Neuroscience 305 (2015) 361–371

A NOVEL METHOD FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE BASED ON PSEUDO ZERNIKE MOMENT FROM STRUCTURAL MRI

H. T. GORJI * AND J. HADDADNIA

Biomedical Engineering Department, Electrical and Computer Faculty, Hakim Sabzevari University, Sabzevar, Iran

Abstract—Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common type of dementia among older people. The number of patients with AD will grow rapidly each year and AD is the fifth leading cause of death for those aged 65 and older. In recent years, one of the main challenges for medical investigators has been the early diagnosis of patients with AD because an early diagnosis can provide greater opportunities for patients to be eligible for more clinical trials and they will have enough time to plan for future, medical and financial decisions. An established risk factor for AD is mild cognitive impairment (MCI) which is described as a transitional state between normal aging and AD patients. Hence an accurate and reliable diagnosis of MCI can be very effective and helpful for early diagnosis of AD. Therefore in this paper we present a novel and efficient method based on pseudo Zernike moments (PZMs) for the diagnosis of MCI individuals from AD and healthy control (HC) groups using structural MRI. The proposed method uses PZMs to extract discriminative information from the MR images of the AD, MCI, and HC groups. Two types of artificial neural networks, which are based on pattern recognition and learning vector quantization (LVQ) networks, were used to classify the information extracted from the MRIs. We worked with 500 MRIs from the database of the Alzheimer's Disease Neuroimaging Initiative (ADNI 1 1.5T). The 1 slice of 500 MRIs used in this study included 180 AD patients, 172 MCI patients, and 148 HC individuals. We selected 50 percent of the MRIs randomly for use in training the classifiers, 25 percent for validation and we used 25 percent for the testing phase. The technique proposed here yielded the best overall classification results between AD and MCI (accuracy 94.88%, sensitivity 94.18%, and specificity 95.55%), and for pairs of the MCI and HC (accuracy 95.59%, sensitivity 95.89% and specificity 95.34%). These results were achieved using maximum order 30 of PZM and the pattern recognition network with the scaled conjugate gradient (SCG) back-propagation training algorithm as a classifier.

Key words: Alzheimer's disease, mild cognitive impairment, magnetic resonance images, Pseudo Zernike moments, pattern recognition network, learning vector quantization.

INTRODUCTION

Dementia is a general term for diseases and conditions that develop when nerve cells in the brain die or no longer function normally. The most common type of dementia is Alzheimer's disease (AD) (Budson and Solomon, 2011) a progressive neurodegenerative disease (Braak and Braak, 1991). The disease is associated with impaired consciousness and memory loss, and it generally occurs in people aged over 65 (Brookmeyer et al., 1998). The disease was first described by German psychiatrist Alois Alzheimer in 1906 (Berchtold and Cotman, 1998). The incidence of AD and other dementia-related disease has been increasing rapidly each year throughout the world. For example, approximately 4.7 million Americans had AD in 2010, and this number had increased to 5.3 million in 2013. It has been estimated that there will be 13.8 million AD patients in the U.S. by 2050 (Hebert et al., 2013). There were 68% more deaths from AD in 2010 than there were in 2000, making AD the sixth leading cause of death in the U.S. (Hebert et al., 2001; Escudero et al., 2011).

In 2012, the National Institute of Aging and the Alzheimer's Association proposed new criteria and guidelines for describing and categorizing the changes in the brain associated with AD and other dementia-related diseases (Hyman et al., 2012). These criteria define the three stages of AD as pre-clinical AD, mild cognitive impairment (MCI) due to AD, and dementia due to AD. MCI is described as a transitional state between normal aging and AD patients, and people with MCI exhibit difficulties with memory or thinking, but the impairment is not severe enough to affect their ability to conduct their daily activities (Petersen et al., 1999). In the new criteria and guidelines that were proposed in 2011 (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011) MCI is actually an early stage of Alzheimer's or other dementia-related diseases. Approximately 50% of the people who see a doctor concerning the symptoms of MCI symptoms will develop dementia in three to four years (Petersen et al., 1999). Therefore, the ability to

*Corresponding author.
E-mail address: hamedtaheri@yahoo.com (H. T. Gorji).

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's disease Neuroimaging Initiative; GM, gray matter; HC, healthy control; LVQ, learning vector quantization; LVQNN, Learning Vector Quantization neural network; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; PET, positron emission tomography; PZMs, pseudo Zernike moments; ROI, regions of interest; SBA, surface-based analysis; SCG, scaled conjugate gradient; VBM, voxel-based morphometry; ZMs, Zernike moments.

http://dx.doi.org/10.1016/j.neuroscience.2015.08.013
0306-4522/

diagnose and classify MCI from AD patients and HC groups precisely will be very effective and helpful for the early diagnosis of AD. In recent years, several neuroimaging techniques have been used in the clinical diagnosis and classification of AD patients, MCI patients, and healthy individuals. These techniques include computed tomography (CT), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT), and several attempts have been made to prove and compare the efficiency and reliability of these techniques (Erkinjuntti et al., 1989; Toyama et al., 2005; Davatzikos et al., 2008; Chaves et al., 2009; Magnin et al., 2009; Cuingnet et al., 2011; Illán et al., 2011; Tripoliti et al., 2011; Ortiz et al., 2013). Usually, researchers use three types of well-known techniques for diagnosing and classifying the AD, MCI, and HC groups. Some previous studies were based on volumetric measurements of segmented regions of interest (ROI), (Convit et al., 1997, 2000; Kaye et al., 1997; Rusinek et al., 2004; Tapiola et al., 2008) such as the hippocampus, entorhinal cortex, and gray matter (GM) in some regions that are affected by AD. Several investigators have proposed the use of Voxel-based Morphometry (VBM) for measuring the spatial distribution of atrophy in the white matter (WM) and GM of the brain as well as in the cerebrospinal fluid (CSF) in MCI and AD patients (Hämäläinen et al., 2007; Guo et al., 2010; Shin et al., 2010; Kim et al., 2011; Li et al., 2012), whereas other groups use cortical thickness as a feature for diagnosing AD, MCI, and HC (Lerch et al., 2008; Hua et al., 2009; McDonald et al., 2009; Lehmann et al., 2011; Grand'Maison et al., 2013).

