Journal of Biomedical Informatics 43 (2010) 307–320

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data

Evanthia E. Tripoliti [a], Dimitrios I. Fotiadis [b,*], Maria Argyropoulou [c], George Manis [a]

[a] Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece
[b] Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, P.O. Box 1186, GR45110 Ioannina, Greece
[c] Department of Radiology, Medical School, University of Ioannina, GR 45110 Ioannina, Greece

ARTICLE INFO

Article history:
Received 31 July 2009
Available online 31 October 2009

Keywords:
Alzheimer's disease
Classification
Functional magnetic resonance imaging
Random Forests
Support Vector Machines

ABSTRACT

The aim of this work is to present an automated method that assists in the diagnosis of Alzheimer's disease and also supports the monitoring of the progression of the disease. The method is based on features extracted from the data acquired during an fMRI experiment. It consists of six stages: (a) preprocessing of fMRI data, (b) modeling of fMRI voxel time series using a Generalized Linear Model, (c) feature extraction from the fMRI data, (d) feature selection, (e) classification using classical and improved variations of the Random Forests algorithm and Support Vector Machines, and (f) conversion of the trees, of the Random Forest, to rules which have physical meaning. The method is evaluated using a dataset of 41 subjects. The results of the proposed method indicate the validity of the method in the diagnosis (accuracy 94%) and monitoring of the Alzheimer's disease (accuracy 97% and 99%).

1. Introduction

Dementia is the progressive decline in cognitive function due to damage or disease of the brain, beyond what might be expected from normal aging. It is a non-specific illness syndrome in which affected areas of cognition may be memory, attention, language, and problem solving. Higher mental functions are affected first in the process of the disease. In the later stages, the affected person may be disoriented in time, place, and persons. They do not know who they are or cannot recognize other people around them [1,2]. The Alzheimer's disease (AD) is the most common form of dementia. More specifically, 10% of the subjects over 65 years old and nearly 50% of those over 85 suffer from AD. Alzheimer's disease is a neurodegenerative disorder caused by neuronal death due two misfolded proteins, β-amyloid and hyperphosphorylated-τ. The first one causes the creation of plaques and the second one causes the creation of neurofibrillary tangles. The development of plaques and tangles blocks the communication among nerve cells and disrupts processes necessary for cells to survive. The gradual loss of memory, motor and language skills, the cognitive impairment and the poor or decreased judgment are, simultaneously, some of the most severe symptoms and consequences of AD [1,2].

Up to now, there is no cure for Alzheimer's disease and no way to stop the underlying death of brain cells. Recently, pharmacological developments aim to slow the development of AD [3]. However, the administration of such drugs prerequisites the diagnosis of the disease. Today, the definite diagnosis is based on a time consuming procedure including interviews of patients and family members, physical exams, laboratory and cognitive tests and neurological exams. Neurological exams include studying of the brain anatomy, using magnetic resonance imaging (MRI) or computerized tomography (CT), and the brain function, using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) [4,5].

Functional magnetic resonance imaging (fMRI) is a non-invasive technique which allows indirect measurement of neuronal activity and imaging of activated cortical areas. The measurements are based on the fact that brain stimulation is correlated with an increased local brain metabolism. This metabolic activity causes local changes of the magnetic properties of blood, which can be imaged by fMRI. The activation of the brain due to a stimulus results into two antidromic processes: (i) the absolute amount of deoxyhemoglobin increases in the capillaries, and (ii) an overcompensation of the oxygen extraction takes place, due to an increase in regional cerebral blood flow (rCBF) and in regional cerebral blood volume (rCBV), resulting in an effective reduction of the deoxyhemoglobin concentration in the blood. Due to the paramagnetic properties of deoxyhemoglobin additional magnetic field gradients are created, resulting in larger local differences of the magnetic field. These differences in the local magnetic field are the source of the fMRI signal [6,7].

Nowadays, fMRI is becoming a diagnostic tool. Researchers working on AD have already used fMRI to detect changes in the brain function of patients suffering from the disease [8–24]. The results of these studies revealed a variety of differences between healthy and demented subjects concerning the intensity and extend of activation, the properties of the BOLD response, the metabolism, the functional synchrony and connectivity of brain regions. The majority of these studies are limited in the detection of those

* Corresponding author. Fax: +30 2651008890.
E-mail address: fotiadis@cc.uoi.gr (D.I. Fotiadis).

1532-0464/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2009.10.004

differences. Only five of them [20–24] quantified some of AD related changes to provide a preclinical marker that will support the diagnosis of the disease.

Supekar et al. [20] studied how the world architecture of the functional brain networks detected in task-free (resting state) functional magnetic resonance imaging is disrupted in Alzheimer's disease. They applied wavelet analysis to fMRI data in order to compute frequency-dependent correlation matrices. The analysis of the correlation matrices led to the production of undirected graphs which represent the functional brain networks. Metrics, such as characteristic path length and clustering coefficient were computed using graph analytical methods. Their study revealed that these metrics present statistically significant differences between AD patients and healthy subjects.

Greicius et al. [21,22] examined the activity of a default mode network including regions of posterior cingulated cortex, inferior parietal cortex, left inferolateral temporal cortex and ventral anterior cingulated cortex. Their study revealed that there is significant coactivation of the hippocampus in the default mode network, the network is abnormal in the mildest stages of AD (hypo metabolism of the regions which consist the network) compared to healthy aging and thus, network activity can distinguish AD subjects from healthy elderly controls. These three findings were the result of the application of independent component analysis (ICA) to mild AD and healthy subjects and the results of the utilization of a goodness-of-fit analysis in the individual subject level.

Li et al. [23] measured changes in functional synchrony in the hippocampus in AD, mild cognitive impairment (MCI) and healthy subjects. Functional synchrony was defined through the COSLOF index. The COSLOF index is the mean of the cross correlation coefficients of spontaneous low frequency components between pairs of voxel time series in a brain region. The values of this index were significantly lower in AD patients than healthy or MCI subjects. Also, the COSLOF index in MCI subjects was lower than those for the healthy controls. For the determination of differences in the COSLOF index between groups the two-tailed Student *t*-test was used.

Finally, Petrella et al. [24] tried to identify brain regions in which task-related changes in activation during a memory encoding task correlate with the degree of memory impairment across AD, MCI and elderly controls. For this purpose, a general linear model was used to assess the magnitude of fMRI signal intensity changes by using a contrast map. The contrast map represented the voxelwise difference in signal intensity magnitude between the two encoding conditions that were used (novel and familiar encoding conditions). The within and between subject analysis lead the authors to the conclusion that compared with activation in the medial temporal lobe (MTL), deactivation in the posteromedial cortices (PMCs) could be a more sensitive marker for early detection of AD using fMRI.

In this work a new approach is presented for the diagnosis of AD using fMRI. It consists of six stages (Fig. 1). First, the fMRI data are preprocessed to remove artefacts due to different sources of noise which corrupt the fMRI signal. Second, the fMRI data are modeled using a Generalized Linear Model. The main output of this step is the activation map that is the core for the extraction of the largest part of the features. Third, the features are extracted from structural and functional images and from demographic and behavioral data. According to the literature the specific features present statistically significant differences between AD and healthy subjects. Fourth, a feature selection algorithm is employed to remove redundant information. Fifth, supervised classification methods are applied. More precisely, the classical Random Forests classification algorithm and four variations of it are employed as well as Support Vector Machines classifiers. Finally, a set of rules are extracted from the trees of the forest and aim at assisting the diagnosis (two class problem) and monitoring of the progression of AD (three and four class problem).

The present approach has several advantages: (i) it employs voting schemes as improvements to Random Forests something that leads to better classification results; the employment of voting schemes involves fine tuning of methods parameters customized for the specific problem, (ii) it cross checks the validity of the method with the use of Support Vector Machines, in an attempt to decouple the problem from a specific classifier and show coherence of the results, (iii) the classification results are improved since in some cases the Support Vector Machines present better classification performance than that of Random Forests, and (iv) it extracts rules from the created forests giving a physical meaning to the produced Random Forest.

Furthermore, it presents a variety of characteristics which differentiate it from other methods already reported in the literature and address the diagnosis of AD: (a) it is independent from the type of fMRI experiment (block design, event related), (b) it is independent from the type of the task (resting state, sensory, motor, cognitive task), (c) it exploits all possibly extracted features that express AD related changes, (d) the features from different categories are fused to express as much medical knowledge as possible, (e) it supports not only the diagnosis of AD (two class problem) but the monitoring of progression of the disease too (three and four class problem), (f) it allows for whole brain analysis and not for region based only.

