Artificial Intelligence In Medicine 97 (2019) 19–26

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Texture descriptors and voxels for the early diagnosis of Alzheimer's disease

Loris Nanni[a], Sheryl Brahnam[b], Christian Salvatore[c], Isabella Castiglioni[c,*], the Alzheimer's Disease Neuroimaging Initiative[1]

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

ARTICLE INFO

Keywords:
Alzheimer's disease
Ensemble of classifiers
Pattern recognition
Feature selection

ABSTRACT

Background and objective: Early and accurate diagnosis of Alzheimer's Disease (AD) is critical since early treatment effectively slows the progression of the disease thereby adding productive years to those afflicted by this disease. A major problem encountered in the classification of MRI for the automatic diagnosis of AD is the so-called curse-of-dimensionality, 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.

Methods: A method for performing early diagnosis of AD is proposed that combines a set of SVMs trained on different texture descriptors (which reduce dimensionality) extracted from slices of Magnetic Resonance Image (MRI) with a set of SVMs trained on markers built from the voxels of MRIs. The dimension of the voxel-based features is reduced by using different feature selection algorithms, each of which trains a separate SVM. These two sets of SVMs are then combined by weighted-sum rule for a final decision.

Results: Experimental results show that 2D texture descriptors improve the performance of state-of-the-art voxel-based methods. The evaluation of our system on the four ADNI datasets demonstrates the efficacy of the proposed ensemble and demonstrates a contribution to the accurate prediction of AD.

Conclusions: Ensembles of texture descriptors combine partially uncorrelated information with respect to standard approaches based on voxels, feature selection, and classification by SVM. In other words, the fusion of a system based on voxels and an ensemble of texture descriptors enhances the performance of voxel-based approaches.

1. Introduction

Today over forty-seven million people around the world are affected by Alzheimer's Disease (AD) [1,2]. Early and accurate diagnosis of AD is critical since early treatment effectively slows the progression of the disease adding productive years to those afflicted by it. Research that investigates the symptomatic predementia stage of AD, most commonly referred to as Mild Cognitive Impairment (MCI), is essential for developing better methods for predicting whether MCI will convert to AD (MCIc) or not (MCInc) and thus for initiating appropriate intervention programs.

Although a definite diagnosis of AD can only be obtained through a post-mortem analysis, clinical diagnosis today relies mostly on neuropsychological assessments of cognitive impairment [3]. In a recent revision of the diagnostic criteria by the National Institute on Aging-Alzheimer's Association workgroup [4], new supportive indicators are now recommended for consideration in the diagnosis of AD, including results on neurogenetic testing and measurement of cerebrospinal fluid (CSF), amyloid, 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 particularly good at providing measurements of atrophy and metabolism/amyloid markers, respectively. Changes in these features are detectable even before dementia is

* Corresponding author.
E-mail addresses: loris.nanni@unibo.it (L. Nanni), sbrahnam@missouristate.edu (S. Brahnam), christian.salvatore@ibfm.cnr.it (C. Salvatore), IsabellaCastiglioni@ibfm.cnr.it, isabella.castiglioni@ibfm.cnr.it (I. Castiglioni).

[1] 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.

https://doi.org/10.1016/j.artmed.2019.05.003
Received 11 April 2018; Received in revised form 10 May 2019; Accepted 16 May 2019

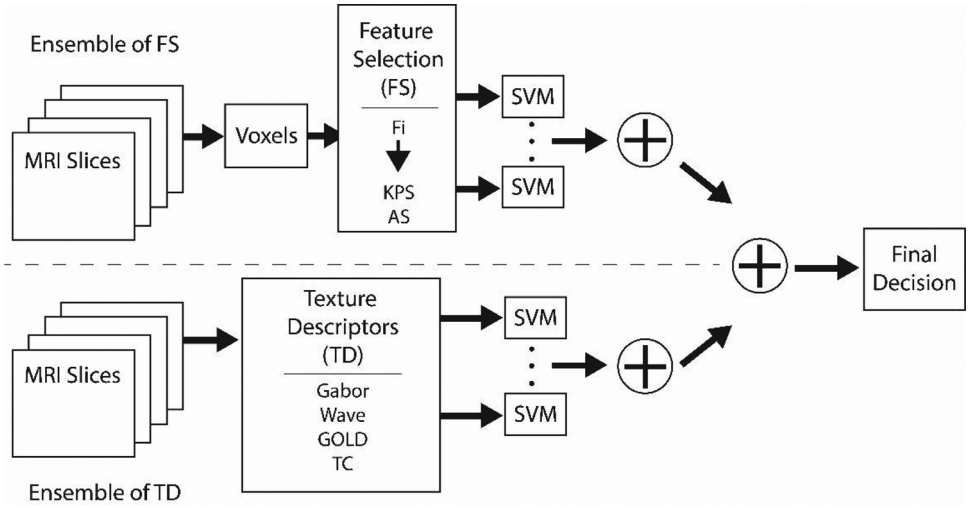

Fig. 1. Schematic of the proposed system.

evident [5,6].

Because MRI is less expensive than PET and is noninvasive, it is 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 [7]. Consequentially, considerable research effort has recently been focused on the development of advanced MRI processing techniques, especially techniques that exploit the power of Machine Learning (ML) for enhancing diagnostic accuracy in the early detection of AD. The goal is to produce ML systems that can detect pathologies through the automatic analysis of brain MRI volumes without apriori hypotheses on the location of relevant information. ML research in this area would thus not only move the detection of AD forward but would also improve our understanding of the disease.

An issue often encountered in the classification of MRI for the automatic diagnosis of AD is the 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. To avoid the curse-of-dimensionality problem, ML approaches must employ methods for selecting a subset of the original MRI features that are powerful enough to achieve high classification performance in the detection of AD.