For many years, moments have been used as descriptors of the properties of the images in pattern recognition, and many researchers still use them in several applications.

Zernike moments (ZMs) have attracted extensive attention as a powerful feature extractor in pattern recognition due to their high robustness to noise and their good performance in recognizing circular shapes, such as faces (Jenkinson et al., 2005). ZMs can be used to extract structural facial features that are local features of face images, for example, the shapes of eyes, nose and mouth in human face images. Since Zernike polynomials are orthogonal to each other, ZMs can represent the properties of an image with no redundancy or overlap of information between the moments, so they are used extensively in various applications (Haddadnia et al., 2003; Li et al., 2009; Liyun et al., 2009; Tahmasbi et al., 2011).

Although the ZMs are very useful in image processing, they have some limitations. Thus, an improved version was developed, which is referred to as pseudo Zernike moment (PZM).

Current research is directed toward the development of a novel method for early diagnosis of AD based on PZM's extraction of features of AD, MCI, and HC from MR images. In this study, we used pattern recognition and learning vector quantization (LVQ) networks as classifiers to discriminate between the three groups.

MATERIALS

Data

Data used in the preparation of this paper were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

Subjects

The general eligibility criteria used for the inclusion of participants were those defined in the ADNI protocol (described in details at http://www.adni-info.org/Scientists/AboutADNI.aspx#). Enrolled subjects were between 55 and 90 (inclusive) years of age, had a study partner able to provide an independent evaluation of functioning, and spoke either English or Spanish. All subjects were willing and able to undergo all test procedures, including neuroimaging and agreed to longitudinal follow up. Between twenty and fifty percent must be willing to undergo two lumbar punctures, spaced one year apart. Specific psychoactive medications were excluded. General inclusion/exclusion criteria were as follows: healthy control subjects (HC) had Mini Mental State Examination (MMSE) scores between 24 and 30 (inclusive), a Clinical Dementia Rating (CDR) of zero. They were non- depressed, non MCI, and non-demented. MCI subjects had MMSE scores between 24 and 30 (inclusive), a memory complaint, had objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. AD patients had MMSE scores between 20 and 26 (inclusive), CDR of 0.5 or 1.0, and met NINCDS/ADRDA criteria for probable AD. In this paper, we used a total of 500 subjects, including 180 AD patients, 172 MCI subjects and 148 HC individuals. The demographic characteristics of all subjects are shown in Table 1.

MRI acquisition

In this paper, we used T1-weighted MR images from 1.5 T scanners acquired according to the ADNI acquisition protocol (Jack et al., 2008). All MR images that we used had undergone specific preprocessing correction steps.

Table 1. Clinical and demographic information of the study population. For each group, the total number of subjects (N), number of men (M) and number of females (F) are shown, along with the average age, Standard deviation (SD), average mini mental state examination (MMSE) score, and clinical dementia rating (CDR)

	HC (N = 148; 67 F/81 M)			MCI (N = 172; 77 F/95 M)			AD (N = 180; 75 F/105 M)		
	Mean	SD	Range	Mean	SD	Range	Me	SD	Range
Age	75.3	5.6	60–86	76.3	6.5	55–88	75.4	6.3	56–89
MMSE	28.39	1.4	25–30	26.8	1.7	24–30	23.8	2.1	20–27
CDR	0	0.0	0–0	0.5	0.0	0.5–0.5	0.7	0.3	0.5–1

First of all, the image geometry distortion due to gradient non-linearity was corrected by a system-specific correction that was called Gradwarp after that B1 non-uniformity was used to for correct the image intensity non-uniformity and in the third step N3 that is a histogram peak sharpening algorithm was applied to all images for reducing the residual intensity non-uniformity due to the wave or the dielectric effect at 3T and 1.5T scans.

Meanwhile, all MR images downloaded from the public ADNI site (www.loni.ucla.edu/ADNI) with Neuroimaging Informatics Technology Initiative (NIfTI) format.

FEATURE EXTRACTION METHOD

In order to design an accurate diagnostic system from MRIs the choice of feature extraction methods is very important. Moment invariants as a statistical-based approach for their invariance properties have received much attention in recent years. The term invariant denotes that an image feature remains unchanged if that image undergoes one or a combination of the changes such as: change of size (scale), change of position (translation), change of orientation (rotation), and reflection. In this paper we used the PZMs for extracting the features from the MRIs. PZM is a powerful shape descriptor and a form of projection of the entire image intensity function onto the space of the pseudo Zernike polynomials of order n with repetition m. As earlier mentioned, the orthogonality implies no redundancy or overlap of information between the moments with different orders and repetitions. This property enables the contribution of each moment to be unique and independent from the information in an image (Hwang and Kim, 2006). There are three steps for computation of the PZM from an input image: first of all computation of radial polynomials, after that computation of pseudo Zernike basis functions and finally computation of PZMs by projecting the image onto the pseudo Zernike basis functions (Hwang and Kim, 2006; Li et al., 2009).

ZMs

Zernike polynomials were introduced by Frits Zernike in 1934 (Zernike, 1934) as a set of orthogonal, complex-valued polynomials, over the interior of the unit circle (Fig. 1).

The complete set of Zernike polynomial or Zernike basis function $V_{n,m}$ is defined as:

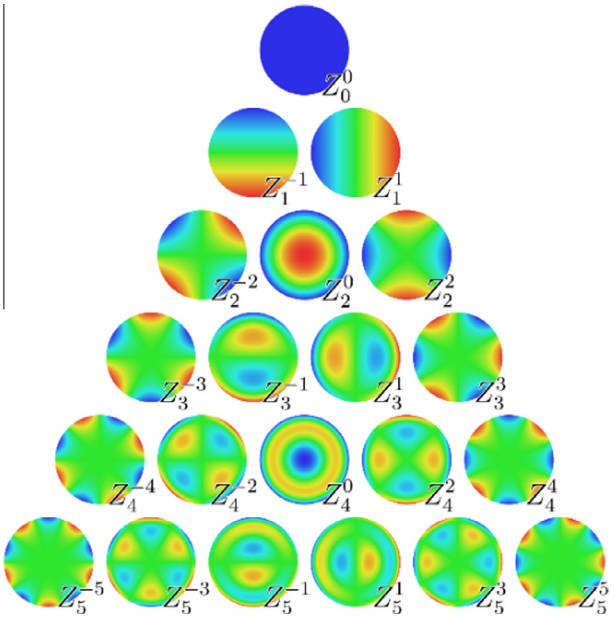

Fig. 1. Plots of Zernike polynomials in the unit disk.