2. Materials and methods

2.1. fMRI experiment

2.1.1. Subjects

Raw structural and functional data from 41 right-handed English speaking individuals were received from the fMRI Data Center maintained at Dartmouth College (Hanover, NH) (http://www.fmridc.com). Fourteen healthy young subjects (5 male and 9 female), 13 elderly subjects (6 male and 7 female) with very mild to mild AD and 14 healthy elderly subjects (5 male and 9 female) were scanned during a simple sensory motor paradigm. The mean age of young participants was 21.1 years (range 18–24 years). The healthy elderly participants ranged from 66 to 89 years old (mean age 74.9 years), while AD subjects ranged from 68 to 83 years old (mean age 77.2 years) [15].

2.1.2. Dementia status characterization

Dementia status of the subjects was established using the Washington University Alzheimer's Disease Research Center (ADRC) recruitment and assessment procedures. Non-demented control subjects and those with mild dementia of Alzheimer type (DAT) were assessed clinically with the Clinical Dementia Rating (CDR): CDR 0 indicates no dementia; CDR 0.5 and CRR 1 indicate very mild and mild dementia of Alzheimer's type, respectively. The ADRC diagnostic criteria for DAT are comparable to those of "probable AD" as described by the work group of the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association [15,25].

2.1.3. Stimuli

The basic task paradigm consisted of the presentation of a 1.5 s duration visual stimulus. Participants pressed a key with their right index finger upon stimulus onset. The visual stimulus was an 8 Hz counterphase flickering (black to white) checkerboard subtending approximately 12° of visual angle (6° in each visual field). The stimulus onset was triggered at the beginning of the image acquisition via the PsyScope button box. Subjects complete four runs. Runs were structured such that for every eight-image acquisition one of two kinds of trial conditions was presented. Task trials

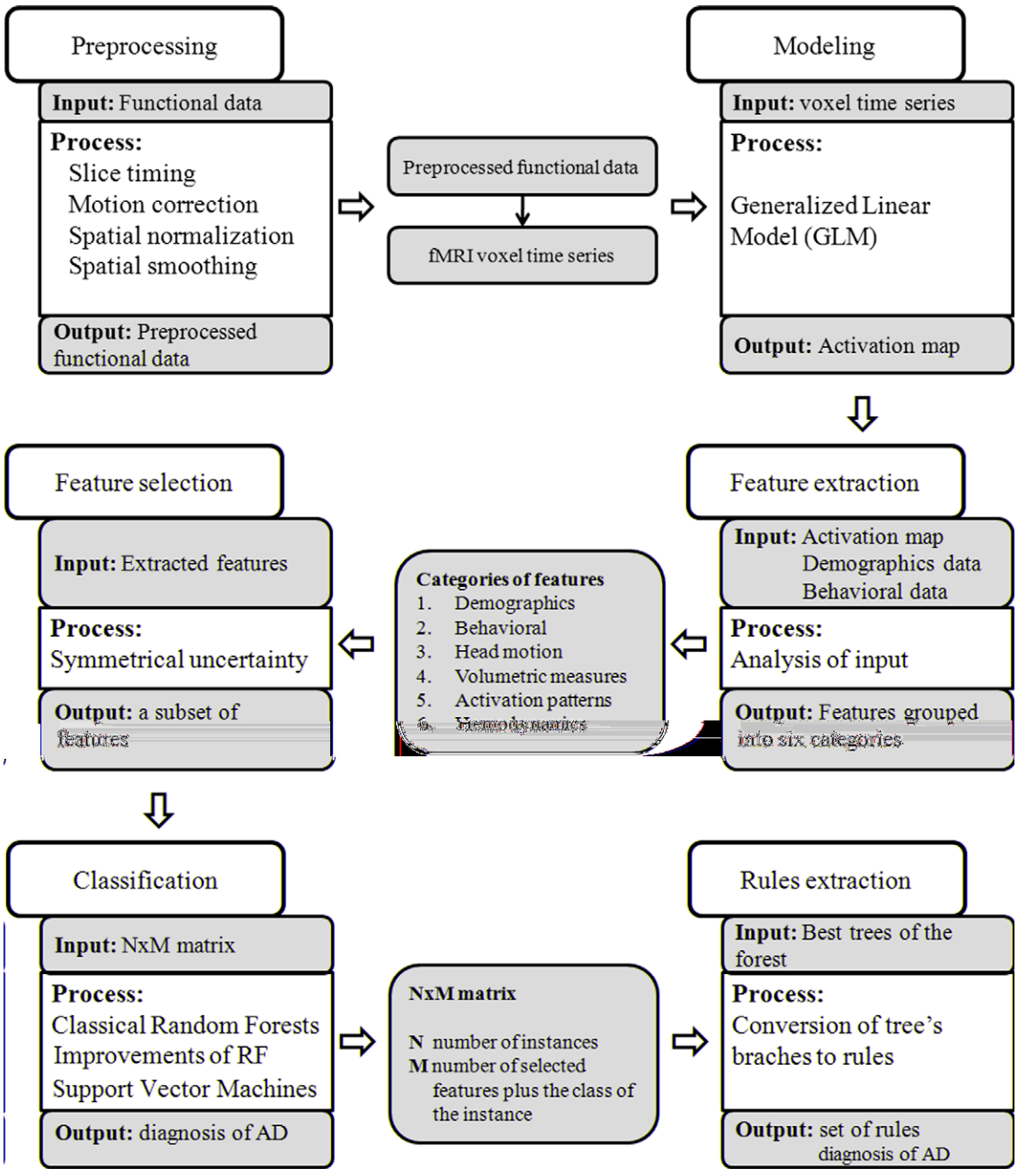

Fig. 1. A flowchart of the proposed method.

either involved stimuli presented in isolation (one trial condition) or in pairs (two trial conditions) with an inter-trial interval of 5.36 s. One trial and two trial conditions were pseudo randomly intermixed such that eight trials of one type and seven of the other appeared in each run. This results to four different functional runs. Each functional run lasted approximately 5.5 min and a 2 min delay existed between runs, during which subjects were permitted to rest [15].

2.1.4. Imaging protocol

Functional and structural images were acquired in a Siemens 1.5T Vision System with an asymmetric spin echo sequence sensitive to BOLD contrast. The protocol parameters are the following: TR = 2.68 s; 3.75 × 3.75 mm in plane resolution; $T2^*$ evolution time = 50 ms; $\alpha = 90^o$. Whole brain imaging was performed using 16 contiguous 8-mm thick axial oblique slices (acquired parallel to the plane of anterior–posterior commissures). High resolution structural images were also acquired in a series of three to four separate T1-weighted MP-RAGE (magnetization prepared rapid gradient echo sequence) anatomic images with the following parameters: 1 × 1 × 1.25 mm resolution; TR = 9.7 ms; flip angle = 10°; T1 = 20 ms; TD = 500 ms [15]. A detailed description of the scanner protocol for functional and structural images is given in Table 1.

2.2. The method

The proposed method consists of six stages. A flowchart of the method is shown in Fig. 1. A detailed description of these stages follows.

Table 1
Parameters of the imaging protocol.

Scanner protocol		
Protocol name	BOLD sensory motor	MP-RAGE structural
Coil type	Standard head	Standard head
Pulse sequence type	Asymmetric spin echo	MP-RAGE
Flip angle	90.0°	10.0°
TE	37.0 ms	4.0 ms
TR	2.68 ms	9.7 ms
No. of time points	128	1
No. of dummy acquisitions	0	0
No. of slices	16	128
Slice thickness	8.0 mm	1.25 mm
Slice skip	0 mm	0 mm
Sequence order	Interleaved	Interleaved
Field of view	240.0 mm	256.0 mm
Receiver bandwidth	Unknown	Unknown
Original acquisition matrix	64 × 64	256 × 256
Reconstructed image matrix	256 × 256	256 × 256
Full of partial *K*	Full	Full
Image orientation	Radiological	Radiological
Ramp sampling	Yes	Yes
Echo shift	50	50

2.2.1. Preprocessing

Functional MRI voxel time series include both signal of interest (evoked responses) and artefacts. The artefacts are due to the subjects' motion, the receiver coil, the amplifiers, the tissue pulsation and the fluctuations in blood oxygenation. Preprocessing aims at removing extraneous sources of variation and isolating the fMRI signal of interest in order activated regions of the brain to be detected. A common applied, by all the prior publications ([21,22]) concerning the specific dataset, preprocessing procedure is performed. More specifically, the preprocessing stage consists of four steps: (a) slice timing, (b) motion correction, (c) spatial normalization, and (d) spatial smoothing (Fig. 2). The preprocessing steps are either performed in the temporal or in the spatial domain.

