Computer aided diagnosis of Alzheimer's disease by automatically obtaining the best coronal slices for multi-classification recognition

Olga Valenzuela[1], Antonio Carrillo[1], Daniel Castillo[1], Juan Manuel Galvez[1] and Ignacio Rojas[1]

Abstract—The goal of this contribution is to find out a set of Y slices (coronal slices) from MRIs of patients with Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Normal images, that provides the maximum accuracy in a multiclass classification system. Images are preprocessed and 2D wavelet coefficients are extracted to form a feature matrix. Using a feature selection algorithm called mRMR, the best features from the matrix are extracted; then, the dimension of the feature vectors is reduced using PCA and finally, it is used to train an SVM to perform multi-class classification. In order to find the best combinations of coronal slices, a multi-objective genetic optimization methodology based on NSGA-II is used and a set of different solutions are extracted from the Pareto front. More relevant solutions are selected using more flexible criteria than that of the Pareto front, and examine what slices and accuracies are achieved. The multi-classification accuracies obtained by the proposed method are 94.2% (the individual of the Pareto front with the highest accuracy) or 91.9% (using the best 13 slices according to their frequency of presence in the Pareto front). It is important to note that not only a good accuracy is obtained in the classification, but also new knowledge about the most relevant coronal slices to distinguish the four MRI classes (AD, Normal, MCI, LMCI).

Keywords— Alzheimer's disease, multiclass classification, feature selection, mild cognitive impairment, Multi-objective genetic algorithm optimization, coronal slices selection.

I. Introduction

There are several possible causes for dementia, but Alzheimer's Disease (AD) is leading cause of dementia in the world. There are around 35.6 million people in the world with dementia and more than 18 million of them have Alzheimer's disease, which represents more than 50% of the total people with dementia. Besides, the number of people with dementia is expected to increase to 65.7 million in 2030 and 115.4 million in 2050. Today, the diagnosis of Alzheimer's disease is made by using clinical criteria; however, these criteria are not capable of diagnosing the disease in its pre-clinical stage, not allowing for an early diagnosis.

Given this situation, it is necessary to develop methods and techniques to be included in the criteria that provide an early diagnosis, which would allow people with dementia to plan ahead while they still have the capacity to make important decisions about their future care.

1: Information and Communications Technology Centre (CITIC-UGR). Univeristy of Granada, Spain.

Besides,it would also allow them to access available drug and non-drug therapies that may improve their cognition and enhance their quality of life.

At this point, it is necessary to identify the condition prior to dementia which is Mild Cognitive Impairment (MCI). Subjects with MCI are in an intermediate clinical situation between normality and dementia, which is characterized by the presence of subjective cognitive complaints but which do not significantly affect their daily activities. As there are studies that show that between the 10% and 15% of patients with MCI have developed dementia within a year, it is really important to be able to identify this pathology. Thus, it is in this field where imaging techniques can play a key role in early detection of patients who can develop dementia, in differential diagnosis of distinct dementias and in monitoring the progression of the disease.

It is important to note that neuroimaging, using mainly magnetic resonance, is a powerful tool that adds a positive predictive value to the diagnosis and includes measurements using structural MRI to evaluate medial temporal lobe atrophy and positron emission tomography using fluorodeoxyglucose (FDG) or amyloid markers [6]. Many studies have focused on quantifying focal atrophy in the temporal lobe [9] and [11] and there are even visual scales to quantify the degree of atrophy, which are quick and easy to use. Recently, validations of computerized methods to measure the degree of temporal atrophy have been published. In comparison, these methods have a similar discriminatory power [19] with the advantage that they would facilitate measurements and would provide more objective results by standardizing the methods of analysis [10].

There are new developments in automatic classification systems based on computer intelligent paradigms, such as support vector machine (SVM), that present new diagnosis tools based on T1-weighted MRI [18],[12],[2], and [4]. These approaches can be divided into three different categories, taking into account the type of features extracted from the MRI (voxel-based, vertex-based, or ROI-based) [6]. For the voxel-based category, the features are the probability of the different tissue classes in a given voxel [13]. In the second category, the vertex-based one, the features are defined at the vertex-level on the cortical surface [16]. The methods based on ROI include mainly the analysis of the hippocampus (study of the volume, shape, and specific characteristic of the hippocampus) [5].