A key area of research investigates the extraction of relevant features from the textural information in MRI images. Some findings support the hypothesis that the accumulated effect of neurofibrillary tangles and Ab plaques on MRI image produce specific textural patterns in MRI images [8], and texture analysis has successfully been used to identify surrogate biomarkers of AD from the medial temporal lobe studied by MRI [9]. Furthermore, the analysis of textures extracted from MRI images of the hippocampus was found to be more effective in the diagnosis and prognosis of AD than hippocampus atrophy [10].

To date, several texture approaches in ML, 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 [11]. In [12] 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 was introduced to diagnose the onset of AD by discriminating images of patients with MCI from Cognitively Normal (CN) subjects. An interesting emerging approach is the one proposed in the paper by Cevik et al. [13], in which the authors introduce a 3-step approach for identifying a subset of significant features and use a Multivariate Adaptive Regression Splines (MARS) method for classifying structural brain MRI of AD.

In this work, we propose a new classification system based on multi-domain sets of features. Separate SVMs are trained on each of these sets and combined. Different feature-extraction and selection approaches are compared for training the SVMs that select different subsets from the whole set of features. Due to computational issues with 3D descriptors, we focus in this study on 2D descriptors.

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

- Two different feature selection methods based on voxels [14];
- 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, an extremely large feature vector is finally obtained. A subset of the feature set is retained using different feature selectors before feeding the vectors into the SVMs;

An important finding in this paper is that ensembles of texture descriptors combine partially uncorrelated information with respect to standard approaches based on voxels, feature selection, and classification by SVM. To the best of our knowledge, this is one of the first papers that presents a study that combines texture descriptors and voxel-based features for the early diagnosis of AD. The most important finding of this paper is that such a combination improves the automatic-classification performance for the early diagnosis of AD. These results further suggest that texture descriptors and voxel intensities return complementary information with respect to the diagnosis of AD.

The remainder of this paper is organized as follows. In Section 2, we describe our approach, including the feature selection algorithms and texture descriptors compared and combined in the proposed system. In Section 3, we compare feature selection algorithms and texture descriptors and experimentally build a high performing ensemble. Finally, in Section 4 we present some conclusions.

2. Proposed system

As illustrated in Fig. 1, the proposed system is an ensemble composed of two other ensembles, one of which combines a set of SVMs (top) trained on markers built from the voxels of MRI (described in Section 2.1) and the other of which combines a set of SVMs (bottom) trained on different texture descriptors extracted from the MRI slices. The four texture descriptors (TD) used in this study are described in Section 2.2. To reduce computation time, voxel-based features are

selected using three different feature selection (FS) algorithms (addressed in Sections 2.1.1–2.1.3): Fisher score (Fi), Kernel Partial Least Squares (KPS), and Aggregate selection (AS). As illustrated in Fig. 1, KPS and AS use Fi features. Each feature selector and each texture descriptor are trained on separate SVMs. The two sets of SVM are combined by weighted sum rule to form their respective ensembles of FS and TD. These two ensembles are then combined by weighted-sum rule for a final decision. In the experimental section, we also compare and combine our approach to that developed by Tong et al. [15] (detailed in Section 2.3) that uses the Global Grading Biomarker (GGB).

SVM [16] is the main classifier used throughout the system (tested using both histogram and the radial basis function kernel) and is implemented using LibSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/).

2.1. Voxels + Feature selection (FS)

A well-established and validated preprocessing procedure [17] was applied to each MRI image before feature selection and classification. Preprocessing includes the following steps: 1) image reorientation; 2) cropping; 3) skull-stripping; 4) image normalization to MNI standard space (MNI152 T1 1 mm brain template); and 5) tissue segmentation into Gray Matter and White Matter tissue-probability maps. Specifically, image normalization aims at transforming all MR images from the original-image space to a standard-reference space defined by the selected template. With this operation, all MR images normalized to a defined standard space characterized by voxel-to-voxel correspondence, making inter-subject analyses possible at a voxel level. Tissue segmentation aims at returning brain spatial distribution of different tissues in a given MRI scan. Tissue segmentation is performed using tissue-probability maps that are available through the SPM software package available at The Wellcome Centre for Human Neuroimaging (https://www.fil.ion.ucl.ac.uk/spm/) and that are referenced to the MNI standard space. It must be noted that GM is known to be the most damaged tissue by the pathophysiological mechanisms of Alzheimer's Disease, and GM-related features have accordingly been shown to have the highest discrimination power in automatically classifying AD (see, e.g. [17]). All the steps of this preprocessing procedure were performed using the VBM8 software package [18]. The final size of the MRI volumes is 121 × 145 × 121 voxels. Raw single-voxel intensities were used as features for the subsequent analyses.

The resulting dimensionality of the feature vector extracted from the preprocessing procedure is huge. Data with high dimensionality is challenging due to the curse-of-dimensionality: in the presence of many irrelevant features, classifiers tend to become overfitted and less comprehensible. FS is one method that identifies relevant features for dimensionality reduction. In this work, the following FS approaches are tested: Fisher score (Fi), Kernel Partial Least Squares (KPS), and Aggregate selection (AS). Each of these FS methods is described below.

2.1.1. Fisher score (Fi)

Fi [16], or the *scoring algorithm*, is a widely used criterion for supervised feature selection. It can be described as follows: given a set $\mathscr{S}$ of d features, the aim is to choose a subset $\mathscr{T}$ of $m < d$ features that maximizes some criterion function $\mathscr{F}$:

$$\mathscr{T}^* = \arg \max_{\mathscr{T} \subseteq S} \mathscr{F}(\mathscr{T}), \quad \text{such that } |\mathscr{T}| = m, \tag{1}$$

where $|\cdot|$ is the cardinality of the feature subset. Because Eq. (1) is NP-hard, a common heuristic is to compute a score for each feature independently using some criterion function, after which the highest scoring m-ranked features are selected.