Slice timing correction uses interpolation between the same slice and the voxel in neighboring acquisition repetition times (TRs) to estimate the signal that would have been obtained if the slices had been acquired at the beginning of each scan. The interpolation time point is typical chosen at TR/2 to minimize relative errors across each TR. This is done by a shift of the phase of the sines which make up the signal. The signal is convolved with a shifter filter to apply the phase shift and the correction is computed using sinc interpolation [7,26].

One of the most important artefacts in fMRI is the movement related one. The severity of movement related artefacts is due to: (a) addition of residual variance causing the activation to vary between adjacent voxels, and (b) movements which may be considered as activations if they correlate the stimuli. Since head movement cannot be eliminated by scanner environmental manipulation it can be corrected afterwards through mathematical transformations. More specifically, motion correction is achieved by a rigid body registration of the whole time series of the image to the image picked first. During the registration, six parameters (three translation and three rotations) are estimated. For the determination of the optimum parameter values a cost function is used, which expresses the mean squared difference between the two images [7,26].

Spatial normalization reduces the anatomical differences between different subjects and enables localization of different functional cortical areas and allows group comparisons. It is performed using a two step approach. The first step involves the determination of the optimum 12-parameter affine registration between the template and the object image. Unlike motion correction, where the images to be matched together are from the same subject, zooms and shears are also needed to register heads of different shapes and sizes. For the determination of the optimum affine transforma-

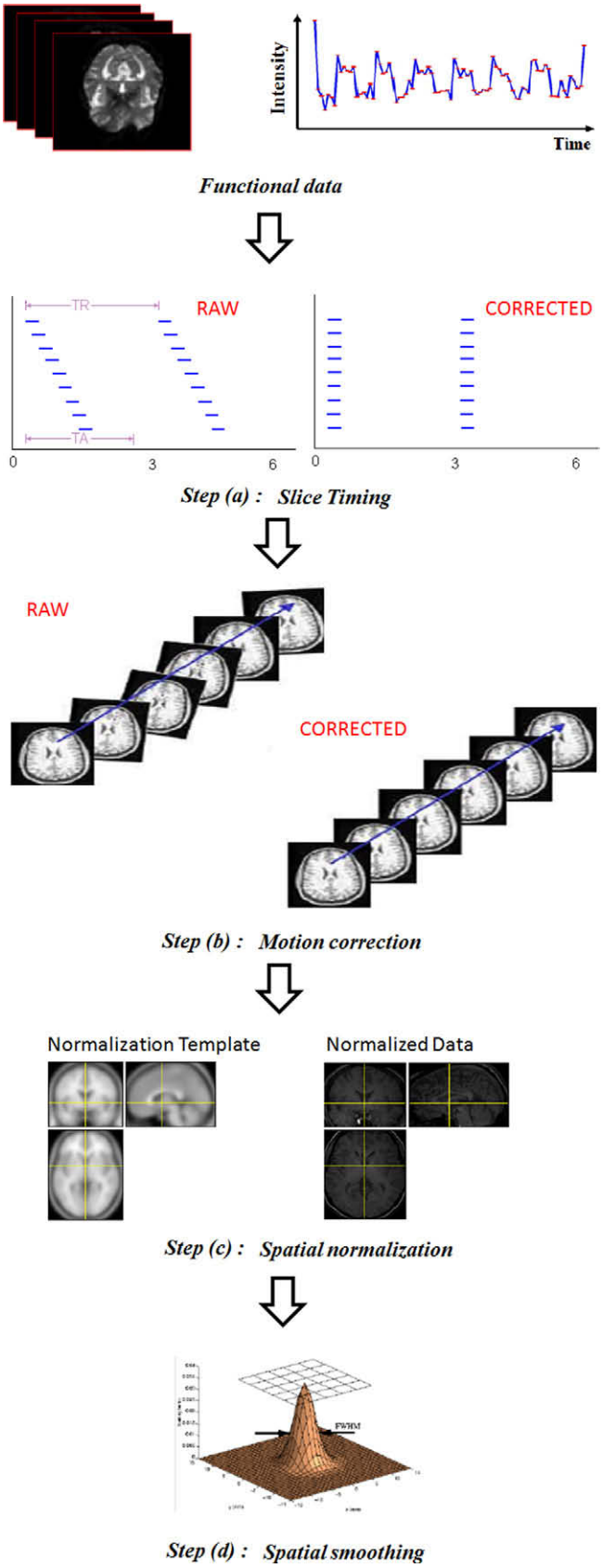

Fig. 2. Schematic representation of the preprocessing stage.

tion the maximum a posteriori (MAP) approach is used. The second step is the nonlinear registration to correct gross differences in head shapes that cannot be accounted for by the first step alone. To map a brain in one space to a brain in another, nonlinear distortions are needed. These distortions are modeled by a linear combination of basis functions, which in this case are the lowest frequency components of the discrete cosine transform [7,26].

Spatial smoothing is the last step of the preprocessing stage. It is performed to improve signal-to-noise ratio, to transform the data to better meet the assumptions of the statistics used in the next stage and to diminish the inter-individual differences in neuroanatomy. It is achieved by a discrete convolution using a Gaussian kernel. The width of the kernel lies between 3 mm and 10 mm of full width half maximum (FWHM) for fMRI. The recommended FWHM is 2–3 times the voxel size [7,26].

2.2.2. Modeling

Statistical analysis is carried out to determine which voxels are activated by the stimulation. For this purpose a univariate method is employed. It is based on the Generalized Linear Model (GLM) to describe the way in which the BOLD response depends on the stimulus [7]. The mathematical form of GLM is:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}. \tag{1}$$

According to this model the voxel time series $\boldsymbol{Y}$ ($T \times V$ matrix, where T is the time points and V is the number of voxels) consists of two parts: the fixed effects $\boldsymbol{X}\boldsymbol{\beta}$, and the random error $\boldsymbol{e}$ (noise). The fixed effects $\boldsymbol{X}\boldsymbol{\beta}$, is the part of the model that do not vary if the experiment is repeated. It is modeled by the product of the design matrix $\boldsymbol{X}$ ($T \times p$ matrix, where p is the number of predictors) and unknown predictor coefficients $\boldsymbol{\beta}$ ($p \times V$ matrix). The columns of the design matrix $\boldsymbol{X}$ reflect how the signal is varying in active areas and contain different types of covariates of interest. The random error $\boldsymbol{e}$ ($T \times V$ matrix) is the part of the model which explains how the observations vary even if the experiment is repeated on the same subject and under the same conditions. The noise has zero mean and covariance $\boldsymbol{V}\sigma^s$. The noise model indicates that the fMRI time series are not independent. Temporal correlations exist due to physiological effects and scanner instability. In order temporal correlations to be addressed Eq. (1) is multiplied by a matrix $\boldsymbol{A}$ which is equivalent to whitening the errors. Thus, the GLM takes the following form:

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{e}}, \quad \tilde{\boldsymbol{e}} \sim N(0,\ \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^T\sigma^2), \tag{2}$$

where $\tilde{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{Y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{A}\boldsymbol{X}$ and $\tilde{\boldsymbol{e}} = \boldsymbol{A}\boldsymbol{e}$ with $\tilde{\boldsymbol{e}} \sim N(0,\ \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^T\sigma^2)$.

The selection of matrix $\boldsymbol{A}$ is discussed in [27]. For the estimation of the parameters $\boldsymbol{\beta}$ of the model the least squares method was applied. The least squares estimator of $\boldsymbol{\beta}$ is given as:

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{X}}^{+}\tilde{\boldsymbol{Y}}, \tag{3}$$

where "+" denotes the Moore–Penrose pseudoinverse and $\tilde{\boldsymbol{X}}^{+} = (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T$.

The estimator of the variance σ^2 is based on the residuals, defined as the difference between the data $\tilde{\boldsymbol{Y}}$ and the estimated fixed effects $\tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}$. More specifically, the estimated variance is given as:

$$\hat{\sigma}^2 = \boldsymbol{r}^T\boldsymbol{r}/\mathrm{trace}(\boldsymbol{R}\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^T), \tag{4}$$

where $\boldsymbol{r} = \boldsymbol{R}\tilde{\boldsymbol{Y}}$, with $\boldsymbol{R} = (\boldsymbol{I} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}})$ and $\boldsymbol{I}$ is the identity matrix. For the estimation of noise variance we assume that the correlation structure $\boldsymbol{V}$ is known. However, the correlation structure is not known *a priori* and it must be estimated. The correct estimation of the correlation structure is crucial since it leads to the best estimator of $\boldsymbol{\beta}$ and the correct estimator of the variance of the estimator, to obtain the correct T or F statistic (in our case F statistic). For the estimation of V an autoregressive model (AR(1)) is used [28].

