Peer-reviewed version available at Sensors 2020, 20, 941; doi:10.3390/s20030941

Emphasis Learning (EL), Feature Repetition in Width Instead of Length, to Obtain Better Classification Performance; Case Study: Alzheimer's Disease Diagnosis

H. Akramifard [1], M. A. Balafar [1], S. N. Razavi [1], Abd Rahman Ramli [2]

[1.] Faculty of Electrical and Computer Engineering, University of Tabriz, East Azerbaijan, Tabriz,Iran
[2.] Faculty of Engineering, University Putra Malaysia, Selangor, Malaysia

Abstract: A method for classification is introduced in this article, and it is tested on ADNI database to diagnose alzheimer's disease (AD). It is obvious that tunning the performance of a classification to get better results is a complicated problem, and when we want model's accuracy or other peformance measurments higher than 90%, the problem will be more complicated. In this study, we tried and succeeded to discover a method to solve this problem. The final feature set can be used clustering too, because outgrowth feature set of the proposed method is invigorated. In the recent years, a lot of activities is done to develop computer aided systems (CAD) for alzheimer's disease diagnosis. Most of these recently developed systems concenterated on extracting and combining features from MRI, PET, CSF, and …; in this article, we made attempt to do so and utilized one more technique to increase classification performance. Finding and producing the best features to solve three binary classification problems of AD vs. Normal Control (NC), Mild Cognitive Impairment (MCI) vs. NC, and MCI vs. AD are the purposes of this article. Experiments indicate performance and effectiveness rates of the proposed method, which are accuracies of 98.81%, 81.61%, and 81.40% for AD vs. NC, MCI vs. NC, and AD vs. MCI classification problems, respectively. As can be seen, using this method increased the performance of the three binary problems incredibly.

Keywords: Alzheimer's disease; Emphasis Learning; Multi-modal classification; SVM; PCA.

Introduction

Alzheimer's disease (AD) could be described by cognitive and memory dysfunctions. It is the major cause of dementia in elderly, now and in the future. AD is one of the main causes of death in the United States (Thies & Bleiler, 2012). Early diagnosis and prognosis of AD are important because of limitation of treatment time. To understand the mechanisms that cause and identify biomarkers for diagnosis of AD, scientists analyzed Magnetic Resonance Imaging (MRI) (Wee et al., 2011), Positron Emission Tomography (PET) (Nordberg, Rinne, Kadir, & Langstrom, 2010), Cerebro-Spinal Fluid (CSF) (Nettiksimmons et al., 2010), etc.

Visual information of tissues atrophy were obtained from structural MRI (results of cellular changes of AD). Cerebral glucose metabolism of functional activity of brain were obtained from PET images (Nordberg et al., 2010).

These brain observing techniques, using machine learning can provide nice tools to diagnosis and overcome brain's dysfunction problems. These combined techniques can use different modalities including MRI, PET, and other neurological data to diagnose AD/Mild Cognitive Impairment (MCI) patients from healthy people (NC) (Garyfallou et al., 2017; Islam et al., 2018; Maqsood et al., 2019; Toro & Gonzalo Martin, 2018).

We can be sure that there exist unsight features among analyzed data in this area, that can help us for diagnosis and prognosis of AD/MCI. We use PCA for reduction dimension of features of data. PCA can efficiently be used to extract more efficient features of data and enhance classification accuracy. Feature representation using PCA will be reduced through processing resources, in addition to the enhancement of the classification accuracy. The steps of this work can be summarized as follows:

1) Feature extraction from MRI and PET images and other data sources (from ADNI dataset).
2) Concatenation of all the features.
2) Preparation of data sets and refining the data.
4) Dimension reduction using PCA.
5) Repeating data in vectors came from step 4 to achieve highest classification performance.

In the next section of this article, the data extraction and data source are presented. After that preprocessing, feature reduction and increment, and classification methods will be explained. In next section, the experimental results are provided. The discussion on the results and conclusions are presented in the final sections.