The evaluation criterion used in Fisher Score ($FS\mathscr{F}$) for a given feature f_i can be formulated as:

$$FS\mathscr{F}(f_i) = \frac{\sum_{j=1}^{c} n_j(\mu_{i,j} - \mu_i)^2}{\sum_{j=1}^{c} n_j \sigma_{i,j}^2}, \tag{2}$$

where μ_i is the mean of the feature f_i, n_j is the number of samples in the j -th class, and $\mu_{i,j}$ and $\sigma_{i,j}$ are the mean and the variance of the f_i on class j, respectively.

2.1.2. Kernel partial least squares (KPS)

KPS [19] is an FS technique that discovers the nonlinear correlation among the features by computing an approximation between a given matrix and a vector of labels. Partial Least Squares (PLS) is often described as a more powerful Principle Component Analysis (PCA) method [20] since the data are transformed into a different set of non-orthogonal basis vectors where only the most important PLS components are used to build a regression model. Whereas the new basis vectors in PCA become a set of successive orthogonal directions that explain the largest variance in the data, in PLS the basis vectors are a set of conjugant gradient vectors to the correlation matrix.

There are two main approaches for kernelizing PLS. One approach is based on the so-called *kernel trick* used in building SVMs. In fact, KPS is closely related to SVM [21] in that each point is mapped nonlinearly to a higher dimensional feature space, and a linear regression function is constructed in the mapped space that corresponds to a nonlinear function in the original input space. In the dual space, the mapped data appear as dot products, and these dot products are replaced by kernel functions in the final K-PLS algorithm. The second approach is to use PLS to factorize the kernel matrix directly. This method is based on a direct factorization of the kernel matrix. KPS is a nonlinear extension of PLS. Once the kernel matrix has been determined, only linear algebra is required [21].

In this paper we use the multivariate feature selection approach proposed in [22]. The KPS algorithm can be stated in terms of the dot products between pairs of inputs and the substitute kernel function $\mathbf{K}(\bullet, \bullet)$. If $\boldsymbol{X} \in \mathbb{R}^{N \times D}$ is the matrix of D-dimensional observed D variables with N the number of observations, and if $\boldsymbol{Y} \in \mathbb{R}^{N \times C}$ is the corresponding matrix of C-dimensional C classes, then we can map a nonlinear transformation $\Phi(\bullet)$ of the data into a higher-dimensional kernel space $\mathbb{K}$, such that $\Phi: x_I \varepsilon \mathbb{R}^D \rightarrow \Phi(x_I) \varepsilon \mathbb{K}$.

The first component for KPS can be determined as the eigenvector of the following square kernel matrix for β^Φ: $\beta^\Phi \lambda = \mathbf{K}_X \mathbf{K}_y \beta^\Phi$, where $\mathbf{K}_X$ is an element of the *Gram Matrix* $\mathbf{K}_X$ in the feature space, and λ is an eigenvalue. The size of the kernel matrix $\mathbf{K}_X \mathbf{K}_y$ is $N \times N$ regardless of the number of variables in the original matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$.

If $= \{t_1, t_2, ..., t_h\}$ is a set of components, with h the desired number of components, then the accumulation of variation explanation of T to $\boldsymbol{Y}$ can be written as:

$$w_i = \sqrt{D \frac{\sum_{l=1}^{h} \Psi(\mathrm{Y}, t_l) v_{il}^2}{\sum_{l=1}^{h} \Psi(\mathrm{Y}, t_l)}}, \quad i \in \{1, 2, ..., D\}, \tag{3}$$

where v_{il} is the weight of the i -th feature for the l -th component, $\Psi(\bullet, \bullet)$ is a correlation function, and $\Psi(\mathrm{y}_j, t_l)$ is the correlation between t_l and $\boldsymbol{Y}$. Larger values of w_i represent more explanatory power of the i -th feature to $\boldsymbol{Y}$.

In kernel space, KPS becomes an optimization problem:

$$\arg \max_{\alpha \subseteq \mathbb{R}^N} \left\{ \frac{\alpha^T \boldsymbol{S}_1^\Phi \alpha}{\alpha^T \boldsymbol{S}_2^\Phi \alpha} \right\}, \tag{4}$$

where α is an appropriate projection vector, and $\boldsymbol{S}_1^\Phi$ and $\boldsymbol{S}_2^\Phi$ are the inter-class scatter matrix and intra-class scatter matrix, respectively.

The calculation of the contribution of the l -th component γ_l can be calculated as:

$$\gamma_l = \frac{\sum_{i=1}^{C} N_i m_{il}^\Phi}{\sum_{i=1}^{C} N_i}, \tag{5}$$

where N_i is the number of samples in the i -th class, and m_{il}^{Φ} is the mean vector of the i -th class with respect to the l -th component in the projection space. The larger γ_l the more significant the classification.

In this paper the same parameters are used as in the MATLAB toolbox provided by the authors of [22]. The source code is available at https://github.com/sqsun/kernelPLS. The number of components, h, is set to 10 (in the source code, we also set the kernel parameters alpha to 1 and coef to 0.1).

2.1.3. Aggregate selection (AS)

AS [23] is an FS approach that combines the feature ranking obtained using Fi [16], the two-sample t-test [16], and the Sparse Multinomial Logistic Regression via Bayesian L1 Regularization [24]. The confidence of each of these criteria is different. Thus, if we take into account the ranking of all the above criteria, the resulting set of features should prove superior.

The two-sample t-test is a parametric hypothesis test that compares whether the average difference between two independent samples is significant or not and is expressed as:

$$t = (\mu_1 - \mu_2)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \tag{6}$$

where μ_1 and μ_2 are the means of the two samples, σ_1 and σ_2 are the respective standard deviations, and n_1 and n_2 are the sample sizes.