Finally, it is important to mention some contributions that jointly analyze different sources of information databases. In the contribution presented by [8], the main goal is to analyze

classification accuracies for detection of AD using FDG-PET, MRI, or combined information from both imaging modes in two independent sets of patients (ADNI database and Leipzig cohort). In this study, the authors randomly extracted FDG-PET and MRI data from 28 AD patients and 28 healthy control subjects from the database provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and compared them to data of 21 patients with AD and 13 control subjects from the Leipzig cohort. The authors have obtained high accuracies for discrimination between AD patients and control subjects in both cohorts using only FDG-PET or combined information from both imaging modalities. An interesting result is that using only MRI, the accuracies were substantially lower.

Even though these approaches achieve high accuracy (over 85%), they were calculated on different studied populations (in fact, the variability between evaluations statistically increases mainly in the case that the number of subjects is low; typically, there are several approaches where less than a hundred subjects are used) making it difficult to compare the obtained results. It is important to note that most of the articles published to date using intelligent classification systems in Alzheimer's disease diagnosis have three major drawbacks:

- They use a low number of data for both training and test (in some cases, less than a hundred).

- Classification is performed between patients with dementia and healthy patients, but it does not make any difference between patients with Mild Cognitive Impairment (MCI), Middle Cognitive Impairment Converters (MCI-C), and Middle Cognitive Impairment Non Converters (MCI-NC), since it is a complex task.

- They only use one type of feature (voxel-based, vertex-based, or ROI-based), but not all of them together.

II. Materials and methodology

A. Subjects cohort

The diagnostic classification was managed using a set of 255 subjects from the ADNI database and grouping them as AD (Alzheimer's disease), MCI (Mild cognitive impairment), Late Mild Cognitive Impairment (LMCI), and Normal.

The AD group contained 68 subjects with ages ranging from 62 to 89 (74.3 $\pm$ 7.6) years. The Mini Mental State Examination (MMSE) and Clinical Dementia Ratio (CDR) scores are mean 23.08 $\pm$ 2.49 and mean 0.72 $\pm$ 0.30, respectively. It is important to remember that a zero CDR represents no dementia, and CDR of 0.5, 1, and 2 indicate very mild, mild, and moderate dementia respectively. The normal group has a total of 69 subjects, the MCI group has 79 subjects, and finally, the LMCI group has 49 subjects.

B. Segmentation and normalization

The images of the subjects are firstly segmented using the Segmentation routine implemented in SPM12. Afterwards, they are normalized to MNI space, using 1x1x1 mm voxels and a Bounding Box of limits: {[}-78 -112 -60; 78 76 85{]}.

Since several articles report that the use of Gray matter achieves better results than the use of White matter or the whole brain, and in order to reduce the continue the experiment using just that Gray matter.

C. Feature extraction, selection, and reduction

Prior to feature extraction, slices used are intensity-normalized and filtered with a Gaussian filter with sigma=0.5.

In order to extract the features from the MRI images, wavelet analysis will be used. In wavelet analysis, a fully scalable modulated window is shifted along the signal for every position, and this process is repeated with shorter or longer windows, yielding a multi-resolution analysis. As a result, when this process is applied to an image, decomposing the original image in its wavelet coefficients (calculated as explained below) will produce a series of images with different scales.

The 1D-DWT can be applied to each dimension of the input data [14]. In the case of 1D input data, each filter yields one sub-band, low-filter outputs the low-frequency sub-band L, and high filter outputs the high-frequency sub-band. There is a consensus in the literature regarding this topic, Db-4 wavelet function up to level 2 has the best trade-off between efficiency and computational cost, and so, the 2D Wavelet coefficients are obtained using Debuchie function of order 4 up to level 2 for each slice. Feature selection is achieved using the Minimum Redundancy Maximum Relevance algorithm by [15]. For discrete variables, mutual information I of two variables x and y measures the level of similarity between them and is defined according to their joint probabilistic function p(x,y) and their respective marginal probabilities $p(x)$ and $p(y)$:

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{1}$$

Minimum redundancy intends to select the features that are mutually maximally dissimilar, making the feature set more representative. This minimum redundancy condition is:

$$\min W_I, \quad W_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(i, j) \tag{2}$$

where |S| denotes the number of features in S, the feature subset we are seeking from the feature set, and i,j are two different features. Maximum relevance condition is met when total relevance of all features is maximized. Relevance measures the level of discriminant power of features according to their difference between target classes. Mutual information is again used between target classes h={h1, h2, … hk}and features i. In this way, I(h,i) quantifies the relevance of feature i for the classification task. Thus, maximum relevance means maximizing the total relevance of all features in S:

$$\max V_I, \quad V_I = \frac{1}{|S|} \sum_{i \in S} I(h, i) \tag{3}$$

The mRMR feature set is obtained optimizing both conditions minWI and VI. This optimization requires combining them into a single criterion function, two of the simplest combination criteria being:

$$MID: \max(V_I - W_I) \quad MIQ: \max(V_I / W_I) \quad (4)$$

The 126 most important features are extracted according to the algorithm using MIQ criteria for minimization. To reduce the feature dimensionality, Principal Components Analysis is used. 95% of the information given by the original set of selected features is selected to create a new feature space with a lower dimensionality.