The estimated parameters of effects of interest $\hat{\boldsymbol{\beta}}$ are used to generate statistical parametric maps. The statistical parametric map is a 3D image, where each voxel has its own statistical value. The statistical value is calculated using the F-test [7]. For the determination of the activated voxels a cortical threshold is applied to the statistical parametric map. The threshold is chosen using the random field theory [7].

2.2.3. Feature extraction

The data which are acquired during the fMRI experiment (MR images, demographics, and behavioral data) and the output of the previous stages of the proposed method (activation maps and motion correction parameters) are the input of the feature extraction stage. The features are extracted from each of four functional runs for each patient and are grouped into the following categories: (1) demographics, (2) head motion, (3) behavioral, (4) volumetric measures, (5) activation patterns, and (6) hemodynamics. The features of the first and the third category are recorded during the conduction of the fMRI experiment. The features of the second category are computed using the parameters of the motion correction algorithm. The features of the fourth category are produced by the segmentation of MR images. The extraction of features of the last two categories is based on the activation maps. More specifically, the features which belong to each one of the above categories are:

2.2.3.1. Demographics. It includes only the age of the patient, since the most known risk factor for Alzheimer's is the increasing age. Most individuals with the illness are 65 and older. The likelihood of developing Alzheimer's approximately doubles every five years after age 65. After age 85, the risk reaches nearly 50% [1].

2.2.3.2. Head motion. The path length (PL) is a measure of head motion defined by D'Esposito et al. [29]. The path length is given as:

$$\mathrm{PL} = \sum_{i=1}^{120} \sqrt{(X_{t,i+1} - X_{t,i})^2 + (Y_{t,i+1} - Y_{t,i})^2 + (Z_{t,i+1} - Z_{t,i})^2}, \tag{5}$$

where i indexes the fMR images obtained for each subject and X_t, Y_t and Z_t are the translational parameter values in the x, y, and z direction, respectively. These parameters are the output of the motion correction algorithm.

2.2.3.3. Behavioral. The median and average reaction times of the subject, which are recorded for each functional run during the fMRI experiment, are utilized [15].

2.2.3.4. Volumetric measures. In Alzheimer's disease, nerve cell death and tissue loss cause areas of the brain to atrophy. Structural MRI allows the visualization of the subtle anatomic changes in the brain and the measurement of atrophy since it is expressed as a loss of gray matter [30]. Structural MR images, which are acquired during the fMRI experiment, are segmented into three clusters: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The segmentation method proposed by Ashburner and Friston [31] is applied. The portion of voxels belonging to the GM class, the mean and the standard deviation of the specific cluster are the features of the current category. In general, the atrophy can be measured in a specific region of the brain (e.g. hippocampus) using specific atrophy indexes proposed in the literature. The avoidance of those indexes and the selection of the features mentioned above make the proposed method to be a non region based approach.

2.2.3.5. Activation patterns. The features of this category are: the number of activated voxels, the value of primary peak (maximum z-score), the size of the cluster where the voxel with the maximum z-score (statistical significant voxel) belongs to, the number of sig-

nificant clusters (clusters of 50 voxels surrounding a primary peak), the percentage of activated regions which belong to a region of interest (regions of visual and motor cortex), and the total activation of these regions [10]. The computation of these features is based on the activation map that is the main output of the modeling stage. The cluster is defined as a contiguous group of activated voxels. The definition of clusters is based on faces and edges, but not corners, so each voxel has 18 neighbors.

2.2.3.6. Hemodynamics. The features of this category can be used to express the hypo perfusion and hypo metabolism. More specifically, the amplitude: of the BOLD response, of the rCBF, of the venous volume, of the vascular signal, of the deoxyHb signal, of the BOLD response undershoot and the transit time, are measured. Those measurements are based on the hemodynamic model proposed by Friston et al. [32]. They are computed for both the significant voxel and the cluster of activation.

All of the above features are extracted from each functional run of each patient, thus four vectors of features are the output of the specific stage. This results to a 164 × 27 (4 runs × 41 patients × 27 features) matrix containing the features from all runs of all subjects.

2.2.4. Feature selection

Feature selection aims at removing redundant information from the data. There are two common approaches: a wrapper approach which uses the intended learning algorithm to evaluate the usefulness of features and the filter approach which evaluates features according to heuristics based on general characteristics of the data [33]. In order the selection of the features to be independent from the learning algorithm and to provide an evaluation of the worth of the extracted features a filter approach is used. We need an independent from the learning algorithm feature selection algorithm since we use more than one classifier. More specifically, a correlation measure, called symmetrical uncertainty, is applied. Symmetrical uncertainty (SU) is based on the information theoretical concept of entropy and is given as:

$$SU = 2.0 \times \frac{H(A) - H(A|B)}{H(A) + H(B)}, \quad (6)$$

where $H(A)$ and $H(A|B)$ are the entropy of variable A before and after observing variable B. In other words, the amount by which the entropy of A decreases reflects the additional information about A provided by B. According to this approach in order to decide if a feature is important or not two issues must be addressed: (a) decide whether a feature is relevant to the class or not, and (b) decide whether such a relevant feature is redundant or not when considering it with other relevant features. A detailed description of all these issues can be found in [34].

Before the application of feature selection algorithm the features which belong to the volumetric measures category are removed in order the method to be comparable with other methods reported in the literature [20–23]. The selected features depend on the dataset and are described in [35].

2.2.5. Classification

2.2.5.1. Random Forests. The selected features are the input in the classification stage. Classification is based on the Random Forests (RF) classification algorithm, where majority and weighted voting schemes are employed. More specifically, classical RF and four variations of it are applied (Fig. 3). Classical RF (Fig. 4) is a classifier that consists of many decision trees. For the construction of each tree of the forest a new subset of samples is selected from the dataset. The tree is built to the maximum size without pruning. Only a subset m of the total set of features M is employed as the candidate splitters of the node of the tree. For each selected feature, the data are sorted by the values of the current feature and the Gini index is computed. The feature with the best value of the Gini index is used for splitting the tree node. The samples that were not selected for the construction of the tree constitute the test set of the tree and are called out-of-bag (OOB) samples. The error of the tree using these samples is called OOB error. The average of the OOB errors of all trees consists the generalization error of the forest. Once the forest is constructed a new sample runs through each tree in

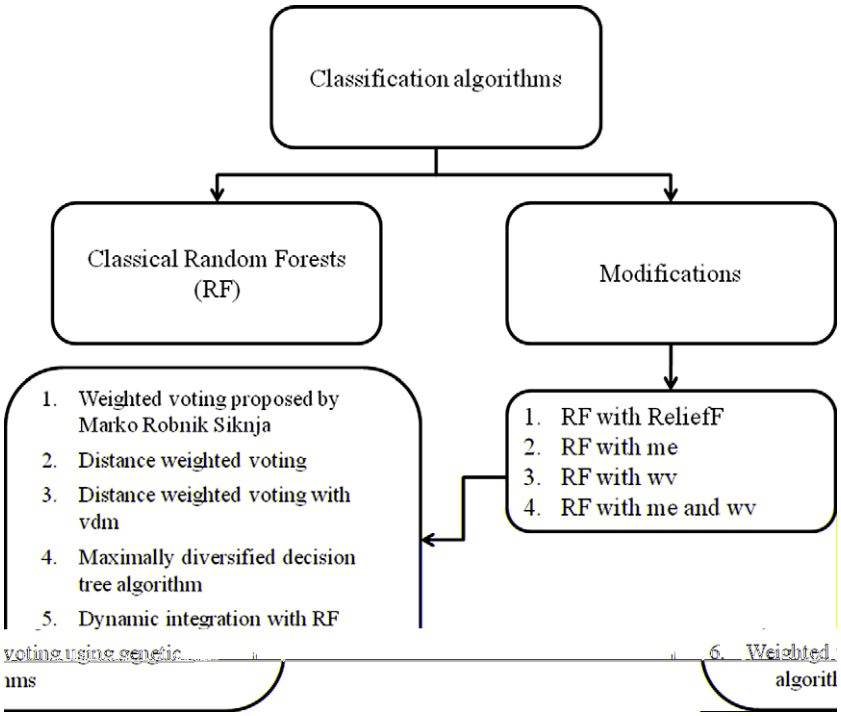

Fig. 3. Classification algorithms based on Random Forests.