The Sparse Multinomial Logistic Regression via Bayesian L1 Regularization provides the standard penalized maximum likelihood solution to multi-class pattern recognition problems that includes sparsity as provided from a Laplace prior. With the Laplace prior, the regularisation parameters can be integrated out analytically. This removes the need for a lengthy cross-validation-based model selection stage. Sparse Multinomial Logistic Regression via Bayesian L1 Regularization is thus a fully automated process, having storage requirements that scale only linearly with the number of model parameters. Because the mathematics describing this algorithm are complex, the reader is referred to [24] for further details.

In practice, it is difficult to combine the ranking of all criteria because the ranges of statistics in the criteria are different, and a criterion that generates a higher range of statistics would dominate those with a lower range. To avoid this problem, AS uses a modified analytic hierarchy process (AHP) that assembles an elite set of features through a systematic hierarchy. This is accomplished by comparing the ranking features of a set of criteria initially by constructing a comparison matrix, whose elements are required to be transitive and consistent. Consistency of the comparison matrix is calculated using the Consistency Index (CI) and the Consistency Ratio (CR) based on large samples of random matrices. Let $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]^T$ be an eigenvector and λ be an eigenvalue of the square matrix $\mathbf{X}$:

$$\mathbf{X}\varepsilon = \lambda\varepsilon. \tag{7}$$

$$CI = \frac{\lambda_{\max} - n}{n - 1}. \tag{8}$$

$$\mathrm{CR} = \mathrm{CI}/\mathrm{RI}, \tag{9}$$

where RI is a random consistency index obtained from randomly generated reciprocal matrices, $\lambda_{max} = \max(\lambda_i), \quad \forall\, i \in [1, n]$ and $\lambda_i = 1/\varepsilon_i[x_{i1}, ...,x_{in}][\varepsilon_1,...,\varepsilon_n]^T$. If the set of judgments is consistent, the CR will not exceed 0.1 (in practice a few CR > 0.1 have to be accepted). If CR = 0, then the judgments are perfectly consistent.

After the comparison matrices are constructed, hierarchical analysis calculates the eigenvectors for each criterion that demonstrate the ranking scores. This becomes the performance matrix. The ranking of features is the multiplication of the performance matrix, and the vector representing the weight of importance for every criterion, i.e., the weight vector, is obtained by evaluating the level of importance each criterion is given regarding some goal. To avoid bias, the weight vector is typically 1 divided by the number of criteria.

2.2. Texture descriptors

Once the final feature vector is provided by concatenating the feature vector extracted from each slice of an MRI, different texture descriptors are extracted and trained with an SVM. Extracting texture features reduces the size of the input vectors. The SVMs are then fused by sum rule. The descriptors tested in this paper are described below.

2.2.1. GABOR

Gabor filter [25] features (GABOR) 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 $5 \times 14 \times 2$.

2.2.2. WAVE

WAVE features are extracted from the horizontal, vertical and diagonal detail coefficients from wavelet decomposition [26] 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), and Coiflets 2 (CO). WAVE- x means that we extract the features using the x wavelet mother.

2.2.3. Gaussian of Local Descriptors (GOLD)

GOLD, proposed by Serra et al. [27], is an improvement of the Bag of Word (BoW) [28]. The canonical BoW descriptor extracts local features that generate a codebook, and this codebook 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 GOLD feature vector is then fed into an SVM with a histogram kernel.

In brief, GOLD descriptors are obtained by extracting a set of feature descriptors $\boldsymbol{F} = \{F_1, F_2,..., F_N\}$, $F_i \in \Re^n$ from an image, collecting and weighting them in a spatial pyramid, and then describing each subregion by the estimated parameters of a multivariate Gaussian distribution. The covariance matrix is projected on a Euclidean space and concatenated to the mean vector to obtain the final descriptor of size $(n^2 + 3n)/2$. This feature vector is then fed into an SVM with a histogram kernel.

Below we describe this four-step process in greater detail:

Step 1. *Extract Features*: dense SIFT descriptors are extracted on a regular grid of the input image. Feature extraction is performed by calculating SIFT descriptors, using the function vl_phow from the vl_feat library 63 [29].

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. In order to take into account the spatial distribution of descriptors, the input image is divided into subregions defined by a multilevel recursive image decomposition: at level zero, the decomposition consists of the entire image, at level one, the image is subdivided into four quadrants, and so on. Instead of performing a hard-assignment of descriptors to regions, a soft assignment is performed according to a weighting strategy that assigns a weight to each of the descriptors on the basis of its distance from the region's center. Given a region, REG, centered in (c_x, c_y) with dimensions $REG_w \times REG_h$ and given a local feature descriptor $F \in \Re^n$ computed at (f_x, f_y), its weighting function is computed as:

$$w(F, REG) = \left(1 - \frac{f_x - c_x}{REG_w}\right) \bullet \left(1 - \frac{f_y - c_y}{REG_h}\right). \tag{10}$$

The function $w(F, REG)$ belongs to the range [0,1] and is designed so that the feature descriptors extracted from the center of the region have the maximum weight, while feature descriptors placed on the borders between two regions are equally considered for both regions. In this work one level of decomposition is used.