D. *Classification*

For classification purposes, a Support Vector Machine (SVM) is used [3]. In order to train and test an SVM model, it is first trained with a training set for which features and target labels are known. Hyperplane parameters are fit so that the model achieves its best accuracy separating the different data points, and then, a testing set for which labels are known is fed into the SVM model and classified, getting a vector of labels as output. This output labels are compared to the actual target labels of the input feature vectors and the SVM model accuracy is obtained.

Mathematically, the construction of the hyperplanes and classification depends on the specific SVM used and its kernel. Given a training set of feature vectors and label pairs x_i, y_i where $i = 1, \dots L$; x in R^n and y is in the interval $\{1, -1\}$, SVM requires the solution of the optimization problem:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{L} \xi_i \quad (5)$$

Subject to:

$$y_i(\mathbf{w}^T \phi(\mathbf{x}_i + b) \geq 1 - \xi_i, \; \xi_i \geq 0 \quad (6)$$

where w is the normal vector of a given hyperplane (w·x+b=0 defines a hyperplane), vectors xi are mapped to a higher dimensional space by the function Φ, and C is the penalty parameter that specifies the cost of misclassifying a data vector (i.e. the error term). Radial Basis Function kernel is used in this contribution.

The selected features are then fed into a Support Vector Machine using 10-fold Cross Validation and a training set of 80% of the whole set. The 10-fold Cross Validation allows us to calculate the best C and Gamma parameters to use, selecting those which obtained the best classification accuracy. On each run, all the subjects conforming the study are shuffled to ensure unbias, so that the model is not overfit. Furthermore, three independent classifications are performed and results are averaged to reduce variance. Ideally, a larger set of subjects would be needed or more classification runs should be performed.

E. *Optimization*

In order to find the best set of coronal slices to perform the classification, a Multi-Objective Genetic Algorithm based on NSGA-II is used in [7], creating a population of individuals consisting on 161 bit strings, each bit coding whether a specific slice should be selected or not. In this contribution, slices from Y=20 to Y=180 are only tested. The two fitness functions that need to be optimized by the algorithm (minimized) are: f(1)=100-accuracy(%); f(2)=nslices.

A population of 100 individuals is used, and the optimization is run for 200 generations. Each individual of each generation was stored in a data-file along with the accuracy achieved, so that a post-optimization analysis could be performed.
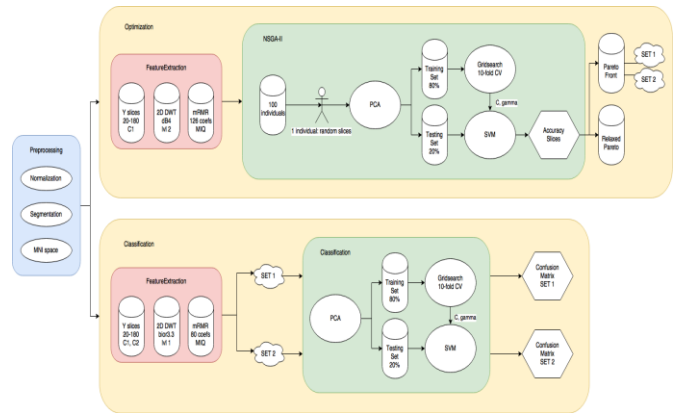

Fig. 1 Full pipeline of the proposed method

Full pipeline of the proposed method is shown in Figure 1. It is important to differentiate between the optimization process and the classification process. The first step is the optimization and selection of the best slices, which is performed by the multi-objective genetic algorithm. The objectives of this step are related to the accuracy (measured by a classifier based on SVM) and the minimization of information required by the system (measured as the number of slices that are necessary to be used by the classifier). The second process, the classification process, is executed once the first optimization process is finished. Using the best coronal slices for multi-classification, an SVM-based classifier is trained and tested.