Materials and Methods

In this method, we emphasized on the strongest and the most influential features. The main idea of this article is rooted in the fact that when a person's good and outstanding features are emphasized, those features along with the person himself improve and his performance gets better. In other words, these features result in the improvement of themselves; and the better and more precise these features, the more effective they can be. To put this another way, it can be said that, either when good features with low quantities repeat for several times, or when such positive features repeat few times, they can result in learning. For instance, which of the two following ways result in higher rates of learning? When a teacher solves a problem with two solutions and repeating that many times? Or when he uses many solutions while repeating them just a few times? The answer is "both"; however, for some problems the first way and for some, the second way works better. This works exactly as computer, and deeper and broader learning takes place when there is the possibility of repeating both approaches.

It is obvious that with increasing the precision of tuning, trying to increase the precision more would be a challenge. For example, usually, increasing the precision of tuning from 80 to 85 percent is less complex and needs less cost and computational burden comparing to increasing it from 95 to 96 percent (if possible), usually. In this study, a method is put forward in order to make this possible and to reduce the cost and computational burden to a great extent. However, when the computed precision using the main data is low, or utilizing the extracted features from the dimension reduction does not cause a change in the precision of the model (in fact, there has been no outstanding feature achieved), repeating these features may not be that much influential in the precision of tuning.

Characteristic of subjects

We only used baseline of MRI, PET images, and CSF data acquired from 156 AD patients, 338 MCI patients and 211 NC subjects from ADNI dataset. Table 1 shows demographic information of patients. All the data were acquired in May 2017. Mini-Mental State Examination (MMSE) scores are added to the extracted data from ADNI database. MMSE scores have ranges that are defined as follows: 1) scores between 24 and 30 represent healthy people; 2) scores between 20 and 24 represent MCI subjects; 3) scores between 13 and 20 represent moderate dementia subjects.

Table 1 Summary of Demographic data of patients and subjects.

	Count	Male	Female	Married	Widowed	Divorced	Never Married	Average of AGE	Average of MMSE
AD	156	76	80	127	18	8	3	74.89	23.32
NC	211	110	101	142	38	17	14	75.91	29.13
MCI	338	215	123	269	39	24	6	74.51	27.05
Total	**705**	**401**	**304**	**538**	**95**	**49**	**23**	**75.01**	**26.85**

MRI and PET Images and CSF Data

The MRI images were in Neuroimaging Informatics Technology Initiative (NIfTI) format. These MRI images were pre-processed for spatial distortion correction. Collection of CSF data was done in the morning right after an overnight using of 20-or 24-gauge spinal needle. The FDG-PET images are average values obtained from 30–60 minute post-injection, their voxel size is converted to standard, and smoothed to 8 mm full width at half of maximum resolution.

MRI Acquisition Parameters

In the multiple ADNI sites, multiple machines as Siemens, Philips, and GE Medical scanners are used. Standard protocol developed to evaluate 3D T1-weighted sequences for morphometric analyses (La Foresta et al., 2019). Structural brain MRI scans were acquired using 1.5 and 3 T MRI scanners. Most of the 1.5T MRIs were obtained from GE Medical scanners, and most of the 3T MRIs were acquired from Siemens machines.

In 1.5T protocol each subject experienced 2*1.5T T1-weighted MRI by 3D sagittal volumetric magnetization prepared rapid gradient echo (MP-RAGE) sequence. The repetition time of typical 1.5T acquisition are 2400ms, and the inversion time is 1000ms. Flipping and field of view are 8°, 24cm, respectively. Dimensionality of MRIs is 256 × 256 × 170, and the voxel size is 1.25 × 1.25 × 1.2 mm3.

For 3T scans, repetition time and inversion time are 2300ms and 900ms respectively. Flipping angle and field of view are 8°, 26cm. Dimensionality of MRIs is 256 × 256 × 170, with voxel size of 1.0 × 1.0 × 1.2 mm3.

For modern systems, scan time at 1.5T is 7.7min, and for 3T systems is 9.3min. This usually happens because of the difference between vulnerability artifacts, spin relaxation and chemical shift properties in 1.5T and 3T.

Preprocessing *of MR Images*

Spatial Parametric Mapping (SPM) software used for preprocessing (Ashburner, 2011). SPM used for realignment, smoothing and spatial normalization and feature extraction from ROI of MRIs. The reprocessing steps using VBM81 tools are as below:

1) Checking format of the images for being in a suitable condition using SPM tools.
2) Segmentation of the images, to identify gray matter (GM) and white matter (WM) and warp GM to the segmented image to Montreal Neurological Institute MNI space using the SPM tools.
3) Deformations estimation to best align the images to each other and create templates by registering the imported images with their average, iteratively using DARTEL tools of SPM.
4) Generation spatially normalised and smoothed GM images, normalised to MNI Space, using the estimated deformations by the DARTEL tools of SPM, generated the smoothed \modulated warped GM and WM images.