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. The set of weighted feature descriptors belonging to each region is then used to infer the parameters of a multivariate Gaussian distribution. If $\boldsymbol{F} = \{F_1, F_2 \ldots, F_N\}$ is the set of weighed local feature descriptors of a region, $\mu \in \Re^n$ the mean vector, and $\boldsymbol{C} \in \Re^{n \times n}$ the covariance matrix of a multivariate Gaussian distribution $\mathcal{N}$ belonging to a region, then:

$$\mathcal{N}(\mathbf{F}; \mu, \mathbf{C}) = \frac{1}{|2\mu\boldsymbol{C}|^2} e^{-0.5(\boldsymbol{F}-\mu)^T C^{-1}(\boldsymbol{F}-\mu)}, \tag{11}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{F}_i, \tag{12}$$

$$\boldsymbol{C} = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{F}_i - \mu)(\boldsymbol{F}_i - \mu)^T. \tag{13}$$

In this work we show that the covariance matrix $\boldsymbol{C}$ can be considered as an image, and standard texture descriptors can be used to describe it. Each of these feature vectors is fed into an SVM (with a radial basis function kernel) [21].

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. The parameters of μ and C provide a good representation of a region, but they are not feature vectors that are suitable for feeding a classifier (e.g., due to the dimensionality). In order to obtain a descriptor suitable for general purpose classifiers, the covariance matrix C is mapped into a point in the Euclidean space and concatenated to the mean μ so that the final region descriptor is a fixed length descriptor appropriate for linear classifiers based on the dot product.

Projection is performed in two steps: first, the covariance matrix $\boldsymbol{C}$ is projected on a Euclidean space tangent to the Riemannian manifold, at a specific tangency matrix $\boldsymbol{T}$; second, the orthonormal coordinates of the projected vector are extracted. In [27], the best choice for $\boldsymbol{T}$ was determined to be the identity matrix $\boldsymbol{I}$ since the neighborhood relation between the points in the new space remain unchanged wherever the projection point p is located. Therefore, the projection formula of the covariance matrix $\boldsymbol{C}$ into a vector point c simply applies the vector operator to the standard matrix logarithm thus:

$$c = vec(log(\boldsymbol{I}^{\frac{1}{2}}\boldsymbol{C}\boldsymbol{I}^{\frac{1}{2}})), \tag{14}$$

where *log* is the matrix logarithm operator and *vec* is the vector operator on the tangent space at identity, which for a symmetric matrix $\boldsymbol{C}$ is defined as $vec(\boldsymbol{M}) = [m_{1,1}, \sqrt{2}\ m_{1,2}, \sqrt{2}\ m_{1,3}, \ldots, m_{2,2}, \sqrt{2}\ m_{2,3}, \ldots, m_{n,n}]$.

2.2.4. Ternary coding (TC)

TC [30] features are extracted from a variant of LBP that addresses a critical limitation of LBP [31], namely its high sensitivity to noise in the near-uniform regions. Because TC offers a higher level of granularity, it is also able to extract a greater number of textural features [32].

LBP is a descriptor that has achieved great success due to its computational efficiency and discriminative power. The traditional LBP [31] is expressed as

$$LBP_{P,R} = \sum_{P=0}^{P-1} s(x)2^P, \tag{15}$$

where $x = q_p - q_c$ is the difference between the intensity levels of a central pixel (q_c) and a set of neighbouring pixels (q_p). A neighborhood is defined by a circular region of radius R and P neighbouring points. The function s(x) in Eq. (15) is defined as:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases}. \tag{16}$$

Since each digit of an LBP code is assigned to a 0 or a 1, LBP codes range in $[0, 2^{P-1}]$. LBP descriptors are the histograms of these binary numbers. The resulting patterns can be divided into two types: *uniform patterns*, which have at most two transitions from 0 to 1 or from 1 to 0, and *nonuniform patterns*, the label given for all other patterns.

TC extends $s(x)$ such that the different x are encoded with three values instead of two using a threshold τ around zero:

$$s(x) = \begin{cases} 1, & x \geq \tau \\ 0, & |x| \leq \tau \\ -1, & x \leq -\tau \end{cases}. \tag{17}$$

To compensate for the larger size of TC (for a given P-pixel neighborhood, there are 3^P possible codes with TC versus 2^P with LBP codes), the TC histograms are split into binary subhistograms that are then concatenated, with each TC code divided into a positive and negative binary pattern according to the sign of its components.

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

2.3. Global grading biomarker (GGB)

GGB, developed by Tong et al. [15], 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 $\boldsymbol{X}^{ADNC} \in \boldsymbol{R}^{K \times N}$, which contains the intensity values of N training images, and $\boldsymbol{X}^{MCI} \in \boldsymbol{R}^{K \times 1}$, which contains the intensity of an MCI image, a sparse representation of the MCI subject is obtained (for details, see [15]).

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 L_1 norm, one subject is selected, while the other is discarded; adding the L_2 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 $\hat{\alpha}$ and the clinical status of the selected training population. The clinical status of a training image is denoted as s_i. If the training image is for an NC subject, s_i is set to 1; if for an AD subject, s_i it is set to -1. A global grading value of the target MCI subject is then calculated by:

$$g^{MCI} = \sum_{i=1}^{N} \hat{\alpha}(i)s_i / \sum_{i=1}^{N} \hat{\alpha}(i), \tag{18}$$

where N is the number of $\boldsymbol{X}^{ADNC}$ training images and $\hat{\alpha}(i)$ is the coding coefficient corresponding to the training image $\boldsymbol{X}^{ADNC}(i)$. 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.

3. Experimental section

3.1. Datasets

The following four datasets belonging to two different brain MRI studies, both based on the ADNI data repository adni.loni.usc.edu, 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 forty-one 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 [33,34]. All images underwent a preprocessing step consisting of 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 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. 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.