2.2.5.2. Improvements of Random Forests. The performance of RF is affected by two factors: (a) the strength of each tree, and (b) the inter tree correlation. In order to strengthen each tree the ReliefF evaluation measure is used, instead of the Gini index (*RF with ReliefF*). The Gini index is fast but it cannot detect strong condition dependencies among attributes [38–40]. As far as it concerns the increase of the variety of the trees multiple estimators are used instead of one (*RF with me*). By using a large number of predictors, the predictor set will be quite different from tree to tree. These estimators are: the Gini index, the Gain ratio, the Minimum Description Length, the ReliefF and the myopic Relief [38–40]. The third improvement concerns the way of "voting". Since the contribution of each tree in the incorrect classification of the instances varies, weights are assigned to the votes of the trees (*RF with wv*). Six different weighted voting schemes are applied. These are: (1) weighted voting proposed by Marko Robnik-Sikonja, (2) distance weighted voting, (3) distance weighted voting using *vdm* distance, (4) maximally diversified multiple decision tree algorithm, (5) dynamic integration with RF, and (6) weighted voting using genetic algorithms. The description of weighted voting schemes is described in [41] and an algorithmic form of them is provided in Appendix A.

None of the above improvements address all the factors that affect the generalization error of the RF algorithm. The proposed approach combines weighted voting and the utilization of multiple estimators to improve the performance of RF (*RF with wv and me*) [42]. More specifically, the first and the third weighted voting schemes are combined with *RF with ReliefF*. The selection of the specific schemes is based on the fact that they provide the best results for the classification (see Section 3 below).

2.2.5.3. Support Vector Machines. In an attempt to decouple the problem from a specific classifier, Support Vector Machines are employed. Support Vector Machines (SVM) is a supervised learning method used for classification. It is based on the definition of an optimal hyperplane, which separates the training data so that the minimum expected risk, and at the same time, the maximum distance (geometric margin) of the data points from the correspond-

the forest and the tree "votes" the class that the sample belongs to. The predicted class is the one that gains most of the votes [36,37].

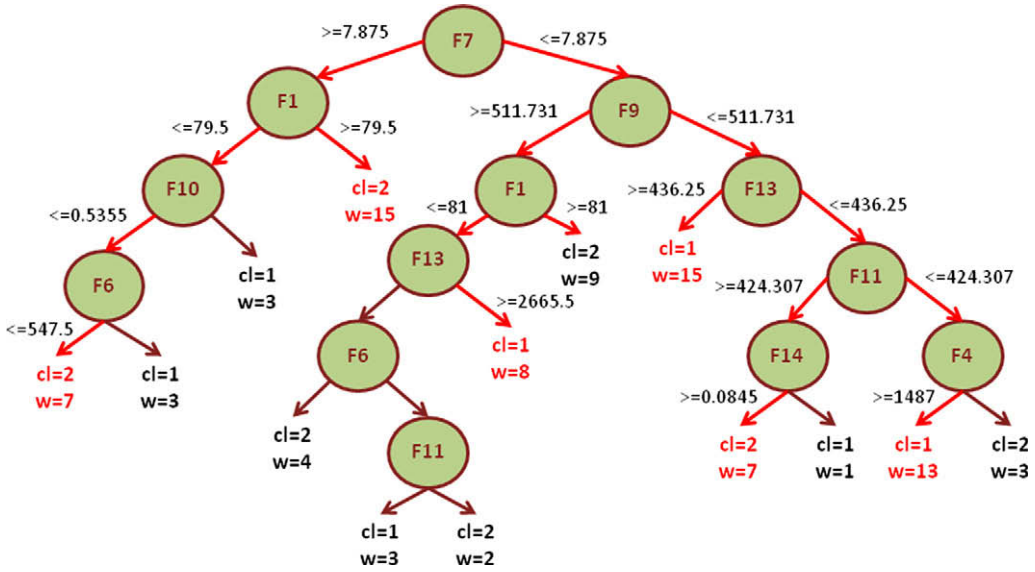

Fig. 5. One of the selected decision trees for the two class problem. Red arrows indicate the extracted rules.

2.2.6. Rule extraction

The forest created in the previous stage is transformed to rules. For each tree of the forest the classification accuracy is computed. Accuracy is defined as the ratio of the number of corrected classified samples to the total number of samples. The trees with accuracy greater than the median accuracy of all trees are selected. The branches of the selected trees, where the weight of the leaf node is larger than five, are converted to rules. The rules are created by pursing the tree from the root node to the specific leaf node. The weight of the leaf node is defined as the number of training samples which are used to classify an instance using the specific branch. This information is provided along with the class label in the leaf nodes. The rules extracted for all the selected trees. The created set of rules can be used to classify a subject as healthy or AD and to classify the stages of AD. An example of a tree and the corresponding rules that are produced is presented in Fig. 5.

In Fig. 5 each node of the tree contain a selected feature, the leaves represent classifications (cl = 1 corresponds to class 1 and cl = 2 corresponds to class 2) and the branches represent conjunctions of features that lead to those classifications. The red arrows indicate the extracted rules. For example one of these rules is the following:

If F7<=7.875 and F9<=511.731 and F13<=436.25 and F11>=424.307 and F14>=0.0845 then class = 2 (cl = 2).

The selected features, and thus the extracted rules, are differentiated depending on the classification problem. The extracted rules provide information concerning how the features belonging to different categories (demographics, behavioral, activation patterns, hemodynamics etc.) can be combined in order the diagnosis to be conducted and also provide a range of values for those features.

A sample set of rules extracted from Random Forests is shown in Table 2. The rules have physical meaning, according to the physicians, something that is also justified by the fact that the rules are composed by features expressing the neurophysiology and the pathology of Alzheimer's disease. It is important to underline that they provide a combination of the medical knowledge concerning the disease which allows to physicians to diagnose the disease and its severity with high specificity.

3. Results

The proposed method is evaluated using a dataset of 41 subjects. This dataset is transformed to an $N \times (M + c)$ matrix containing the M features (in our case 27) extracted from the data of the fMRI experiment and from each one of the four functional runs for each subject (totally $N = 4 \times 41 = 164$ samples). The last column, c, of the matrix contains the class where the sample belongs to. This matrix is the input of the method in the case of the four class problem (healthy young, healthy elderly, demented elderly with mild AD, demented elderly with mild AD subjects). For the two and three class problems the healthy young subjects are excluded. This results to 27 subjects, a number which is comparable

Table 2
A sample set of the extracted rules.

Rule	Decision	Features
IF N1>45 AND N9 > 0.161 AND N5 <=35 AND N13<=2586 AND N11<=1453 AND N6<=0.1315 THEN	Healthy old	• N1: age
IF 45 < N1<=70.5 AND N9<=0.3555 AND N3<=19324 AND N4 > 29 AND N12<=-0.008 THEN	Healthy old	• N2: path length
IF 45 < N1<=73 AND N13 > 212 AND N6<=0.1265 AND N9<=0.231 AND N11<=2702 THEN	Very mild AD	• N3: number of activated voxles
IF N2 > 7.82 AND N1<=82.5 AND N13<=463 AND N8<=0.043 AND N10<=0.857 THEN	Very mild AD	• N4: number of best clusters
IF N7<=0.2175 AND N6 > 0.0095 AND N5 < 34 AND N8<=0.0165 THEN	Mild AD	• N5: maximum z-score
IF 45 < N1<=69.5 AND N7<=0.2725 AND 6.8 < N2 < 9.6 AND 0.011 < N8<=0.034 THEN	Mild AD	• N6: venous volume
		• N7: rCBF
		• N8: deoxyHb
		• N9: vascular signal
		• N10: amplitude of BOLD response
		• N11: activation of significant clusters
		• N12: undershoot
		• N13: size of best cluster

Table 3
Sensitivity, specificity and accuracy of the proposed method for the diagnosis of AD for the five variations of the RF algorithm.

with that employed by other methods reported in the literature. Each patient undergoes four functional runs resulting into 108 instances. The difference between the two and three class problems is the last column of the matrix.