$$V_{nm}(x,y) = R_{nm}(x,y)e^{jm\theta}$$

where $j = \sqrt{-1}$, $\theta = \tan^{-1}\left(\frac{y}{x}\right)$

And $R_{n,m}$ is a set of real-valued radial polynomial which is defined as:

$$R_{nm}(x,y) = \sum_{s=0}^{\frac{n-|m|}{2}} S_{n,|m|,s}(x^2+y^2)^{n-2s}$$

$$S_{n,|m|,s} = (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}$$

where $n \geqslant 0$, $|m| \leqslant n$, $x^2+y^2 \leqslant 1$ and $n-|m| = even$.

The complex ZMs of order n and repetition m are defined as:

$$ZM_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)V^*_{nm}(x,y)$$

where $f(x,y)$ is the image function and * denotes the complex conjugate.

The equation of ZMs by scale invariant central moments is described as follows:

$$ZM_{nm} = \frac{n+1}{\pi}\sum_{\substack{k=|m|\\ n-k=even}}^{n}\sum_{a=0}^{b}\sum_{d=0}^{m}(-j)^d\binom{|m|}{d}\binom{b}{a}S_{n,|m|,s}\mathrm{CM}_{k-2a-d,2a+d}$$

where CM_{pq} is the scale invariant Central moments and $b = (N - |M|)/2 - S$.

PZMs

Zernike polynomials contain only $\frac{1}{2}(n+1)(n+2)$ linear, independent polynomials, and this is their main limitation. Therefore, fewer features will be generated by ZMs, hence PZMs were proposed by Bhatia and Wolf (1954) to solve this problem. The difference between ZM and PZM is in the real-valued radial polynomials, which are defined as:

$$R_{n,m}(x,y) = \sum_{s=0}^{n-|m|} D_{n,|m|,s}(x^2+y^2)^{\frac{n-s}{2}}$$

where

$$D_{n,|m|,s} = (-1)^s \frac{(2n+1-s)!}{s!(n-|m|-s)!(n-|m|-s+1)!}$$

The PZM of order n with repetition m can be computed by the scale invariant central moments (CM_{pq}) and the Radial Geometric Moments (RM_{pq}), which are described by the following equations: (Belkasim et al., 1991; Bailey and Srinath, 1996):

$$\begin{aligned} PZMI_{nm} = {} & \frac{n+1}{\pi} \sum_{(n-m-s)even,s=0}^{n-|m|} D_{n,|m|,s} \\ & \times \sum_{a=0}^{k}\sum_{b=0}^{m} \binom{k}{a}\binom{m}{b} (-j)^b \mathrm{CM}_{2k+m-2a-b,2a+b} \\ & + \frac{n+1}{\pi} \sum_{(n-m-s)odd,s=0}^{n-|m|} D_{n,|m|,s} \\ & \times \sum_{a=0}^{d}\sum_{b=0}^{m} \binom{d}{a}\binom{m}{b} (-j)^b \mathrm{RM}_{2d+m-2a-b,2a+b} \end{aligned}$$

where $k = (n-s-m)/2$ and $d = (n-s-m+1)/2$; CM_{pq} and RM_{pq} are as follows:

$$\mathrm{CM}_{pq} = \frac{\mu_{pq}}{M_{00}^{(p+q+2)/2}}$$

where μ_{pq} denotes the central moment of the connected component, and M_{pq} is the geometric moments of order $p + q$ of a digital image that can be described as:

$$\mu_{pq} = \sum_x \sum_y f(x,y)(x-x_0)^p (y-y_0)^q$$

$$M_{pq} = \sum_x \sum_y f(x,y) x^p y^q$$

where $p,q = 0,1,2,\ldots$, $f(x,y)$ is the gray scale digital image at x,y location, $x_0 = M_{10}/M_{00}$ and $y_0 = M_{01}/M_{00}$ are the centers of the image. And

$$\mathrm{RM}_{pq} = \frac{\sum_x \sum_y f(x,y)(\hat{x}^2+\hat{y}^2)^{\frac{1}{2}} \hat{x}^p \hat{y}^q}{M_{00}(p+q+2)/2}$$

where $\hat{x} = x - x_0$ and $\hat{y} = y - y_0$.

CLASSIFICATION METHODS

Artificial neural networks (ANNs) have been used extensively as classifiers in diagnostic systems, the analysis of medical images, predicting and forecasting diseases in multiple areas, such as cardiology, radiology, pathology, and pulmonology (Itchhaporia et al., 1996; Flores-Fernández et al., 2012; Sartakhti et al., 2012; Vieira et al., 2013; Chankong et al., 2014). In this paper, we used pattern recognition and LVQ neural networks to classify people into the AD, MCI, and HC groups.

Pattern recognition neural network (PRNN)

Pattern recognition neural networks are feed-forward networks that can be trained to classify inputs according to target classes.

Information always moves in only one direction in feed-forward neural networks, and there is no feedback. The information moves forward from the input layer through the hidden layer to the output layer (Fig. 2). A two-layer feed-forward network, with sigmoid hidden and softmax output neurons, can classify vectors arbitrarily well, given enough neurons in its hidden layer.