3.2. Experimental results

In Tables 1 and 2 we report the error under the ROC curve performance indicator (100-area under the ROC curve), accuracy (AC), sensitivity (SE), and specificity (SP) obtained using the following methods:

- Tong: SVM trained considering only two features used in [15], i.e., the GGB and the age of the subject. Notice that in [15], as in other published studies (e.g [35,36].), the cognitive measures used for the clinical diagnosis of patients (i.e., the measures used to label patients as belonging to AD or MCI classes, e.g. MMSE) are also used for the training of the classifier at time points preceding the clinical diagnosis. This can produce an over performance in the automatic-classification task [37] if these time points are too close to the time of the clinical diagnosis. For this reason, we have not used the cognitive features, but only the two above cited features.
- Feat(x): the voxel approach (see Section 2.1) based on the feature selection named x. To reduce the computation time, the feature selection named x is performed on 10,000 features selected by Fi.
- FeatFUS: a fusion between SVM trained with Feat(KPS) and Feat (AS).
- Random: random subspace is a method for managing the high dimensional feature vector. We use a random subspace of 50 SVMs, each trained with a subspace built using 2000 features randomly extracted from the set of 10,000 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) and GOLD. All the methods are weighted by 1 except GOLD that is weighted by 5; the weight of 5 given to GOLD means that it has the same importance in the fusion as (TC + GABOR + WAVE-H + WAVE-DB + WAVE-CO).
- VoxIMG, sum rule among FeatFUS, IMG, and Random.
- FULL, weighted sum rule between VoxIMG (weight 1) and Tong (weight 2).

Table 1
Proposed Ensembles, Moradi dataset. Reporting error under the ROC curve (ROC), accuracy (AC), sensitivity (SE), and specificity (SP).

Method	ROC	AC	SE	SP
FeatFUS	26.1	68.8	80.9	53.4
Feat(KPS)	32.1	66.6	75.0	52.3
Feat(AS)	26.4	67.9	77.6	51.1
Random	28.7	62.1	55.9	72.7
IMG	27.6	65.0	82.2	42.1
GOLD	26.7	66.3	61.8	73.8
FeatFUS + IMG	24.8	67.1	72.4	58.0
VoxIMG	24.7	69.7	69.1	62.5
FULL	**13.8**	**78.8**	**78.9**	**77.4**
Tong	15.2	76.9	78.6	74.0

Best performance is highlighted.

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 standard deviation 1 before applying 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 the other three datasets obtained from the ADNI data repository.

The method proposed in [15] performs exceptionally well, but the fusion of that method with our approach improves performance even more.

It is interesting to note that in all four datasets IMG improves the performance of (FeatFUS + Random). The new method proposed here (VoxIMG) obtains a performance that is slightly better than that reported in Nanni et al. [14], 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 a list of the state-of-the-art, see [17] and [38]).

Using the Q-statistic [39], the error committed by FeatFUS + Random and IMG are (partially) independent, and 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 that 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.

In Fig. 2 we provide a comparison of ROC curves for AD vs. CN, MCI vs. CN, and MCIc vs. MCInc.

What is important to note here is that methods based on voxels extract information that is not captured completely by methods that extract texture directly from MRI slices; in other words, combining texture descriptors with voxel-based methods will improve performance.

3.3. Discussion

Our proposed method that combines texture descriptors and voxel-based features for the early diagnosis of AD shows classification performance in line with the state-of-the-art research [40–42], slightly improving the results obtained using other published feature-extraction and classification methods (see Tables 1 and 2). As reported in a recent review on ML techniques for the automatic diagnosis of AD [43], the mean percentage AUC for discriminating AD vs. CN was found to be 94 ± 4. For the discrimination of MCIc vs. MCInc, the mean percentage AUC was 70 ± 5. Other studies in the literature that investigated the use of ML techniques for the automatic classification of MCIc vs. MCInc report classification performances in the range 51–75%, lower than or comparable to our proposed method [14,17,35,37,44–51]. Cevik et al. [13], mentioned in the introduction and who use a Multivariate Adaptive Regression Splines (MARS) method for classifying structural brain MRI of AD, obtained a maximum average AUC of 87% for the classification of AD vs CN (84% sensitivity, 86% specificity), 70% for MCIc vs CN (78% sensitivity, 89% specificity) and 55% for MCIc vs MCInc (62 sensitivity, 60% specificity).

The main limitation of our work is that the proposed method was validated and tested only on binary comparisons of clinical interest

Table 2
Proposed Ensembles, Salvatore datasets. Reporting error under the ROC curve (ROC), accuracy (AC), sensitivity (SE), and specificity (SP).

Method	AD vs SN				MCIc vs. CN				MCIc vs. MCInc			
	ROC	AC	SE	SP	ROC	AC	SE	SP	ROC	AC	SE	SP
Feat(KPS)	6.7	86.9	85.4	90.1	9.2	84.0	91.9	63.1	36.5	66.1	28.9	88.8
Feat(AS)	7.0	87.3	83.2	90.1	10.4	82.7	93.8	64.4	35.4	66.1	30.2	86.5
FeatFUS	6.8	87.3	84.6	89.5	9.4	84.0	92.6	65.8	35.8	66.7	27.6	87.3
Random	7.2	87.2	83.2	90.7	9.3	85.1	94.4	68.4	34.0	66.6	35.5	85.8
IMG	7.4	84.6	81.8	87.0	12.4	79.0	77.2	82.9	32.7	59.5	65.8	56.0
GOLD	10.2	81.5	80.0	85.2	13.4	77.5	75.2	80.1	35.4	59.0	64.0	55.8
FeatFUS + Random	6.7	87.3	83.9	90.1	9.7	85.3	93.8	67.1	35.1	66.2	34.2	85.8
VoxIMG	**5.9**	**87.6**	84.1	90.3	**8.8**	**85.6**	94.0	68.5	32.7	67.1	34.5	86.5
[14]	6.7	87.0	83.3	89.5	11.1	84.0	92.6	64.4	**31.4**	**67.2**	35.5	86.0
[17]	24.0	—	—	—	28.0	—	—	—	34.0	—	—	—

Best performance is highlighted.

