches:

GABOR: Gabor filter (Fogel & Sagi, 1989) features are extracted from several different values (experimentally evaluated) for scale level and orientation. The best result obtained was with five different scale levels and fourteen different orientations. The mean-squared energy and the mean amplitude were calculated from each possible combination between scale and orientation. This method resulted in a feature vector of size 5×14×2.

WAVE: features are extracted from the horizontal, vertical and diagonal detail coefficients from wavelet decomposition (Meyer, 1992) at level 0 to 9. For each level we use the square root of the sum of all the horizontal, vertical, and diagonal coefficients as the feature. Three wavelet mothers are used: Haar (H); Daubechies 4 (DB); Coiflets 2 (CO). WAVE-x means that we extract the features using the 'x' wavelet mother.

GOLD: Gaussian Of Local Descriptors is a method that extracts a set of features proposed by Serra et al. (2015). This method is an improvement of Bag of Word (BoW) (Csurka, Dance, Fan, Willamowski, & Bray, 2004). The canonical BoW descriptor extracts local features that generate a codebook, and this codebook in turn encodes the local features into codes that form a global image representation. The codebook generation step is performed through clustering methods on the training set. GOLD generates the codebook via a flexible local feature representation obtained through a parametric probability density estimation that requires neither quantization nor a training set. The feature vector is fed into an SVM with a histogram kernel.

Specifically, the extraction of the feature vector can be described as a four-step process:

Step 1 *Extract Features*: dense SIFT descriptors are extracted on a regular grid of the input image;

Step 2 *Apply Spatial Pyramid Decomposition*: the image is decomposed into subregions by a multilevel recursive image decomposition; features are then softly assigned to regions according to a local weighting;

Step 3 *Estimate parametric probability density*: each region is represented as a multivariate Gaussian distribution of the extracted local descriptors by inferring local mean and covariance;

Step 4 *Project on the tangent Euclidean space*: After the covariance matrix is projected on the tangent space, it is concatenated to the mean to obtain the final region descriptor.

TERNARY CODING (TC) (Tan & Triggs, 2007): features are extracted from a variant of LBP that addresses some critical limitations of LBP (Ojala, Pietikainen, & Maeenpaa, 2002), especially its high sensitivity to noise in the near-uniform regions. Because TC offers a higher level of granularity, it is able to extract a greater number of textural features (Paci et al., 2013).

Since each digit of a LBP code is assigned to a 0 or a 1, codes range in [0, 2P-1]. The descriptors are the histograms of these binary numbers. TC extends s(x) such that the different *x* are encoded with three values instead of two using a threshold around zero.

Features are extracted using the Moore neighborhood (P=8,R=1) as well as (P=16,R=2), both with = 0.1 and by extracting normalized uniform bins.

CONVOLUTIONAL NEURAL NETWORKS (CNNs*):*CNNs the repeated concatenation of five classes of layers: convolutional (CONV), activation (ACT), pooling (POOL), fully-connected (FC), and classification (CLASS). We test and combine the following CNN architectures:

* AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) is the 2012 winner of the ImageNet ILSVRC challenge.
* GoogleNet (Szegedy et al., 2015) is the 2014 winner of the ImageNet ILSVRC challenge.
* VGGNet (Simonyan & Zisserman, 2014) placed second in ILSVRC 2014. The two best-performing VGG models (viz. VGG-16 and VGG-19), with 16 and 19 weight layers, respectively, are available as pretrained models.
* ResNet (He, Zhang, Ren, & Sun, 2016) is the winner of ILSVRC 2015. ResNet is a network that is about twenty times deeper than AlexNet and eight times deeper than VGGNet.

Each pre-trained CNN has been fine-tuned with the training set of the target problem. The training procedure sets the maximum number of epochs for training to 20, fixes the learning rate to 0.0001, and uses a mini-batch with 10 observations at each iteration.

From each MRI we extract the 100 central slides. Thus, each MRI is represented with 100 2D images, and each of these images is classified by CNN. The final classification score of a MRI is given by the sum rule of these 100 scores.