Training functions. After initializing the weights and biases of the network, it will be ready for training. There are several algorithms that can be used to train artificial neural networks (Hagan et al., 1996). In the present study, we used three types of training algorithms. The first was the Levenberg–Marquardt (LM) back-propagation algorithm (Marquardt, 1963), which often has the fastest convergence of the training algorithms and usually performs better with function-fitting problems than with pattern recognition problems. The LM is very useful when accurate training is required, but the advantage of LM decreases when the number of weights in the network increases. The second algorithm is a resilient back-

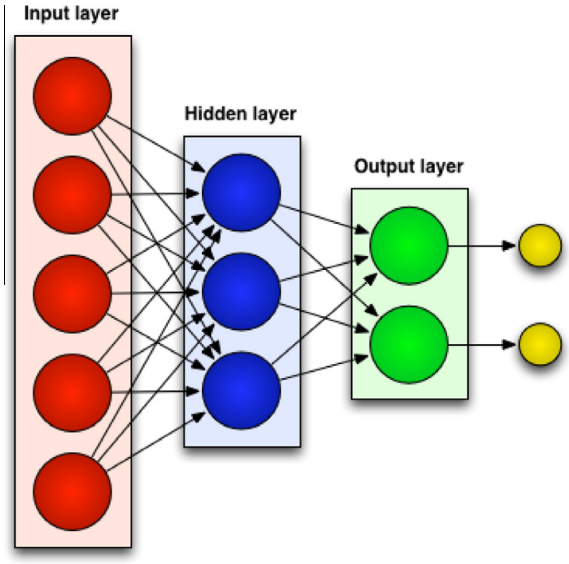

Fig. 2. Feed-forward neural networks.

propagation (RB) algorithm that was introduced by Martin Riedmiller and Heinrich Braun (Riedmiller and Braun, 1992). The RB algorithm is a local, adaptive learning scheme that was proposed to eliminate the harmful influence of the magnitudes of the partial derivatives on the weight step (Baykal and Erkmen, 2000). The scaled conjugate gradient (SCG) algorithm (SCG) (Møller, 1993) is the third algorithm that was used in the current study, and it is based on conjugate directions and designed to avoid the time-consuming line search. The SCG algorithm is almost as fast as the LM algorithm, and it will be faster for large networks on function approximation problems.

The term 'back propagation' (BP) that is used in all of these methods refers to an algorithm that is a supervised learning method for multi-layer, feed-forward networks. This network was proposed by Rumelhart and McClelland (Rumelhart et al., 1988) for determining the optimal weight parameters. The back-propagation method is also known as the error back-propagation algorithm, which is based on the error-correction learning rule. In this algorithm, network weights are selected randomly, and, in each iteration, the difference between network output and desired output (error) is calculated; therefore the goal is to minimize the error by adjusting the weights.

Learning Vector Quantization neural network (LVQNN)

LVQ is a supervised neural network that was proposed by Teuvo Kohonen (Kohonen and Maps, 1995). LVQ is a combination of competitive learning with supervision, and it is well known and extensively used in processing medical images and in bioinformatics or pattern classification (Ganesh Murthy and Venkatesh, 1998; Kara and Güven, 2007; Martín-Valdivia et al., 2007; Hung et al., 2011). An LVQ network is made up of two layers. The first layer is a competitive layer that maps input vectors into clusters that are identified by the network during training. The second layer is a linear layer that maps the merging of the groups of first-layer clusters into classes defined by the target data.

EXPERIMENTS AND RESULTS

As mentioned above, all MR images that we downloaded from ADNI were pre-processed. We considered 1 slice of all MR images in axial view (Fig. 3), and PZMs was calculated as features for each MRI.

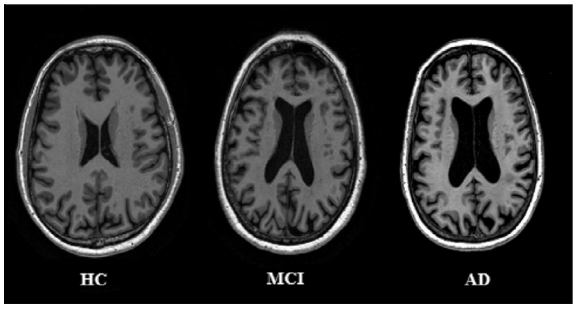

Fig. 3. Axial view of HC, MCI and AD.

For analyzing the effect of the orders of PZM on the performance of the overall system, four categories were defined as feature vectors based on PZM's orders. Maximum order of the first category is 5 and includes 12 moments of PZM that are considered as feature vector elements, which satisfy the following conditions:

$$\text{Category1} = \left\{ \begin{array}{c} 0 \leqslant n \leqslant 5 \\ |m| \leqslant n \\ n - |m| = 2k \\ k \in N \end{array} \right\}$$

In the second category, maximum order of the second category is 10 and the number of moments is 36, and these moments satisfy the following conditions:

$$\text{Category2} = \left\{ \begin{array}{c} 0 \leqslant n \leqslant 10 \\ |m| \leqslant n \\ n - |m| = 2k \\ k \in N \end{array} \right\}$$

In the third and fourth categories, maximum orders are 20 and 30 and there are 121 and 256 moments, respectively, that satisfy the following conditions:

$$\text{Category3} = \left\{ \begin{array}{c} 0 \leqslant n \leqslant 20 \\ |m| \leqslant n \\ n - |m| = 2k \\ k \in N \end{array} \right\}$$

$$\text{Category4} = \left\{ \begin{array}{c} 0 \leqslant n \leqslant 30 \\ |m| \leqslant n \\ n - |m| = 2k \\ k \in N \end{array} \right\}$$

Table 2 shows feature vector elements of PZM for four categories.

To demonstrate the applicability of our proposed method, an experimental study was conducted on the 1 slice of 500 ADNI MRI database, which includes 180 AD patients, 172 MCI subjects, and 148 HCs. We chose 50 percent of images (250 images) at random and used them in the training set (90 AD, 86 MCI and 74 HC); 25 percent of MR images (125 images) were used as the validation set (45 AD, 43 MCI and 37 HC), and the rest of 25 percent MR images (125 images) used as the testing set for evaluating the classification performances.

Training set is presented to the network during training, and the network is adjusted according to its error. Validation set is used to measure network generalization, and to halt training when generalization stops improving and the testing set has no effect on training and so provide an independent measure of network performance during and after training (Hai-Jew,

Table 2. Feature vector elements of PZM

Category	Maximum Order of PZM	No. of moments
1	5	12
2	10	36
3	20	121
4	30	256

2015). We used the validation set to avoid the overfitting problem. In order to prevent overfitting, it is necessary to use some additional techniques (e.g. cross-validation, regularization, early stopping) (Li, 2008). In this paper we used the early stopping technique. For using this technique the available data should be divided into three subsets. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned (Du and Swamy, 2013).