III. **Results**

The most usual representation of the results obtained by a GA algorithm is that of a Pareto front, shown in Figure 2. This graph shows the dominant individuals.
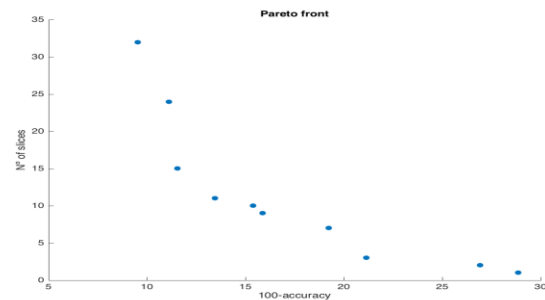

Fig. 2 Pareto front. This graph shows the trade-off between accuracy and number of slices used

The histograms of slices have been obtained, as shown in Figure 3. Observing the histograms, it is possible to get an idea of the set of best coronal slices to classify subjects. Since the genes of the individuals that are outperformed by others tend to disappear with further generations, the dominant individuals, which have dominant genes, exhibit a similar genome, so that the frequency of each slice in the Pareto front is somewhat an approximation of the relevance of that slice when extracting features for classification. The ranking of slices obtained from the Pareto front is shown in Table1
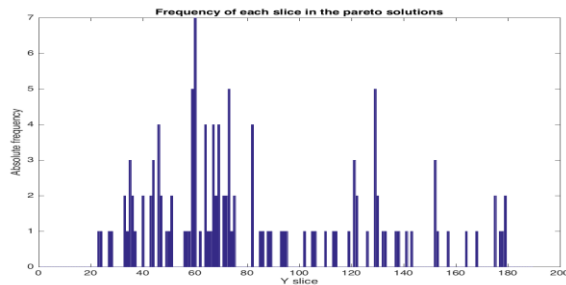

Fig. 3 Histogram for the Pareto front. There are several dominant slices: 60, 59, 73, 129...

Frequency	Slices
7	60
5	59, 73, 129
4	46, 64, 67, 69, 82
3	35, 44, 121, 152
2	33, 36, 40, 43, 47, 51, 68, 71, 72, 75, 122, 130, 175, 179
1	23, 24, 27, 28, 34, 37, 49, 50, 56, 57, 58, 62, 65, 66, 74, 85, 86, 88, 89, 93, 94, 95, 102, 105, 106, 110, 113, 114, 119, 126, 132, 133, 137, 138, 141, 143, 153, 157, 164, 168, 177, 178

TABLE I. Frequency of slices in the Pareto front.

IV. **Consistency of ROIs detected**

In order to contrast the results obtained in this contribution, comparisons with those of other experts and researches presented in the bibliography have been carried out. Figure 4 and 5 clearly show consistency among different studies and researches. Specifically, the slices from the Pareto front that appear on at least 3 of the individuals conforming the Pareto have been selected, which happens to be 13 slices (sorted by relevance: 60, 129, 73, 59, 82, 69, 67, 64, 46, 152, 121, 44, 35).

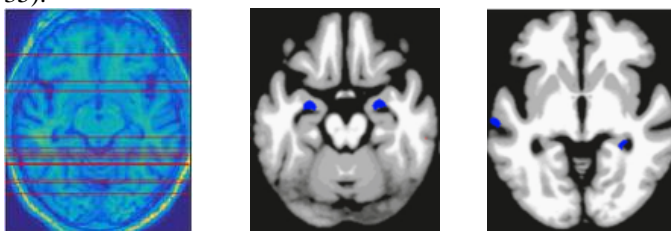Fig. 4 Left: The best coronal slices. Right: ROIs used by Dukart et al. 2012.

V. **Performance comparison**

Some important measures of the performance of a classifier are the Sensitivity and the Specificity, which are defined for a two class classification problem as: SEN=TP/(TP+FN) and SPE=TN/(TN+FP), where TP, TN, FP,

and FN are true positives, true negatives, false positives, and false negatives, respectively. Because of the nature of the proposed method, which is that of a multi-class classifier, these definitions are not met, and instead, confusion matrices are used. A confusion matrix is a graphic representation of the performance of a multi-class problem where rows represent the output class given by the classifier and columns represent the target class, ie. the actual class.
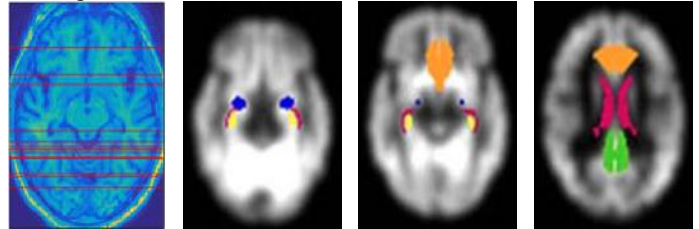

Fig. 5 Left: The best coronal slices. Right: ROIs used by Aggarwal et al. 2015 [1]. Note, especially in rows two and three, the normalizations are different, but ROIs used by Aggarwal are very close to the slices found in our experiment.