* 1. Global Grading Biomarker

The Global Grading Biomarker (GGB), tested in Tong et al. (2017), is a grading function that propagates disease levels of CN and AD, which, when combined, form the training population. The relationship of the training population to each MCI subject is modeled through a weighting function using a sparse representation method.

Since each MCI subject is assumed to lie in the space of the training population, it can be represented through a linear combination of CN and AD. After performing an FS process, *K* discriminative voxels and *K* intensity values are extracted from each image. Given , which contains the intensity values of *N* training images, and , which contains the intensity of an MCI image, a sparse representation of the MCI subject is obtained (for details, see (Tong et al., 2017)).

The nonzero coefficients indicate that the corresponding training image has been selected to propagate its clinical label information to the target MCI subject. Using the L1 norm in, one subject is selected, while the other is discarded; adding the L2 norm produces a grouping effect over the sparse coding coefficients so that both subjects can be used to calculate GGB.

The scoring of each MCI subject is based on the coding coefficients and the clinical status of the selected training population. The clinical status of a training image is denoted as . If the training image is for an NC subject, is set to 1; if for an AD subject, it is set to -1. A global grading value of the target MCI subject is then calculated by

where *N* is the number of training images and is the coding coefficient corresponding to the training image If GGB is close to -1, then this indicates that the MCI subject is likely to convert to AD within the given time period. If GGB is close to 1, then the MCI is less likely to convert to AD within the given time period.

* 1. Combined Biomarker

GGB is calculated for each MCI subject and then combined with the age of that subject and the results of six cognitive tests, resulting in a *combined biomarker*. As stated in Section 1, the cognitive measures used in the combined biomarker are MMSE, CDR-SB, RAVLT, FAQ, and ADAS-cog. These standard cognitive measurements, which are widely used in assessing cognitive and functional performance of dementia patients in clinical practice, are described in the ADNI General Procedures Manual (The Alzheimer’s disease neuroimaging initiative (ADNI)).

* 1. Voxels

The following preprocessing methods are applied to a given MRI image (for details, see (Salvatore et al., 2015) ): 1) image re-orientation; 2) cropping; 3) skull-stripping; 4) image normalization to MNI standard space (MNI152 T1 1mm brain template); and 5) tissue segmentation into Gray and White Matter tissue probability maps. The final size of the MRI volumes is 121x145x121 voxels. This large feature vector is reduced using an FS algorithm (see section 2.1) before training the SVM.

1. Experimental Section

The following four datasets belonging to two different brain MRI studies (both based on Alzheimer’s Disease Neuroimaging Initiative data repository) have been used to test the performance of our methods.

**Salvatore:** this set is composed of 509 subjects obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data repository (adni.loni.usc.edu). Data were collected from 41 radiology centers and include 137 AD, 76 MCIc, 134 MCInc, and 162 CN. The follow-up period to observe the conversion to AD was 18 months. For each patient, data from the screening or the baseline were considered. The final dataset is made up of T1-weighted structural MR images of the patients acquired at 1.5 T according to the standard ADNI acquisition protocol (Beheshti & Demirel, 2015; Jack et al., 2008). All images underwent a preprocessing step consisting in a 3D-gradwarp geometry correction for gradient nonlinearity and a B1 intensity correction for non-uniformity.

The following three classification tasks are performed: AD vs. CN, MCIc vs. CN, and MCIc vs. MCInc. The validation of the classifier is performed using a 20-fold cross-validation (CV). More information about the dataset and the splitting indexes can be found at <https://github.com/christiansalvatore/Salvatore-509>.

**Moradi:** this is the dataset used in Moradi et al. (2015) . Patients in the MCI group were classified as progressive MCI (164) if a conversion to AD during a three-year follow-up was observed. Patients were classified as stable MCI if the diagnosis was MCI at both baseline and 36-months follow-up. The validation of the classifier is performed using a 10-fold CV approach.