(i.e., AD vs. CN, MCIc vs. CN, and MCIc vs. MCInc). Recent studies are trying to address the problem of multigroup classification obtaining encouraging results [52,53]. It must be underlined that the proposed method can be translated to the discrimination of more than two diagnostic classes. However, further research is needed in this field.

In this paper, we compared the proposed method to a limited number of different automatic classification and feature-extraction techniques. However, new interesting techniques are emerging in the recent literature, including 1) studies on the minimization of uncertainty in prediction models (e.g. [54]), 2) studies addressing modeling problems even in different research fields (e.g [55–57].), and 3) deep-learning techniques, which are strongly emerging as potential automatic-discrimination algorithms given their high performance in different classification tasks (e.g [58].), but whose application to medicine (especially to neuroimaging) is still limited, mainly due to the high number of samples per diagnostic class needed to obtain stable predictive models. All these emerging approaches should be considered in future research work, and a comparison of the proposed method with a more inclusive set of techniques could help trace future research directions.

4. 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 [59]. 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 machine learning methods for the early detection of AD. Our system combines a set of SVMs trained on different texture descriptors extracted from slices of MRI with a set of SVMs trained on markers built from the voxels of MRIs. To reduce computation time, voxel-based features are selected using different feature selection algorithms, each of which trains an SVM. These two sets of SVMs are then combined by weighted-sum rule for a final decision. Our system was tested using two different brain MRI studies (for a total of four datasets extracted from ADNI dataset). Results demonstrate that an ensemble of texture descriptors extracts information that is partially uncorrelated to SVMs trained with a subset of the voxels. In other words, their fusion improves performance. We also show that the high-performance methods used in [15] are improved when combined with the ensemble proposed in this paper.

In future studies, we plan on investigating the performance of IMG using more descriptors (in particular, we will perform more tests using Convolutional Neural Networks. We will also evaluate using some powerful 3D texture descriptors).

For researchers interested in reproducing our experiments, our code will be available at https://www.dropbox.com/s/bguw035yrqz0pwp/ElencoCode.docx?dl=.

Conflict of interest

None.

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;

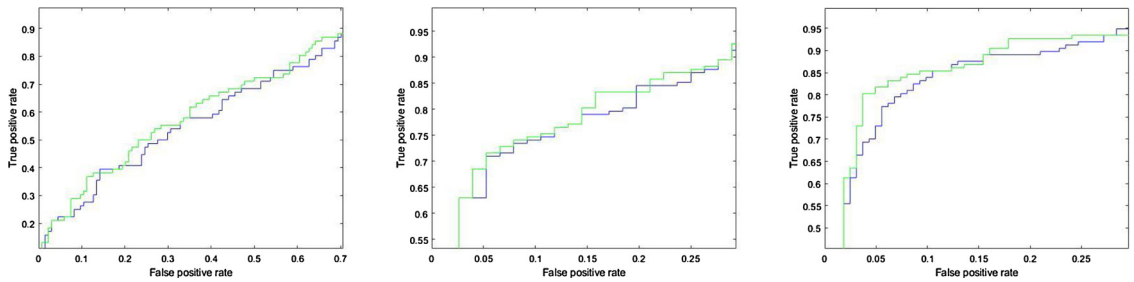

Fig. 2. Comparison of ROC curves of our new ensemble (green) and FeatFUS, the method based on voxels + feature selection + SVM (blue) for AD vs. CN, MCI vs. CN, and MCIc vs. MCInc (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