Three steps were performed. First, PZMs of 1 slice of 500 MRIs were computed and used as features. Second, the extracted features were used to feed the pattern recognition and LVQ networks. Third, the effect of PZMs' orders and three training algorithms were assessed to identify relevant pairs of diagnostic groups (AD/MCI, and MCI/HC).

The performance of the proposed method was evaluated by the percentages of three measurements, i.e., sensitivity, specificity and accuracy. The respective definitions of these parameters are described as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

where TP, TN, FP, and FN denote true positives (i.e., the number of patients who were correctly classified), true negatives (i.e., the number of HCs who were correctly classified), false positives (i.e., the number of patients who were classified as HCs), and false negatives (i.e., the number of members of the HC group who were classified as patients), respectively. After training the pattern recognition and LVQ networks using the training set, we tested our proposed method using the test set. To increase reliability we train, validate and test our proposed method 5 times for each AD/HC, AD/MCI and MCI/HC group and the average of results considered. Tables 3–5 present the results of the classification experiments for the pairs of AD/HC, AD/MCI, and MCI/HC, respectively. Figs. 4–6 show the accuracy rates for the classifications of the AD/HC, AD/MCI, and MCI/HC groups with respect to maximum order of PZM, PRNN training algorithms and LVQ neural network.

Classification of AD and HC

The classification results for pairs of AD/HC are shown in Table 3, and Fig. 4.

The best results that were obtained when the maximum order of PZM was selected as 30 (256 features) and the pattern recognition network was trained by SCG training algorithm were sensitivity 96.64%, specificity 97.79%, and accuracy 97.27%, whereas classification based on the LVQ network resulted in 89.72% sensitivity, 90.76% specificity, and 90.30% accuracy.

Classification of AD and MCI

The best overall classification results (sensitivity 94.18%, specificity 95.55%, and accuracy 94.88%) for groups of AD and MCI were obtained when PZM was used with maximum order 30 (256 features) and the pattern recognition network was trained by the SCG algorithm. The LVQ network obtained lower results (sensitivity 90.01%, specificity 90.27%, and accuracy 90.62%) than a pattern recognition network. The classification results are shown in Table 4, and Fig. 5.

Table 3. Classification results of AD vs. HC

	Maximum order of PZM	No. of features	Classification methods	Training algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
AD/HC	5	12	PRNN	SCG	93.41	92.43	92.89
				LM	92.68	90.42	91.47
				RB	93.12	89.06	90.90
			LVQNN		85.22	88.63	86.93
	10	36	PRNN	SCG	92.10	95.50	93.93
				LM	92	94.44	93.33
				RB	90.27	90.32	90.30
			LVQNN		88.81	88.77	88.78
	20	121	PRNN	SCG	94	96.11	95.15
				LM	9.55	95.08	94.84
				RB	91.78	92.39	92.12
			LVQNN		86.92	91.52	89.32
	30	256	PRNN	SCG	96.64	97.79	97.27
				LM	95.97	97.23	96.66
				RB	94.59	95.60	95.15
			LVQNN		89.72	90.76	90.30

Table 4. Classification results of AD vs. MCI

Maximum order of PZM	No. of features	Classification methods	Training algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
AD/MCI	5	12	SCG	89.75	88.70	89.20
		PRNN	LM	90.36	89.24	89.77
			RB	87.05	87.91	87.50
		LVQNN		85.54	87.70	86.64
	10	36	SCG	89.77	93.18	91.47
		PRNN	LM	89.71	92.65	91.19
			RB	89.47	90.60	90.05
		LVQNN		89.95	86.77	87.78
	20	121	SCG	92.39	93.37	92.89
		PRNN	LM	92.35	92.85	92.61
			RB	92.07	89.89	90.90
		LVQNN		85.39	89.65	87.50
	30	256	SCG	94.18	95.55	94.88
		PRNN	LM	95.23	94.56	94.87
			RB	91.27	92.77	92.04
		LVQNN		90.01	90.27	90.62

Table 5. Classification results MCI vs. HC

	Maximum order of PZM	No. of features	Classification methods	Training algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
MCI/HC	5	12		SCG	89.65	89.59	89.62
			PRNN	LM	88.80	90.79	88.36
				RB	85.62	89.69	87.73
			LVQNN		83.44	86.82	85.22
	10	36		SCG	88.96	93.29	91.19
			PRNN	LM	92.30	90.85	91.50
				RB	89.18	90.58	89.93
			LVQNN		86.75	89.82	88.36
	20	121		SCG	93.37	95.80	94.65
			PRNN	LM	93.28	94.67	94.02
				RB	91.78	91.86	91.82
			LVQNN		89.04	89.53	89.30
	30	256		SCG	95.89	95.34	95.59
			PRNN	LM	95.20	94.76	94.96
				RB	92.36	91.37	91.82
			LVQNN		90.97	90.22	90.56

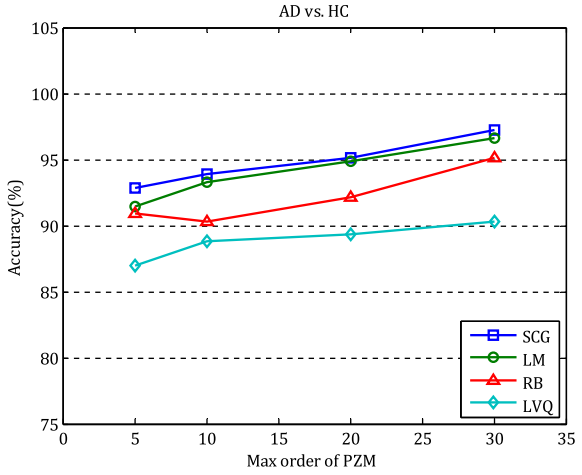

Fig. 4. Classification accuracy of AD vs. HC with respect to Max orders of PZM, 3 PRNN Training algorithm and LVQ.