In Tables 1-2 we report the error under the ROC curve performance indicator (100-area under the ROC curve) obtained using the following methods:

* FULL: all the features (the combined biomarker) proposed in (Tong et al., 2017), see sections 2.3 and 2.4 above;
* NoCogn: SVM trained considering only two features used in (Tong et al., 2017), i.e., the biomarker and the age of the subject (no cognitive measures are used in this case, as they are very expensive to calculate).
* NoBiom: only age and cognitive measures are used as features (the complex (in terms of difficulty of implementation) biomarker proposed by Tong et al. (2017) is not considered).
* Feat(x): VOX approach (see section 2.5) based on the feature selection named x. To reduce the computation time, the feature selection named x is performed on the 10000 features selected by Fi.
* FeatFUS: fusion between SVM trained with Feat(KPS) and Feat(AS).
* CNN: It is the fusion by sum rule among the scores obtained by the five tuned convolutional neural networks reported in section 2.2.
* Random: random subspace method for managing the high dimensional feature vector, random subspace [5] of 50 SVMs each trained with a subspace built using 2000 features randomly extracted from the set of 10000 features selected by Fi.
* IMG: our image based approach proposed in this work. IMG is given by the weighted sum rule among (TC+GABOR+WAVE-H+WAVE-DB+WAVE-CO)+5×GOLD (the weight 5 used with GOLD means that it has the same importance in the fusion as (TC+GABOR+WAVE-H+WAVE-DB+WAVE-CO).
* IMG+CNN, the weighted sum rule among (TC+GABOR+WAVE-H+WAVE-DB+WAVE-CO)+5×GOLD+CNN.

Both the number of retained features and the parameters of SVM are selected with an internal 5-fold cross-validation on the training data (i.e., for each of the 20 folds of the CV, a further internal 5-fold CV is performed using only the training data). Note: the test set is always blind. Also, when we combine two methods (e.g. FeatFUS+IMG), the scores of each method (e.g. IMG and FeatFUS) are normalized to mean 0 and std 1 before the sum rule.

We want to stress that IMG coupled with FeatFUS obtains an interesting performance improvement in all the MRI case studies (see Table 2 as well). Clearly more studies should be performed on texture descriptors for improving the performance of IMG. In Table 2 we validate our approaches in other three datasets obtained from the ADNI data repository.

**Table 1. Proposed Ensembles**

|  |  |
| --- | --- |
| **FULL** | **11.64** |
| **FeatFUS** | **26.06** |
| **Feat(KPS)** | **32.19** |
| **Feat(AS)** | **26.41** |
| **IMG** | **27.62** |
| **GOLD** | **26.70** |
| **TC** | **43.40** |
| **CNN** | **36.70** |
| **IMG+CNN** | **28.45** |
| **Random** | **28.74** |
| **FeatFUS+IMG** | **24.84** |
| **FeatFUS+IMG+Random** | **24.76** |
| **FeatFUS+IMG+Random+6×FULL** | **10.66** |
| **NoCogn** | **15.19** |
| **FeatFUS+IMG+Random+6×NoCogn** | **13.80** |
| **NoBiom** | **16.61** |
| **FeatFUS+IMG+Random+6×NoBiom** | **13.29** |

The method proposed in (Tong et al., 2017) obtains very high performance. The fusion with our approaches permits only a slight performance improvement. In our opinion, the fusion of our ensemble with NoBiom or NoCogn produces more interesting result since the fusion is more useful.