To measure the performance, two sets of ROIs have been tested: SET1 involves the slices conforming the individual of the Pareto front with the highest accuracy whereas SET2 involves the top 13 slices according to their frequency of presence in the Pareto front. For feature extraction, the 2D wavelet coefficients of gray and white matter for each coronal slice have been calculated, using the transform bior3.3 up to level 1. For feature selection, the top 80 mRMR features for each slice have been selected, together with the use of PCA (kept 95% of the variance in the data is used). This gives us vector features for SET1 and SET2 of 228 and 219 features respectively. For classification, SVM with RBF kernel is computed. Summary of results is shown in Table II and confusion matrices for both sets of ROIs are shown in Figure 7 for SET1 and Figure 8 for SET2. Note that each classification was performed five times in order to unbias the classification accuracy and results were merged so that they could be later on averaged.

VI. **Conclusions**

In this paper, a new procedure to find the most significant ROIs when determining mental state in the development of Alzheimer's Disease have been described. In order to find the best set of coronal slices to perform the classification, a Multi-Objective Genetic Algorithm based on NSGA-II is used. In this way, the training time required and the necessary computing time in order to find the best ROIs is significant. Prior researches and papers show good methods which achieve high accuracies, but the literature lacked high classification accuracy when multi-class classification is performed. In most researches, subjects are classified class vs class. It would be an interesting development to design a method in which, given an MRI input, the software determined the class to which it belongs and the most relevant coronal slices for this purpose. This could help medical experts in the diagnosis of different states in the development of AD. Usually, ROIs determined by medical experts are used as the source of features. This contribution faces these issues and tries to investigate this computationally, so that maybe-hidden ROIs which experts

had not noticed could be found. From the experiment carried out in this report, it is possible to assert that the obtained ROIs in our experiment are consistent with the existing bibliography. The proposed method (without the requirement of prior knowledge from experts) is scalable and reproducible to further analyse any other neurological disease that affects gray or white matter; hence, the same procedure could be suitable to find relevant ROIs in other diseases.

Fig. 6 SET1 confusion matrix, individual in the Pareto front with the highest accuracy

Fig. 7 SET2 confusion matrix, top 13 slices from the Pareto frequency ranking.

Authors	Imaging	Cohort	Classification	Method	Acc (%)	Sen (%)	Spe (%)
Aggarwal et al., 2015 [1]	MRI	OASIS	99 AD vs 99 Normal	3DWT - FDR - mRMR - SVM	68.94	81.13	75.02
Beheshti et al., 2015 [2]	MRI	ADNI	130 AD vs 130 Normal	VBM,PCA,SVM	90.76	90	91.53
			68 AD vs 68 Normal		78.23	77.34	79.11
Dukart et al., 2013 [8]	MRI PET	ADNI	28 AD vs 28 Normal	VOIs - SVM	85.7	89.3	82.1
Yudong Zhang et al., 2015 [20]	MRI	OASIS	28 AD vs 98 Normal	ICV-Eigenbrain-WTT-SVM	91.47	90.17	91.84
Yudong Zhang et al., 2015 [20]	MRI	OASIS	28 AD vs 98 Normal	DF-PCA-TSVM	92.75	90.56	93.37
Tong Tong et al. 2017 [17]	MRI and PET	ADNI	37 AD vs 35 Normal	Nonlinear graph fusion	91.8	88.9	94.7
			75 MCI vs 35 Normal		79.5	85.1	67.1
			37 AD vs 75 MCI vs 35 Normal		60.2	-	-
Xiaofeng Zhu et al. 2016 [22]	MRI	ADNI	51 AD vs 52 Normal vs 99 MCI	PCA,LDA,LPP,SVM	68.31	-	-
			51 AD vs 52 Normal vs 43 MCIc vs 56 MCInc		59.74	-	-
Proposed method	MRI	ADNI	58 AD vs 49 LMCI vs 79 MCI vs 69 Normal	DWT-mRMR-PCA-SVM	94.2	Figure 6	
Proposed method	MRI	ADNI	58 AD vs 49 LMCI vs 79 MCI vs 69 Normal	DWT-mRMR-PCA-SVM	91.9	Figure 7	