Data cleanizing and selection was done in the preprocessing step. In the second step (feature extraction) the input data is converted into small vectors (Duin, 2000). The classification algorithm determines that the vectors are more similar to Mild Cognitive Impairment (MCI) or to AD patient, or to Normal Control (NC) subject.

[1] http://www.neuro.uni-jena.de/vbm/download/

Feature Extraction

To extract the features of all the images, we adopted standard procedures of anterior commissure (AC)–posterior commissure (PC) correction, skull-stripping, and cerebellum removal for preprocessing and preparing. MIPAV software is used for AC-PC correction. FAST tools is used in FSL package (Y. Zhang et al., 2001) to segment structural MRI images into: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) images. To extract features of ROI of all the images VBM-SPM is used. After that, volumetric changes in specific regions such as entorhinal cortex, hippocampus, temporal and parietal lobes have been used. For each ROI, a mask has been made using WFU Pick Atlas[2] tools.

The PET images were aligned to the corresponding MR images strictly. Recognized areas from MRI Gray Matter tissue volume, and average voxel values, and average of PET voxel values, were used as features. In the literature these are used for AD/MCI diagnosis (Davatzikos et al., 2011);(Hinrichs et al., 2011);(D. Zhang & Shen, 2012);(F. Liu et al., 2013). Three CSF biomarkers of Ab42, t-tau, and p-tau were used for making features set, too. Therefore, 144 features were used to form the final feature set involving 132 MRI, 1 MMSE, 4 personal information, 3 CSF, and 4 PET images. At last, the vectors of extracted features were normalized by applying natural logarithm. K-fold cross validation method was used for testing and evaluation purposes.

Classification methods

For diagnosis and prognosis of AD and MCI, some classification algorithms are common and some algorithms are supporting. Among them, SVM and PCA are mostly in use.

Feature reduction method

One of the most common techniques for dimension reduction of data is Principal Component Analysis (PCA). PCA was introduced by Karl Pearson for the first time. It maps the data to a lower dimension, while maintaining the variance of the data. To use this method, the covariance matrix of the data and the eigenvectors of this matrix must be computed. The eigenvectors from the largest eigenvalues, that are the principal components, reconstruct the highest variance of the primary data. The first few eigenvectors often have the most information of the primary data.

Hence, we are left with a lesser number of eigenvectors, and there might have been some data loss in the process. But, the most important variances should be retained by the remaining eigenvectors. Figure 1 shows what eigenvectors of dataset are.

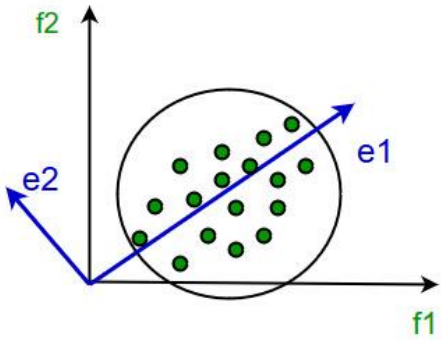


Figure 1 Schematic representation of eigenvectores of dataset.

[2] https://www.nitrc.org/projects/wfu_pickatlas/

Reduction in storage space and computation time and elimination of redundant features are advantages of using PCA, and reduction in some information of the original data, failing when mean and covariance are not enough to define data, and Indetermination of the number of principal components to keep information of data are disadvantages of using it.

Increasing Dimention of data to achieve better classification results

The main idea of this paper is repeating the most efficient features in classification. Theoriticaly and in the first look, increasing dimension of data by repeating features some times give us better and some times can make worse results in classification; but, what if we find and repeat good features of data to make our model for classification? Our experiment showes this theory works excelently! For diagnosis of Alzheimer's Disease, we tested this method after extracting best features for data set using PCA, dimension reduction, and repeating these new features. We repeated these features as input data for the classification algorithm (here SVM) until this action had no positive effect on classification performance or did not reduce it significantly. Figure 2 shows the proposed method diagram.