**Table 2. Proposed Ensembles**

|  |  |  |  |
| --- | --- | --- | --- |
| Method | **AD vs CN** | **MCIc vs. CN** | **MCIc vs. MCInc** |
| Feat(KPS) | 6.7 | 9.2 | 36.5 |
| Feat(AS) | 7.0 | 10.4 | 35.4 |
| FeatFUS | 6.8 | 9.4 | 35.8 |
| Random | 7.2 | 9.3 | 34.0 |
| IMG | 7.4 | 12.4 | 32.7 |
| GOLD | 10.2 | 13.4 | 35.4 |
| TC | 17.8 | 20.0 | 41.8 |
| CNN | 11.5 | 21.8 | 37.7 |
| IMG+CNN | 7.3 | 12.7 | 37.4 |
| FeatFUS + Random | 6.7 | 9.7 | 35.1 |
| FeatFUS + IMG+ Random | **5.9** | **8.8** | 32.7 |
| (Nanni et al., 2016) | 6.7 | 11.1 | **31.4** |
| (Salvatore et al., 2015) | 24.0 | 28.0 | 34.0 |

It is interesting to note that in all the four datasets IMG permits to improve the performance of (FeatFUS+Random). IMG+CNN obtains performance similar to IMG, anyway it is clear that CNNs can be used in this classification problem, also if more tests (e.g. to choose the parameters tuning) should be performed to optimize them.

The new method proposed here obtains a performance that is slightly better than that reported in Nanni et al. (2016), where their approach compared favorably with respect to different state-of-the-art classification methods that were applied to the same dataset (ADNIset) used in this study (for the state-of-the-art see (Cuingnet et al., 2011).

Using the *Q*-statistic (Kuncheva, 2005), the error committed by FeatFUS+Random and IMG are (partially) independent. *Q* varies between −1 and 1. For statistically independent classifiers, *Qi*,*k* = 0. Classifiers that tend to recognize the same patterns correctly will have *Q* > 0, and those which commit errors on different patterns will have *Q* < 0. The Q-statistic between FeatFUS+Random and IMG is 0.786. Because of this partial independence, their fusion improves performance.

1. Conclusion

It is essential that methods be developed for early diagnosis of Alzheimer Disease (AD). Drugs now in trial show the greatest benefit to patients who have been diagnosed early; AD patients in a milder disease stage tend to respond better (Sevigny et al., 2016). It is hoped that even better results will be obtained if proper drugs are taken before the onset of the clinical symptoms.

In this paper, we propose robust ML methods for the early detection of AD. These methods combine MRI features with a so-called *combined biomarker*, which includes age and the results of cognitive assessments. To validate our approach, the system was tested using two different brain MRI studies (for a total of four datasets).

In future studies we will attempt to improve the performance of IMG by testing more descriptors (i.e., we will perform more tests using CNNs).

To reproduce our experiments, our code will be available at <https://www.dropbox.com/s/bguw035yrqz0pwp/ElencoCode.docx?dl=0>

Acknowledgments

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**

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment 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 & Dementia, 7*(3), 270-279.

Alzheimer's Disease International. (2015a). *Dementia statistics: Numbers of people with dementia*. Retrieved from <https://www.alz.co.uk/research/statistics>:

Alzheimer's Disease International. (2015b). *The global impact of dementia: An analysis of prevalence, incidence, cost and trends*. Retrieved from <https://www.alz.co.uk/research/WorldAlzheimerReport2015-sheet.pdf>.

Beheshti, I., & Demirel, H. (2015). Probability distribution function-based classification of structural MRI for the detection of Alzheimer’s disease. *Computers in Biology and Medicine, 64*, 208-216.

Bustamam, A., Sarwinda, D., & Ardenaswari, G. (2018). Texture and gene expression analysis of the mri brain in detection of alzheimer’s disease. *Journal of Artificial Intelligence and Soft Computing Research, 8*(2), 111-120.

Cawley, G. C., Talbot, N. L., & Girolami, M. (2007). Sparse multinomial logistic regression via bayesian L1 regularisation. *Nips, 19*, 209-216.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., & Bray, C. (2004). *Visual categorization with bags of keypoints*. Paper presented at the ECCV International Workshop on Statistical Learning in Computer Vision. <http://doi.org/10.1234/12345678>

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M. O., . . . Colliot, O. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage, 56*(2), 766-781.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification* (2nd ed.). New York: Wiley.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics, 61*(2), 103-113. doi:10.1007/BF00204594

Gutkin, M., Shamir, R., & Dror, G. (2009). SlimPLS: a method for feature selection in gene expression-based disease classification. *PLoS ONE, 4*(7), e6416.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., . . . Weiner, M. W. (2008). The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging, 27*(4), 685–691.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 1097-1105). Red Hook, NY: Curran Associates, Inc.