The results of the proposed method for the two, three and four class problem are reported in Table 3. The parameters which affect the performance of the method are: (a) the number of trees of the forest, (b) the number of features which are retained by the feature selection algorithm, and (c) the number of similar instances which are utilized in the weighted voting schemes for the classification of a new instance. Various combinations of those parameters are utilized to determine the one which produces the best results. The optimum parameter values for each one of the classification problems are reported in Table 4. The maximum number of trees that is used in all variations of Random Forests is 100, 95, 89 for the two, three and four class problem respectively, while the maximum number of features selected from the initial ones (27 features) is 14, 17 and 18 for the three classification problems, respectively. The utilization of the extracted rules, for the diagnosis of AD and the determination of the stage of the disease, produces the results reported in Table 5, Table 6, and Table 7 for the three classification problems, respectively.

A 10-fold stratified cross validation procedure is employed and the results are presented in Table 3. In this evaluation procedure each time a training set and a test set are selected which do not overlap. The results of the method iteration are averaged to produce a single estimation shown in Table 3. The results indicate that *RF with wv* can assist in the diagnosis of AD and the classification of its stages.

Table 7
Confusion matrix and evaluation measures using rules for the four class problem.

4 class problem		Healthy subject		Demented elderly subjects with very mild AD	Demented elderly subjects with mild AD
		Young	Elderly		
Healthy subjects	Young	56	0	0	0
	Elderly	0	56	0	0
Demented elderly subjects with very mild AD		0	1	27	0
Demented elderly subjects with mild AD		0	1	0	23
Evaluation measures					
Sensitivity $c = 1$ 100%		Sensitivity $c = 2$ 100%	Sensitivity $c = 3$ 96.43%	Sensitivity $c = 4$ 95.83%	Accuracy 98.78%
Specificity $c = 1$ 100%		Specificity $c = 2$ 98%	Specificity $c = 3$ 100%	Specificity $c = 4$ 100%	
PPV $c = 1$ 100%		PPV $c = 2$ 96.55%	PPV $c = 3$ 100%	PPV $c = 4$ 100%	

The combination of weighted voting with multiple estimators is evaluated only for the first and the third weighted voting schemes for two reasons: (a) they are the two schemes with the best results and (b) to test if the combination of the two variations affects the performance of the RF algorithm. Fig. 6 indicates that the combination improves the results of classification in the case of two and three class problem. However, in the case of the four class problem a slight decrease (1% and 6%) of the accuracy is observed. The comparison of the four classifiers (first weighted voting scheme, third weighted voting scheme, *me with wv-1* and *me with wv-3*) provided in Fig. 7 supports that the *me with wv-3* is the best classifier for the two and the three class problem, while the third weighted voting scheme gives the best result for the four class problem.

The contribution of the proposed method to the diagnosis of AD and to the classification of the stages of the disease is supported by the results produced by the application of the extracted rules. Although, a slight decrease, 4%, is observed in the accuracy of the two class problem, an important increase, 10%, in the accuracy of the three and four class problem is reached. The number of the extracted rules depends on two factors: (a) the number of the trees which are selected (trees where the classification accuracy is larger than the median accuracy of all trees), and (b) the structure of the tree (the number of branches where the weight of the leaf node is larger than five).

As already mentioned in the introduction, the validity of the method is cross checked by using the Support Vector Machines, as an alternative classification method. The results produced by the application of SVM in the classification stage are reported in Table 8. Comparing the results of SVM with those of the classical RF, the improved version of RF and the rules, we conclude that although the SVM outperforms the classical RF, the variation of RF, which produce the best results, and the rules are superior to the SVM in all classification problems. Table 9 presents a comparison of SVM models produced in multi category problems (see Section 2.2.5) and classical Random Forests algorithm. The results indicate that SVM and the Random Forests algorithm give comparable results with small deviations of order of 2%.

A comparison of the proposed method with those reported in the literature is shown in Table 10. The comparison is feasible only for the two class problem since there are no methods reported in the literature which classify the stages of AD. The results indicate that the proposed method provides better results than other meth-

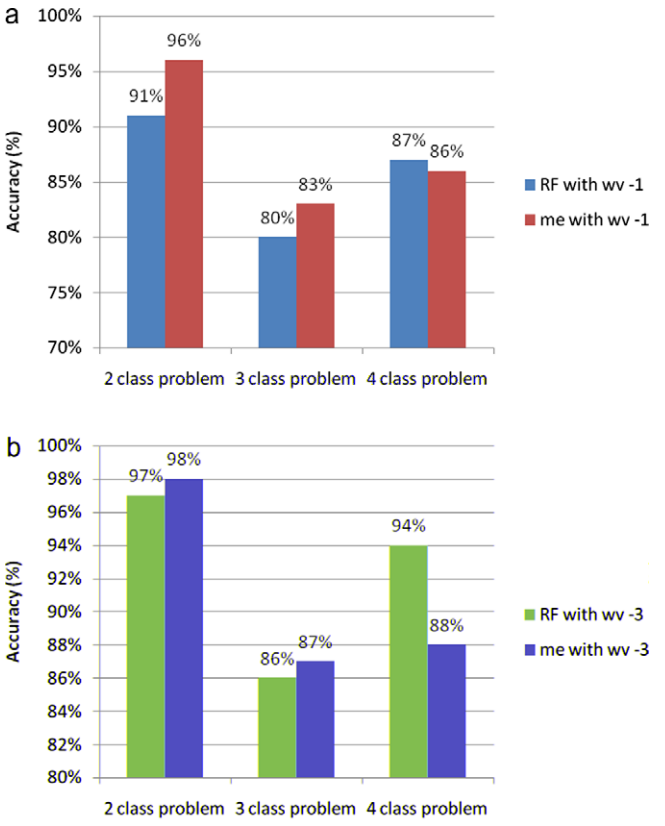

Fig. 6. (a) Comparison between the first weighted voting scheme and the me with wv-1, and (b) Comparison of the third weighted voting scheme and the me with wv-3.

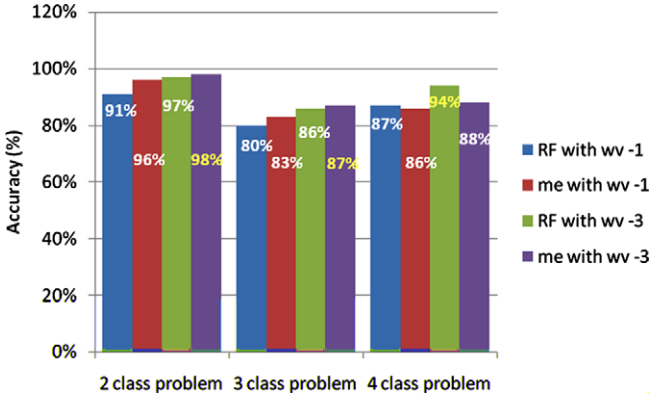

Fig. 7. Comparison between the first and the third version of the weighted voting schemes, me with wv-1 and me with wv-3.

Table 8
Accuracy of the SVM and other classification methods for the two, three and four class problem.

Problem	Method			
	SVM (%)	Classical RF (%)	RF with improvement (%)	Rules (%)
2 class	88	84	98	94
3 class	94	78	87	97
4 class	73	85	94	99

Table 9
Comparison of SVM models with classical RF.

Problem	Method	
	SVM (%)	Classical RF (%)
Healthy[1] versus demented[2] (SVM model—HD)	88	84
Very mild AD versus mild AD (SVM model—DD)	84	82
Healthy[3] versus demented (SVM model—HD)	91	93

[1] Healthy elderly subjects.
[2] Demented subjects with very mild and mild AD.
[3] Healthy young and elderly subjects.

ods, even if the evaluation is based on the same (Washington University Data) or different datasets.

4. Discussion

We propose a supervised method for the diagnosis and monitor the progression of the Alzheimer's disease. The method is based on features recorded during an fMRI experiment (imaging finding, behavioral and demographics data) and provides interpretation of the decision through the rules that are extracted. The method employs six stages: preprocessing, modeling, data modeling, feature extraction, feature selection, classification and rule extraction.

In the past, several methods have been reported in the literature utilizing the fMRI modality in order to extract new knowledge about the Alzheimer's disease. These methods can be grouped in two broad categories. The first category includes methods that study how a feature, which is extracted from fMRI data, is differentiated between demented and control groups. The second category includes methods that try to quantify Alzheimer's disease related changes in order to provide an index that can serve as a clinical marker for the diagnosis of the disease.