TABLE II. For those authors who used SVM among different classifiers, the results of their SVM approaches are shown. If different configurations were carried out for the experiment, the most similar ones to ours or those that

achieved the best accuracies are shown

Acknowledgment

References

[1] N. Aggarwal, et al. "3D discrete wavelet transformfor computer aided diagnosis of Alzheimer's disease using t1-weighted brainMRI. International Journal of Imaging Systems and Technology, 25(2):179–190, 2015.

[2] I Beheshti and H Demirel. Probability distribution function-based classificationof structural MRI for the detection of Alzheimer's disease. Computers inbiology and medicine, 64:208–16, 2015.

[3] Bernhard E. Boser, et al. "A TrainingAlgorithm for Optimal Margin Classifiers". Fifth Annual ACM Computational Learning Theory, 1992.

[4] Edward Challis, et al. "Gaussian process classification of Alzheimer's disease andmild cognitive impairment from resting-state fMRI". NeuroImage, 112:232–243,2015.

[5] M. Chupin, et al. Neuroimaging Initiative.

[6] Rémi Cuingnet, et al. "Automatic classification of patients with Alzheimer's diseasefrom structural MRI: A comparison of ten methods using the ADNI database"NeuroImage, 56(2):766–781, 2011.

[7] Kalyanmoy Deb. Multi-Objective Optimization Using Evolutionary Algorithms,2001.

[8] Juergen Dukart, et al. "Meta-analysis based SVM classificationenables accurate detection of Alzheimer's disease across different clinicalcenters using FDG-PET and MRI. Psychiatry Research - Neuroimaging, 2013.

[9] G B Frisoni, et al. "Neuroimaging toolsto rate regional atrophy, subcortical cerebrovascular disease, and regional cerebralblood flow and metabolism: consensus paper of the EADC". Journal of neurology, neurosurgery, and psychiatry, 74(10):1371–81, 2003.

[10] CR Jr Jack, et al. "Steps to standardization and validation of hippocampalvolumetry as a biomarker in clinical trials and diagnostic criterionfor Alzheimer's disease" . Alzheimers Dement., 4(7):474–485, 2011.

[11] EL Koedam, et al. "Visual assessment of posterior atrophydevelopment of a MRI rating scale". Eur Radiol., 21:2618–2625, 2011.

[12] Lin Li, et al. Automated detection of mildcognitive impairment through MRI data analysis . Int. Journal On Artificial Intelligence Tools, 21(5), 2012.

[13] B. Magnin, L. et al. Support vector machinebasedclassification of Alzheimer's disease from whole-brain anatomical MRI.Neuroradiology, 2009.

[14] Loris Nanni et al."Wavelet decomposition tree selectionfor palm and face authentication". Pattern Recognition Letters, 29(3):343–353,2008.

[15] H C Peng, et al. "Feature selection based on mutual information:Criteria of max-dependency, max-relevance, and min-redundancy". IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8):1226–1238,2005.

[16] Olivier Querbes, et al. "Early diagnosis of Alzheimer's disease using cortical thickness: impactof cognitive reserve". Brain, 2009.

[17] Tong Tong, et al. "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion".Pattern Recognition, 63:171–181, 2017.

[18] Rik Vandenberghe, et al. "Binary classificationof F-18-flutemetamol PET using machine learning: Comparison with visual reads and structural MRI" Neuroimage, 64:517–525, 2013.

[19] Eric Westman, et al. "Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease". PLoS ONE,6(7), 2011.

[20] Yudong Zhang, et al. "Detection of subjects and brain regions relatedto Alzheimer's disease using 3D MRI scans based on eigenbrain and machinelearning". Frontiers in Computational Neuroscience, 2015.

[21] Yudong Zhang, et al. "Detection of Alzheimer's disease and mild cognitiveimpairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC". Biomedical Signal Processing and Control,21:58–73, 2015.

[22] Xiaofeng Zhu, et al. "SubspaceRegularized Sparse Multitask Learning for Multiclass Neurodegenerative Disease Identification". IEEE Transactions on Biomedical Engineering, 63(3):607–618, 2016.32