Kuncheva, L. I. (2005). Diversity in multiple classifier systems. *Information Fusion, 6*(1), 3-4.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease Report of the NINCDS‐ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology, 34*(7), 939-939.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. J., Kawas, C. H., . . . Phelps, C. H. (2011). 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 & Dementia, 7*(3), 263-269.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, 398-412.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

Nanni, L., Salvatore, C., Cerasa, A., & Castiglioni, I. (2016). Combining multiple approaches for the early diagnosis of Alzheimer’s Disease. *Pattern Recognition Letters, 84*(December), 259-266.

Nguyen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015). A novel aggregate gene selection method for microarray data classification. *Pattern Recognition Letters, 60-61*(August), 16-23.

Ojala, T., Pietikainen, M., & Maeenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971-987.

Paci, M., Nanni, L., Lathi, A., Aalto-Setälä, K., Hyttinen, J., & Severi, S. (2013). Non-binary coding for texture descriptors in sub-cellular and stem cell image classification. *Current Bioinformatics, 8*(2), 208-219.

Salvatore, C., Cerasa, A., Battista, P., Gilardi, M. C., Quattrone, A., & Castiglioni, I. (2015). Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Frontiers in Neuroscience, 1*(9), 307.

Serra, G., Grana, C., Manfredi, M., & Cucchiara, R. (2015). Gold: Gaussians of local descriptors for image representation. *Computer Vision and Image Understanding, 134*(May), 22–32.

Sevigny, J., Chiao, P., Bussière, T., Weinreb, P. H., Williams, L., Maier, M., . . . Sandrock, A. (2016). The antibody aducanumab reduces Aβ plaques in Alzheimer’s disease. *Nature, 537*, 50. doi:10.1038/nature19323.

Simoes, R., Slump, C., Marie, A., 2012. Using local texture maps of brain MR images to detect Mild Cognitive Impairment, 21st International Conference on Pattern Recognition, Japan.

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. Retrieved from arXiv:1409.1556v6

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennettd, D. A., Craft, S., Fagan, A. M., . . . Phelp, C. H. (2011). 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 & Dementia, 7*(3), 280-292.

Sperling, R. A., Jack, C. R. J., Black, S. E., Frosch, M. P., Greenberg, S. M., Hyman, B. T., . . . Schindler, R. J. (2011). Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: Recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimer's & Dementia, 7*(4), 367-385.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Tan, X., & Triggs, B. (2007). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Analysis and Modelling of Faces and Gestures, LNCS 4778*, 168-182.

The Alzheimer’s disease neuroimaging initiative (ADNI). *ADNI2: Defining Alzheimer's disease procedures manual*. Retrieved from <https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf>:

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2017). A novel grading biomarker for the prediction of conversion from mild cognitive impairment to alzheimer's disease *IEEE Transactions on Biomedical Engineering, 64*(1), 155-165.

Zhang, J., Chunsui , J., Y., & Lian, G. (2012). 3D texture analysis on MRI images of Alzheimer’s disease. *Brain Imaging and Behavior, 6*, 61-69.

Zhou, X., Liu, Z., Zhou, Z., & Xia, H. (2010). *Study on texture characteristics of hippocampus in mr images of patients with alzheimer’s disease*. Paper presented at the 3rd Annual Conference on Biomedical Engineering and Informatics, Yantai, Beijing.

1.  Tel.: +39-02-21717552; fax: +39-02-21717558; e-mail: [isabella.castiglioni@ibfm.cnr.it](mailto:isabella.castiglioni@ibfm.cnr.it).

   + The data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed both to the design and the implementation of ADNI and provided data but did not participate in the analysis or the writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how\_to\_apply/ADNI\_Acknowledgement\_List.pdf. [↑](#footnote-ref-1)