Although the proposed method can be categorized in the second group of methods has several features that differentiate it from other related methods of the two categories reported above. More specifically, the proposed approach not only assists in the diagnosis of Alzheimer's disease but also determines the stage of the progression of the disease. In order this to be achieved, it uses features that can be extracted from all the types of fMRI experiments (block design, event related) and all the types of stimulus or cognitive tasks and are not based on a certain brain region. The features that are employed belong to different categories such as demographics, behavioral data, head motion, activation patterns and hemodynamics and according to the literature, express as much medical knowledge as possible about AD. In addition, the proposed method adds new knowledge to the field since: (a) produces a set of rules which can be considered by the physician and help him in making a justifiable decision, and (b) adds new knowledge in the field leading to a set of rules and not single rule knowledge. It is noted that the proposed approach can be evaluated not only using common measures but also by the physician who knows the origin of the disease. Those features make our method advantageous compared to other methods presented in the literature.

The evaluation of the proposed method is based on a dataset of 41 subjects. This dataset is divided into three sub-datasets, one for each classification problem (two class, three class and four class problem, respectively). In each one of the above cases, a number of iterations are conducted using different values of the parameters which affect the performance of the classification stage. This procedure is followed in order to select the combination of parameter values which provide the highest values of the used evaluation measures. For the evaluation of the proposed method 10-fold stratified cross validation procedure was employed. For the two and three class problem 108 instances were used while for the four class problem 164 instances were used, four instances per patients. The four instances result from the four functional runs that each

Table 10
Comparison with other methods reported in the literature for the two class problem.

Method	Dataset used	Approach	Sensitivity (%)	Specificity (%)
Our work Random Forests	*Washington University data* 14 healthy elderly subjects 7 elderly subjects with very mild AD 6 elderly subjects with mild AD	1. Preprocessing 2. Modeling 3. Feature extraction 4. Feature selection 5. Classification with Random Forests	94	97
Our work Rules		1. Preprocessing 2. Modeling 3. Feature extraction 4. Feature selection 5. Classification with Random Forests 6. Rules extraction	96	93
Our work SVM		1. Preprocessing 2. Modeling 3. Feature extraction 4. Feature selection 5. Classification with SVM	88	87
Supekar et al. [18]	18 healthy subjects 21 demented subjects	1. Preprocessing 2. Anatomical parcellation 3. Construction of the brain networks 4. Computation of small-world metrics	72	78
Greicius et al. [19,20]	*Washington University data* 14 healthy elderly subjects 7 elderly subjects with very mild AD 6 elderly subjects with mild AD *Resting state data* 7 healthy elderly subjects 9 demented elderly subjects	1. Preprocessing 2. ICA 3. Selection of best fit component 4. Goodness-of-fit analysis	85 89	77 100
Li et al. [21]	9 healthy elderly subjects 10 elderly subjects with AD 5 subjects with mild cognitive impairment	1. Selection of ROI 2. Extraction of voxels time courses 3. Filtering of voxels time courses 4. COSLOF index	80	90

patient performed. The four functional runs were different since two types of stimulation were randomly intermixed such that eight trials of one type and seven of the other appeared in each run. This was also verified by experimental measurements which show that the instances of the same subject present low degree of correlation. The method achieved to classify a patient as healthy or AD (two class problem) and to classify the stages of the disease (three class and the four class problem) with satisfying accuracy either using variation of Random Forests or SVM. More specifically, the accuracy of RF for the three problems is 98%, 86%, and 94%, respectively, while the accuracy of SVM is 88%, 94%, 73%, respectively. Once the forest of the decision trees is constructed, rules are extracted from certain trees of the forest. The application of extracted rules give 94%, 97%, and 99% accuracy for the two, three and four class problem, respectively. Although, the results are very promising, the extracted rules should be further evaluated on different datasets and should be estimated by radiologists in clinical practice.

5. Conclusions

We propose a supervised method for the diagnosis of AD and the classification of the stages of AD in very mild and mild. The method it is fully automated and it addresses both diagnosis and monitoring of the progression of the AD. It combines both imaging findings, behavioral and demographic data and it provides interpretation of the decision through the rules which are extracted in the last stage. In addition, the selection of the features is independent from the learning algorithm (Random Forests, SVM) providing to the physicians a combination of features indicative for the diagnosis and monitoring of the disease. Although, the obtained results are satisfactory (94%, 97%, 99% for the two, three and four class problem, respectively) further improvements, especially in the last two stages (classification and rules extraction), might enhance them. The proposed method can find applications as a diagnostic tool but also in the evaluation of therapeutic procedures of AD.

Appendix A. Weighted voting schemes

RF with wv-1 [38]

INPUT
Training set and a new instance
Step 1:
For each new instance
Compute the similarity with all training instances
Similarity measure: The number of times a tree predictor places then in the same terminal node
Step 2:
Select t most similar instances
Step 3:
Classify t most similar instances with each tree where they are in the OOB set
Compute the margin of each tree
Step 4:
Discard the trees with negative margin
Step 5:
Classify the new instance using weighted voting of the remaining trees
Weights: the average margins on the similar instances
OUTPUT
Classification of the new instance.

RF with wv-2 [44]

INPUT
Training set and a new instance
Step 1:
Compute the distance between the new instance and all training instances using:

$$d(q,x_i)=\sum_{f\in F} w_f\delta(q_f,x_{if}),$$

where q is the unknown instance, x_i is the training instance i, f is a feature of the set of features F, w_f is the weight assigned to feature f, q_f is the value of feature f in the unknown instance q, x_{if} is the value of feature f in the training instance x_i and δ is given as:

$$\delta(q_f.x_{if})=\begin{cases}0 & \text{if } q_f=x_{if}\\ 1 & \text{if } q_f\neq x_{if}\end{cases}.$$

Step 2:
Select k nearest neighbors
Step 3:
Determine the class of the new instance using:

$$\text{Vote}(y_j)=\sum_{n=1}^{k}\frac{1}{d(q,x_c)}I_n(y_j,y_c),$$

where I_n is equal to one if the class label are matched and zero otherwise
OUTPUT
Classification of the new instance

RF with wv-3

INPUT
Training set and a new instance
Step 1:
Compute the distance between the new instance and all training instances using:

$$d(q,x_i)=\sqrt{\sum_{f\in F} vdm_f(q_f,x_{if})^2},$$

where two forms of vdm_f were used [45]:

$$vdm_{af}=\sum_{c=1}^{C}|P_{f,q_f,c}-P_{f,x_{if},c}|^2$$

and

$$vdm_{bf}=\sqrt{\sum_{c=1}^{C}|P_{f,q_f,c}-P_{f,x_{if},c}|^2}.$$

$P_{f,x_{if},c}$ is the conditional probability such that the output class is c given that the attribute f has value x_{if}. C is the number of classes
Step 2:
Select k nearest neighbors
Step 3:
Determine the class of the new instance using:

$$\text{Vote}(y_j)=\sum_{n=1}^{k}\frac{1}{d(q.x_c)}I_n(y_j,y_c),$$

where I_n is equal to one if the class label are matched and zero otherwise

OUTPUT

Classification of the new instance

RF with wv-4 [46]

INPUT

Accuracy of the trained trees and a new instance

Step 1:

For each tree in the forest:

Compute the class cl that tree predicts for the new instance

Increase the votes for the class cl according to the following:

$$\text{vote}(\text{cl}) = \text{vote}(\text{cl}) + \text{accuracy}(T_j),$$

where $\text{accuracy}(T_j)$, is the accuracy of the trained tree T_j

Step 2:

The class of the new instance is the class that maximizes vote(cl)

OUTPUT

Classification of the new instance

RF with wv-5 [47]

INPUT

Training instances and a new instance

Step 1:

Compute the distance between the new instance q and all training instances using:

$$\sigma(q, x_i) = \frac{1}{heom(q, x_i)},$$

where x_i is the ith training instance and

$$heom(q, x_i) = \sqrt{\sum_{f=1}^{m} heom_f^2(q, x_i)}$$

where $heom_f(q, x_i) = \frac{|q - x_i|}{max_f - min_f}$

Step 2:

Select k nearest neighbors of the new instance

Step 3:

Compute the margin of each tree on the jth nearest neighbor of q using:

$$mr_t(x_j) = \left\{ \begin{array}{ll} 1 & \text{correct prediction} \\ -1 & \text{wrong prediction} \end{array} \right\}$$

Step 4:

Compute the weight of each tree using:

$$w_t(q) = \frac{\sum_{j=1}^{k} I(x_j \in \text{OOB}_t)\sigma(q, x_j)mr_t(x_j)}{\sum_{j=1}^{k} I(x_j \in \text{OOB}_t)\sigma(q, x_j)},$$

where I is an indicator function, x_j is the jth nearest neighbor of q and OOB_t is the set of "out-of-bag" instances of tree t

Step 5:

Discard the trees with the highest local errors (the classifiers with error that fall into the upper half of the error interval)

Step 6:

Locally weighted voting is applied to the remaining trees

OUTPUT

Classification of the new instance

RF with wv-6 [48]

INPUT

Whole dataset and a new instance

Step 1:

Initialization of starting population

Step 2:

Fitness evaluation of the population. The fitness function is defined as the recognition rate of the forest when using weighted voting

Step 3:

Examine if the termination criterion is fulfilled

If yes, the algorithm is terminated else go to the next step

Step 4:

Generation of a new population

Parent selection

Crossover

Mutation

Go to Step 3

OUTPUT

Weights for each tree of the forest

References

[1] Carr DB, Goate A, Phil D, Morris JC. Current concepts in the pathogenesis of Alzheimer's disease. Am J Med 1997;103(Suppl 3A):3–10S.