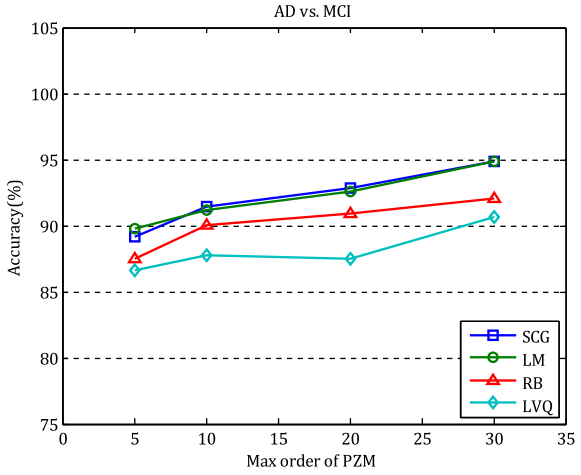

Fig. 5. Classification accuracy of AD vs. MCI with respect to Max orders of PZM, 3 PRNN Training algorithm and LVQ.

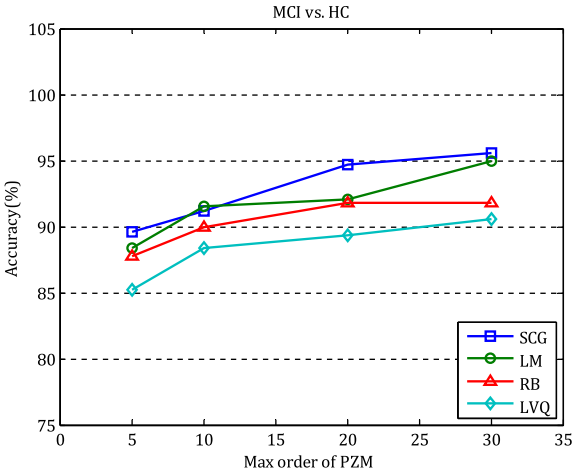

Fig. 6. Classification accuracy of MCI vs. HC with respect to Max orders of PZM, 3 PRNN Training algorithm and LVQ.

Classification of MCI and HC

The best classification result of maximum order 30 (256 features) of PZM was achieved (sensitivity 95.89%, specificity 95.34%, and accuracy 95.59%) when we used the pattern recognition network and the SCG training algorithm as was the case for the AD/MCI classification results mentioned above. As expected, the LVQ classification result was lower than the other classification method (sensitivity 90.97%, specificity 90.22%, and accuracy 90.56%). The results are shown in Table 5, and Fig. 6.

DISCUSSION

Several methods have been proposed for the diagnosing dementia in patients. Some of the recent studies have used segmented ROI, VBM, and surface-based analysis (SBA) for diagnosis of AD, MCI, and HC.

Recent Studies based on ROI, VBM, and SBA have reported classification rates between 75% and 94.3% (Fan et al., 2008a; Magnin et al., 2009; Zhang et al., 2011; Gray et al., 2013; Ortiz et al., 2013). The high variation in these reported results is partly due to different databases and study populations being used but also due to the fact that different studies using different classifiers and common types of cross-validation approaches, e.g. K-fold or leave-one-out cross-validation.

To the best of our knowledge, the present study is the first in which the PZM approach was used with the aim of early diagnosing AD, based on T1-weighted structural MRI data. PZM is a powerful shape descriptor. A relatively small set of PZMs can characterize the global shape of a pattern effectively. Lower order moments represent the global shape of a pattern and higher order moments represent the details. PZM descriptor has such desirable properties that make it one of the most suitable choices for our work: rotation invariance, robustness to noise, expression efficiency, fast computation and multi-level representation for describing the various shapes of patterns. With a proper normalization method, scale invariance can also be achieved.

Therefore PZMs have been used to extract the AD, MCI, and HC features from MRIs with the aim of generating feature vectors for the task of diagnosing and classifying the AD/MCI, and MCI/HC pairs. Also, to evaluate and compare the performances of our proposed method, two types of classifiers (pattern recognition with the three training algorithm and LVQ networks) were used for the classification of the AD, MCI, and HC groups. Sensitivity, specificity, and accuracy are three parameters that determine the results of the classification. The ADNI MRI data were separated randomly into two groups of the same size, i.e., the training set and the test set. After training the classifiers by the features that were extracted by PZM, the performance of our proposed method was evaluated using the test set.

As shown in 'Experiments and results' section, our proposed method obtained high values of sensitivity, specificity and accuracy. The performance of the proposed method improved significantly when the maximum order of PZMs increased from 5 (12 features) to 30 (256 features) and pattern recognition network with the SCG back-propagation training algorithm used as a classifier (4.38% for AD/HC, 5.68% for AD/MCI and 5.97% for MCI/HC).

Tables 3–5 shows classification results for AD/HC, AD/MCI and MCI/HC by PRNN and LVQ network respectively for four categories of maximum orders of PZMs. The Classification accuracy of AD/HC, AD/MCI and MCI/HC with respect to Maximum orders of PZMs, 3 PRNN Training algorithm and LVQNN is shown in Figs. 4–6 respectively.

This technique yielded the best overall classification results between AD and MCI (accuracy 94.88%, sensitivity 94.18%, and specificity 95.55%), and for pairs of the MCI and HC (accuracy 95.59%, sensitivity 95.89% and specificity 95.34%) which is higher compared with state-of-the-art methods.

In several studies classification between AD/HC and MCI/HC was evaluated. Therefore, we also assessed the classification performance of AD over HC using our proposed method.

The lower classification results that we obtained by PZM with a maximum order 5 and LVQNN were for the AD/HC pairs (sensitivity 89.72%, specificity 90.76%, and accuracy 90.30%), which were comparable to the best results that have been reported in Table 6.