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 Disease International. Dementia statistics: numbers of people with dementia. 2015https://www.alz.co.uk/research/statistics.
[2] Alzheimer's Disease International. The global impact of dementia: an analysis of prevalence, incidence, cost and trends. 2015https://www.alz.co.uk/research/WorldAlzheimerReport2015-sheet.pdf.
[3] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984;34(7):939.
[4] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CRJ, Kawas CH, 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 Dement 2011;7(3):263–9.
[5] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. 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 Dement 2011;7(3):270–9.
[6] Sperling RA, Aisen PS, Beckett LA, Bennettd DA, Craft S, Fagan AM, 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 Dement 2011;7(3):280–92.
[7] Sperling RA, Jack CRJ, Black SE, Frosch MP, Greenberg SM, Hyman BT, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer's Association Research Roundtable Workgroup. Alzheimer Dement 2011;7(4):367–85.
[8] Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol 2004;59(12):1061–9.
[9] Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage 2011;58(2):469–80.
[10] Sørensen L, Igel C, Nielsen M. MCI trial enrichment using MRI hippocampus texture. Alzheimer Dement: J Alzheimer Assoc 2016;12(7):P108–9.
[11] Bustamam A, Sarwinda D, Ardenaswari G. Texture and gene expression analysis of the mri brain in detection of Alzheimer's disease. J Artif Intell Soft Comput Res 2018;8(2):111–20.
[12] Zhang J, Shunsui Y, Lian G. 3D texture analysis on MRI images of Alzheimer's disease. Brain Imaging Behav 2012;6:61–9.
[13] Çevik A, Weber GW, Eyüboglu BM, Oguz KK. Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI. Ann Oper Res 2017;251(1):31–57.
[14] Nanni L, Salvatore C, Cerasa A, Castiglioni I. Combining multiple approaches for the early diagnosis of Alzheimer's Disease. Pattern Recognit Lett 2016;84(December):259–66.
[15] Tong T, Gao Q, Guerrero R, Ledig C, Chen L, Rueckert D, et al. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. IEEE Trans Biomed Eng 2017;64(1):155–65.
[16] Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley; 2000.
[17] Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front Neurosci 2015;1(9):307.
[18] Ashburner J, Friston KJ. Voxel-based morphometry—the methods. NeuroImage 2000;11(6):805–21.
[19] Lindgren F, Geladi P, Wold S. The kernel algorithm for PLS. J Chemom 1993;7:45–9.
[20] Han L, Embrechts MJ, Szymanski B, Sternickel K, Ross A. Random forests feature selection with kernel partial least squares: detecting ischemia from magnetocardiograms. European symposium on artificial neural networks. Belgium: Burges; 2006. p. 221–6.
[21] Vapnik VN. The nature of statistical learning theory. New York: Springer-Verlag; 1995.
[22] Sun S, Peng Q, Shakoor A. A kernel-based multivariate feature selection method for microarray data classification. PLoS One 2014;9(7):e102541.
[23] Nguyen T, Khosravi A, Creighton D, Nahavandi S. A novel aggregate gene selection method for microarray data classification. Pattern Recognit Lett 2015;60–61(August):16–23.
[24] Cawley GC, Talbot NL, Girolami M. Sparse multinomial logistic regression via bayesian L1 regularisation. NIPS 2007;19:209–16.
[25] Fogel I, Sagi D. Gabor filters as texture discriminator. Biol Cybern 1989;61(2):103–13.
[26] Meyer Y. Wavelets and operators. Cambridge: Cambridge University Press; 1992.
[27] Serra G, Grana C, Manfredi M, Cucchiara R. Gold: Gaussians of local descriptors for image representation. Comput Vis Image Underst 2015;134(May):22–32.
[28] Csurka G, Dance CR, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. ECCV international workshop on statistical learning in computer vision. 2004. p. 59–74.
[29] Vedaldi A, Fulkerson B. VLFeat – an open and portable library of computer vision algorithms. Design 2010;3(1):1–4.
[30] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. Anal Modell Faces Gestures 2007;LNCS 4778:168–82.
[31] Ojala T, Pietikainen M, Maeenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24(7):971–87.
[32] Paci M, Nanni L, Lathi A, Aalto-Setälä K, Hyttinen J, Severi S. Non-binary coding for texture descriptors in sub-cellular and stem cell image classification. Curr Bioinform 2013;8(2):208–19.
[33] Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685–91.
[34] Beheshti I, Demirel H. Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease. Comput Biol Med 2015;64:208–16.
[35] Runtti H, Mattila J, Van Gils M, Koikkalainen J, Soininen H, Lötjönen J. Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort. J Alzheimer Dis 2014;39(1):49–61.
[36] Segovia F, Bastin C, Salmon E, Górriz JM, Ramírez J, Phillips C. Combining pet images and neuropsychological test data for automatic diagnosis of alzheimer's disease. PLoS One 2014;9(2):e88687.
[37] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, et al. Identification of conversion from mild cognitive impairment to alzheimer's disease using multivariate predictors. PLoS One 2011;6(7). https://doi.org/10.1371/journal.pone.0021896.
[38] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 2011;56(2):766–81.
[39] Kuncheva LI. Diversity in multiple classifier systems. Inf Fusion 2005;6(1):3–4.
[40] Vaithinathan K, Parthiban L. A novel texture extraction technique with T1 weighted MRI for the classification of Alzheimer's disease. J Neurosci Methods 2019;318:84–99.
[41] Khedher L, Illán IA, Górriz JM, Ramírez J, Brahim A, Meyer-Baese A. Independent component analysis-support vector machine-based computer-aided diagnosis system for Alzheimer's with visual support. Int J Neural Syst 2017;27(03):1650050.
[42] Salvatore C, Cerasa A, Castiglioni I. MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis. Front Aging Neurosci 2018;10(135).
[43] Salvatore C, Battista P, Castiglioni I. Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. Curr Alzheimer Res 2016;13(5):509–33.
[44] Koikkalainen J, Pölönen H, Mattila J, Van Gils M, Soininen H, Lötjönen J. Improved classification of Alzheimer's disease data via removal of nuisance variability. PLoS One 2014;7:e31112.
[45] Ye J, Farnum M, Yang E, Verbeeck R, Lobanov V, Raghavan N, et al. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC Neurol 2012;12(46).
[46] Casanova R, Hsu F-C, Sink KM, Rapp SR, Williamson JD, Resnick SM, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. PLoS One 2013:e77949. p. e8.
[47] Peters F, Villeneuve S, Belleville S. Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors. J Alzheimer Dis 2014;38:307–18.
[48] Dukart J, Sambataro F, Bertolino A. Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. J Alzheimer Dis 2015;49:1143–59.
[49] Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, Collins DL. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. Neurobiol Aging 2015;36:S23–31.
[50] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 2015;104:398–412.
[51] Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes J-D. Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. Alzheimers Dement (Amst, Neth) 2015;1(2):206–15.
[52] Nanni L, Lumini A, Zaffonato N. Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment. J Neurosci Methods 2018;302:42–6.
[53] Salvatore C, Castiglioni I. A wrapped multi-label classifier for the automatic diagnosis and prognosis of Alzheimer's disease. J Neurosci Methods 2018;302:58–65.
[54] Özmen A, Weber GW, Batmaz I, Kropat E. RCMARS: robustification of CMARS with different scenarios under polyhedral uncertainty set. Commun Nonlinear Sci Numer Simul 2011;16(12):4780–7.
[55] Weber G-W, Batmaz I, Köksal G, Taylan P, Yerlikaya-Özkurt FC. CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. Inverse Probl Sci Eng 2012;20(3):371–400.
[56] Savku E, Weber G-W. A stochastic maximum principle for a Markov regime-switching jump-diffusion model with delay and an application to finance. J Optim Theory Appl 2018;179(2):696–721.
[57] Onak ON, Serinagaoglu Dogrusoz Y, Weber GW. Effects of a priori parameter selection in minimum relative entropy method on inverse electrocardiography problem. Inverse Probl Sci Eng 2018;26(6):877–97.
[58] Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognitio. CoRR 2014. 1403.6382.
[59] Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 2016;537:50.