[2] Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. Global prevalence of dementia: a Delphi consensus study. Lancet 2005;366:2112–7.

[3] Alzheimer's Society Drug treatments for Alzheimer's disease. Available online at: http://alzheimers.org.uk/factsheet/407.

[4] Petrella JR, Edward Coleman R, Murali Doraiswamy P. Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. Radiology 2003;226:315–36.

[5] Scheltens P. Early diagnosis of dementia: neuroimaging. J Neuro 1999;246:16–20.

[6] Burggren AC, Bookheimer SY. Structural and functional neuroimaging in Alzheimer's disease: an update. Curr Top Med Chem 2002;2:385–93.

[7] Jezzard P, Matthews PM, Smith SM, editors. Functional MRI: an introduction to methods. USA: Oxford University Press; 2001.

[8] Bookheimer SY, Strojwas MH, Cohen MS, Saunders AM, Pericak-Vance MA, Mazziotta JC, et al. Patterns of brain activation in people at risk of Alzheimer's disease. N Engl J Med 2000;343:450–6.

[9] Thulborn KR, Martin C, Voyvodic JT. Functional MR imaging using a visually guided saccade paradigm for comparing activation patterns in patients with probable Alzheimer's disease and in cognitively able elderly volunteers. AJNR Am J Neuroradiol 2000;21:524–31.

[10] Bassett SS, Yousem DM, Cristinzio C, Kusevic I, Yassa MA, Caffo BS, et al. Familial risk for Alzheimer's disease alters fMRI activation patterns. Brain 2006;129:1229–39.

[11] Sperling RA, Bates JF, Chua EF, Cocchiarella AJ, Rentz DM, Rosen BR, et al. FMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease. J Neurol Neurosurg Psychiatry 2003;74:44–50.

[12] Gronn G, Bittner D, Schmitz B, Wunderlich AP, Riepe MW. Subjective memory complaints: objective neural markers in patients with Alzheimer's disease and major depressive disorder. Ann Neuro 2002;51:491–8.

[13] Grossman M, Koenig P, DeVita C, Glosser G, Moore P, Gee J, et al. Neural basis for verb processing in Alzheimer's disease: an fMRI study. Neuropsychology 2003;17:658–74.

[14] Lustig C, Snyder AZ, Bhakta M, O'Brien KC, McAvoy M, Raichle ME, et al. Functional deactivations: change with age and dementia of the Alzheimer type. PNAS 2003;100:14504–9.

[15] Buckner RL, Snyder AZ, Sanders AL, Raichle ME, Morris JC. Functional brain imaging of young nondemented, and demented older adults. J Cogn Neurosc 2000;12:24–34.

[16] Burns DH, Rosendahl S, Bandila D, Maes OC, Chertkow HM, Shipper HM. Near-infrared spectroscopy of blood plasma for diagnosis of sporadic Alzheimer's disease. J Alzheimer's Dis 2009;17:391–7.

[17] McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ, Holland D, Karow DS, et al. Alzheimer's disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 2009;251:195–205.

[18] Hampel H, Burger K, Teipel SJ, Bokde ALW, Zetterberg H, Blennow K. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimer's Dementia 2008;4:38–48.

[19] Wierenga CE, Bondi MW. Use of functional magnetic resonance imaging in the early identification of Alzheimer's disease. Neuropsychol Rev 2007;17(2):127–43.

[20] Supekar K, Menon V, Rubin D, Musen M, Greicius MD. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol 2008;4(6):1–11.
[21] Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. PNAS 2004;101:4637–42.
[22] Greicius MD, Seeley WW, Miller BL, Glover GH, Parthasarathy S, Beckmann VL, et al. Early diagnosis and intervention. Washington, DC: Alzheimer's Association International Conference on Prevention of Dementia; 2005.
[23] Li SJ, Li Z, Wu G, Zhang MJ, Franczak M, Antuono PG. Alzheimer disease: evaluation of a functional MR imaging index as a marker. Radiology 2002;255:253–9.
[24] Petrella JP, Wang L, Krishnan S, Slavin MJ, Prince SE, Tran TT, et al. Cortical deactivation in mild cognitive impairment: high-field strength functional MR imaging. Radiology 2007;245:224–35.
[25] Alzheimer's Disease Research Center. Available online at: http://alzheimer.wustl.edu/cdr/default.htm.
[26] Strother S. Evaluating fMRI preprocessing pipeline. IEEE Eng Med Biol 2006;25:27–41.
[27] Friston K, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, et al. Analysis of fMRI time-series revisited. NeuroImage 1995;2:45–53.
[28] Friston K, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J-B. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 2000;12:196–208.
[29] D' Esposito M, Zarahn E, Aguirre GK, Rypma B. The effect of normal aging on the coupling of neural activity to the BOLD hemodynamic response. Neuroimage 1999;10:6–14.
[30] Chetelat G, Baron JC. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. Neuroimage 2003;18:525–41.
[31] Ashburner J, Friston K. Multimodal image corregistration and partitioning—a unified framework. Neuroimage 1997;6:209–17.
[32] Friston K, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other hemodynamics. Neuroimage 2000;12:466–77.
[33] Hall MA, Smith LA. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: Proceedings of the Florida artificial intelligence symposium. Orlando, Florida: 1999; p. 235–9.
[34] Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the 20th international conference on machine learning. 2003; p. 856–63.
[35] Tripoliti EE, Fotiadis DI, Argyropoulou M. A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. Artificial Intelligence in Medicine (under review).
[36] Breiman L. Random Forests. Mach Learn 2001;45:5–32.
[37] Tan PN, Steinbach M, Kumar V. Introduction to data mining. Addison Wesley Higher Education; 2006.
[38] Robnik-Sikonja M. Improving Random Forests. In: Proceedings of the European conference on machine learning. 2004; p. 359–69.
[39] Kononeko I. Estimating attributes: analysis and extensions of RELIEF. In: Proceedings of the European conference on machine learning. 1994; p. 171–82.
[40] Kononeko I. On biases in estimating multi-valued attributes. In: Proceedings of IJCAI-95. 1995; p. 1034–40.
[41] Tripoliti EE, Fotiadis DI, Argyropoulou M. An automated supervised method for the diagnosis of Alzheimer's disease based on fMRI data using weigthed voting schemes. In: Proceedings of the IEEE international workshop on imaging systems and techniques—IST 2008. 2008; p. 340–5.
[42] Tripoliti EE, Fotiadis DI, Argyropoulou M. Diagnosis of Alzheimer's disease using fMRI data and modifications of Random Forests algorithm. In: Proceedings of the World congress and biomedical engineering Munich Germany—WC2009 (is going to be presented).
[43] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. USA: Cambridge University Press; 2000.
[44] Cunningham P. A taxonomy of similarity mechanisms for case-based reasoning. Technical Report UCD-CSI-2008-01; 2008.
[45] Wilson DR, Martinez TR. Improved heterogeneous distance functions. JArtif Intell Res 1997;6:1–34.
[46] Hu H, Li J, Wang H, Daggard G, Shi M. A maximally diversified multiple decision tree algorithm for microarray data classification. In: The 2006 workshop on intelligent systems for bioinformatics (WISB2006). Hobart, Australia.
[47] Tsymbal A, Pechenizkiy M, Cunningham P. Dynamic integration with Random Forests. Machine learning: ECML 2006. Berlin/Heidelberg: Springer; 2006.
[48] Gunter S, Bunke H. Optimization of weights in a multiple classifier handwritten word recognition system using a genetic algorithm. Electron Lett Comput Vis Image Anal 2004;3:25–41.