Classification between MCIc (MCI individuals that converted to AD) and MCInc (MCI non-converters to AD) has been proposed in a few studies (Misra et al., 2009; Zhang et al., 2011; Aguilar et al., 2013; Ota et al., 2014; Suk et al., 2014). However, since the preprocessing steps were different, separate training and test sets were used, the classification rate varied between 74% and 86% for the pairs of MCIc/MCInc. Because the aim of this study is the early diagnosis of AD, and as mentioned above MCI is a transitional state between HC people and AD patients, we achieved classification results between the pairs of AD/MCI and MCI/HC to distinguish between MCI and two other groups. However, for discrimination between AD/MCI and MCI/HC, our proposed

Table 6. Comparison of the proposed method with other studies

Reference	Year	Feature extraction method	Database	Acc of AD/HC (%)	Acc of AD/MCI (%)	Acc of MCI/HC (%)
Our proposed methods	2014	Pseudo Zernike moment	ADNI	**97.27**	**94.88**	**95.59**
Shen et al. (Shen et al., 2014)	2014	GM, WM, CSF	ADNI	93.2	–	86.5
Wang et al. (Wang et al., 2013)	2013	GM, WM, CSF	ADNI	93.3	–	78.9
Yang et al. (Yang et al., 2013)	2013	Volume and shape by SPM	Private	94.12	–	88.89
Gray et al.	2013	VBM and biomarkers	ADNI	89	–	75
Ortiz et al.	2013	Tissue distribution of GM and WM	ADNI	90	–	–
Zhang et al.	2011	(ROIs) and CSF biomarkers	ADNI	93.2	–	76.4
Escudero et al. (Escudero et al., 2011)	2011	Anatomical analysis by Freesurfer	ADNI	89.2	–	72.7
Ramírez et al.	2010	Partial least squares	Private	96.9	–	–
López et al. (López et al., 2009)	2009	Voxel of ROI	Private	92.31	–	–
Colliot et al. (Colliot et al., 2008)	2008	Hippocampus Segmentation	Private	84	–	73

method can achieve a classification accuracy of 94.88%, and 95.59% respectively.

The best classification results we obtained for MCI vs. HC, i.e., sensitivity 95.89%, specificity 95.34%, and accuracy 95.59%), were lower than those in the study reported in the original paper by (Fan et al., 2008b). They reached 100% accuracy when they chose 68 or 75 brain regions as features, but, as mentioned in their paper, for a reasonable range of the brain region numbers (between 50 and 80), they obtained an average accuracy of about 93%, whereas the average result of our proposed method was 92.76%, is comparable to their results. However, for an exact comparison of the two studies, we should consider an important point, i.e., number of subjects used in the two experiments. In our study, we used 320 subjects (172 MCI and 148 HC), while just 30 subjects (15 MCI and 15 HC) were used in the other study. Thus, we can claim that the current study achieved better performance than the other study. Table 6 provides a brief comparison between our proposed method and the other diagnosis and classification results.

CONCLUSION

A novel method based on PZMs was presented in this paper for extracting discriminative information from structural MRI with the aim of the early diagnosis of AD and classification between patients with AD, patients with MCI, and HC subjects. The results on ADNI MRI data for 500 subjects showed that our feature extraction and classification methods achieved high accuracy, high sensitivity, and high specificity for the AD and MCI groups as well as the HC group. A pattern recognition network with three training algorithms and a LVQ network were used as classifiers in this work. The best results were obtained when the maximum order of PZM was 30 and the pattern recognition network with the SCG back-propagation training algorithm was used as the classifier.

The proposed method yielded 97.27% classification accuracy (96.64% sensitivity and 97.79% specificity) for the pairs of AD/HC, 94.88% classification accuracy (94.18% sensitivity and 95.55% specificity) for the pairs of AD/MCI, and 95.59% classification accuracy (95.89% sensitivity and 95.34% specificity) for classification of the pairs of MCI/HC. These results demonstrate the high precision and reliability of our proposed method and indicate that the proposed method could provide a good choice for the early diagnosis of AD.

CONFLICT OF INTEREST

We declare no conflict of interest.

REFERENCES

Aguilar C et al (2013) Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Res: Neuroimaging 212 (2):89–98.

Albert MS et al (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dementia 7(3):270–279.

Bailey RR, Srinath M (1996) Orthogonal moment features for use with parametric and non-parametric classifiers. Pattern Analysis Machine Intel, IEEE Trans on 18(4):389–399.

Baykal N, Erkmen AM (2000). Resilient backpropagation for RBF networks. Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. In: Proceedings. Fourth International Conference on, IEEE.

Belkasim SO et al (1991) Pattern recognition with moment invariants: a comparative study and new results. Pattern Recogn 24 (12):1117–1138.

Berchtold NC, Cotman CW (1998) Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. Neurobiol Aging 19(3):173–189.

Bhatia A, Wolf E (1954) On the circle polynomials of Zernike and related orthogonal sets. Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge Univ Press.

Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82(4):239–259.

Brookmeyer R et al (1998) Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. Am J Public Health 88(9):1337–1342.

Budson AE, Solomon PR (2011) Memory loss: a practical guide for clinicians. Elsevier Health Sciences.

Chankong T et al (2014) Automatic cervical cell segmentation and classification in Pap smears. Comput Methods Programs Biomed 113(2):539–556.

Chaves R et al (2009) SVM-based computer-aided diagnosis of the Alzheimer's disease using *t*-test NMSE feature selection with feature correlation weighting. Neurosci Lett 461(3):293–297.
Colliot O et al (2008) Discrimination between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus 1. Radiology 248(1):194–201.
Convit A et al (1997) Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. Neurobiol Aging 18 (2):131–138.
Convit A et al (2000) Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease☆. Neurobiol Aging 21(1):19–26.
Cuingnet R et al (2011) Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56(2):766–781.
Davatzikos C et al (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 29(4):514–523.
Du K-L, Swamy M (2013) Neural networks and statistical learning. Springer Science & Business Media.
Erkinjuntti T et al (1989) White matter low attenuation on CT in Alzheimer's disease. Arch Gerontol Geriatr 8(1):95–104.
Escudero J, et al. (2011). Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials. Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE.
Fan Y et al (2008a) Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. NeuroImage 39(4):1731–1743.
Fan Y et al (2008b) Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. NeuroImage 41(2):277–285.
Flores-Fernández JM et al (2012) Development of an optimized multi-biomarker panel for the detection of lung cancer based on principal component analysis and artificial neural network modeling. Expert Syst Appl 39(12):10851–10856.
Ganesh Murthy C, Venkatesh YV (1998) Encoded pattern classification using constructive learning algorithms based on learning vector quantization. Neural Net 11(2):315–322.
Grand'Maison M et al (2013) Early cortical thickness changes predict β-amyloid deposition in a mouse model of Alzheimer's disease. Neurobiol Dis 54:59–67.
Gray KR et al (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage 65:167–175.
Guo X et al (2010) Voxel-based assessment of gray and white matter volumes in Alzheimer's disease. Neurosci Lett 468(2):146–150.
Haddadnia J et al (2003) An efficient feature extraction method with pseudo-Zernike moment in RBF neural network-based human face recognition system. EURASIP J Appl Signal Proc 2003:890–901.
Hagan MT et al (1996) Neural network design. Boston: Pws.
Hai-Jew S (2015). Enhancing Qualitative and Mixed Methods Research with Technology.
Hämäläinen A et al (2007) Voxel-based morphometry to detect brain atrophy in progressive mild cognitive impairment. Neuroimage 37 (4):1122–1131.
Hebert LE et al (2001) Annual incidence of Alzheimer disease in the United States projected to the years 2000 through 2050. Alzheimer Dis Assoc Disord 15(4):169–173.
Hebert LE et al (2013) Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology 80 (19):1778–1783.
Hua X et al (2009) Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 48(4):668–681.
Hung W-L et al (2011) Suppressed fuzzy-soft learning vector quantization for MRI segmentation. Artif Intell Med 52(1):33–43.
Hwang S-K, Kim W-Y (2006) A novel approach to the fast computation of Zernike moments. Pattern Recogn 39 (11):2065–2076.
Hyman BT et al (2012) National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimer's Dementia 8(1):1–13.
Illán I et al (2011) Computer aided diagnosis of Alzheimer's disease using component based SVM. Appl Soft Comput 11 (2):2376–2382.
Itchhaporia D et al (1996) Artificial neural networks: current status in cardiovascular medicine. J Am Coll Cardiol 28(2):515–521.
Jack CR et al (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 27(4):685–691.
Jenkinson M, et al. (2005). BET2: MR-based estimation of brain, skull and scalp surfaces. In: Eleventh annual meeting of the organization for human brain mapping.
Kara S, Güven A (2007) Training a learning vector quantization network using the pattern electroretinography signals. Comput Biol Med 37(1):77–82.
Kaye JA et al (1997) Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. Neurology 48(5):1297–1304.
Kim S et al (2011) Voxel-based morphometric study of brain volume changes in patients with Alzheimer's disease assessed according to the Clinical Dementia Rating score. J Clin Neurosci 18 (7):916–921.
Kohonen T, Maps S-O (1995). Springer series in information sciences. Self-organizing maps 30.
Lehmann M et al (2011) Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. Neurobiol Aging 32(8):1466–1476.
Lerch JP et al (2008) Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging 29(1):23–30.
Li H (2008). Classifying Atomicity Violation Warnings Using Machine Learning, ProQuest.
Li S et al (2009) Complex Zernike moments features for shape-based image retrieval. Syst Man Cybernetics, Part A: Syst Hum, IEEE Trans on 39(1):227–237.
Li J et al (2012) A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer's disease. Neurosci Biobehav Rev 36(2):757–763.
Liyun W et al (2009) Spermatogonium image recognition using Zernike moments. Comput Methods Programs Biomed 95 (1):10–22.
López MM et al (2009) SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. Neurosci Lett 464(3):233–238.
Magnin B et al (2009) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51(2):73–83.
Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11(2):431–441.
Martín-Valdivia MT et al (2007) The learning vector quantization algorithm applied to automatic text classification tasks. Neural Net 20(6):748–756.
McDonald C et al (2009) Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. Neurology 73 (6):457–465.
McKhann GM et al (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dementia 7(3):263–269.
Misra C et al (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. NeuroImage 44 (4):1415–1422.
Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Net 6(4):525–533.

Ortiz A et al (2013) LVQ-SVM Based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. Pattern Recogn Lett 34(14):1725–1733.
Ota K et al (2014) A comparison of three brain atlases for MCI prediction. J Neurosci Methods 221:139–150.
Petersen RC et al (1999) Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 56(3):303–308.
Riedmiller M, Braun H (1992). RPROP-A fast adaptive learning algorithm. Proc. of ISCIS VII, Universitat, Citeseer.
Rumelhart DE, et al. (1988). Learning representations by back-propagating errors. Cognitive modeling.
Rusinek H et al (2004) Atrophy rate in medial temporal lobe during progression of Alzheimer disease. Neurology 63 (12):2354–2359.
Sartakhti JS et al (2012) Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). Comput Methods Programs Biomed 108 (2):570–579.
Shen D et al (2014) Machine Learning Techniques for AD/MCI Diagnosis and Prognosis. Machine Learning in Healthcare : Springer. pp. 147–179.
Shin J et al (2010) Voxel-based analysis of Alzheimer's disease PET imaging using a triplet of radiotracers: PIB, FDDNP, and FDG. Neuroimage 52(2):488–496.
Sperling RA et al (2011) Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dementia 7(3):280–292.
Suk H-I et al (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 101:569–582.
Tahmasbi A et al (2011) Classification of benign and malignant masses based on Zernike moments. Comput Biol Med 41 (8):726–735.
Tapiola T et al (2008) MRI of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study. Neurobiol Aging 29 (1):31–38.
Toyama H et al (2005) PET imaging of brain with the β-amyloid probe,[11C] 6-OH-BTA-1, in a transgenic mouse model of Alzheimer's disease. Eur J Nucl Med Mol Imaging 32(5):593–600.
Tripoliti EE et al (2011) A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. Artif Intell Med 53(1):35–45.
Vieira SM et al (2013) Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. Appl Soft Comput 13(8):3494–3504.
Wang Y, et al. (2013). Kernel-based multi-task joint sparse classification for Alzheimer's disease. In: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, IEEE.
Yang S-T, et al. (2013). Discrimination between Alzheimer's Disease and Mild Cognitive Impairment Using SOM and PSO-SVM. Computational and mathematical methods in medicine 2013.
Zernike VF (1934) Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. Physica 1 (7):689–704.
Zhang D et al (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55(3):856–867.

(Accepted 5 August 2015)
(Available online 8 August 2015)