SVM

The one binary classification method that is used in medicine is support vector machine (SVM) (Fan et al., 2008; Kloppel et al., 2008; Mourao-Miranda et al., 2005). Classification efficiency of SVM in training high dimensional data has been proven. Moreover, SVM has been applied to voice activity detection, pattern recognition, classification and regression analysis (Burges, 1998; Shawe-Taylor & Cristianini, 2000). It is used to separate a set of training data with a hyperplane that is maximally distant from the two classes. SVM is the most common and efficient classifier in binary classification. Here, SVM was used to distinguish between AD and MCI patients and NC objects, pairwise.

Data Normalization vs. Data Standardization

Normalization maps values into a range of [0, 1] and it will be effective in the applications that require positive values. In this study, we used Normalization method. Equation (1) shows normalization fomula:

$$\boldsymbol{X}_{norm} = \frac{\boldsymbol{X} - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where X_{norm} is Normalized input data. The problem with normalization is outlier elimination.

on the other hand, Standardization maps original data to have a mean of 0. Equation (2) demonstrates standardization. In some cases, data standardization is recommended.

$$\boldsymbol{X}_{std} = \frac{\boldsymbol{X} - X_{mean}}{X_{dev}} \qquad (2)$$

Where $\mathbf{X}_{std}$ implies standardized data, X_{mean} is the mean of original data and X_{dev} is the standard deviation of data.

Evaluation criteria

Accuracy is a well-known evaluation measure for classification methods. Using accuracy, we compute the correctly classified samples and all samples' ratio. The other common evaluation metrics are sensitivity and specificity. Area Under Curve (AUC), and Receiver Operating Characteristic (ROC) are the other performance parameters for diagnosis procedures. Positive predictive value (PPV) and Negative predictive value (NPV) are widely used measures to describe

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

Peer-reviewed version available at Sensors 2020, 20, 941; doi:10.3390/s20030941

the performance of a classifier. The accuracy, sensitivity, specificity, PPV, and NPV are defined in the following equations, respectively:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Sen = \frac{TP}{TP+FN} \quad (4)$$

$$Spec = \frac{TN}{TN+FP} \quad (5)$$

$$PPV = \frac{TP}{TP+FP} \quad (6)$$

$$NPV = \frac{TN}{TN+FN} \quad (7)$$

Where, TP is all about the number of true positives (number of correctly classified as patients); TN is the number of true negatives (number of correctly classified as non-patients); FP stands for the number of false positives (number of non-patients classified as patients wrongly); FN is the number of false negatives (number of patients classified as non patients wrongly, too).

Two measures that were used to evaluate the rate of actual positives or negatives are "Specificity and sensitivity", e.g. the ratio of correctly classified AD or MCI patients or NC subjects and total subjects. These measures show the method detection power between AD, MCI and NC. Here, these metrics were measured using K-fold cross validation (with k = 10). Using this method, 10 selected sets of AD, MCI and NC will be sampled randomly. One set for testing and nine for training the classifier. This will be done for all the 10 sets, and the average of the evaluation parameters will be used to show the performance of classification method. In this article, we repeated K-fold method 100 times and average of averages was used to represent the performance of the method.

The steps of the algorithm of the method is as bellow:

1) Feature extraction
    a. **MF** <- MRI ROI Features
    b. **PF** <- PET ROI Features
    c. **CD** <- CSF data
    d. **PI** <- Personal information
    e. MS <- MMSE Score
2) **Concatenated Features** <- [**MF**, **PF**, **CD**, **PI**, MS]
3) **Imputaded data** <- Imputation(**Concatenated Features**)
4) **data** <- ln(**Imputaded data**) 'data normalization using Natural Logarithm'
5) **Normalized data** <- (**data-min(data)/(max(data)-min(data))**) 'data normalization'
6) **Reduced data** <- PCA(**Normalized data**)
7) **Data** <- [**Reduced data**]
8) **Diagnosis <-** SVM(**Data**) 'classificaion using SVM'
9) **Data** <- [**Data, Reduced data**]
10) Go to (8) until getting no better performance or significant decrease

Experimental results

Here, we evaluate our proposed method's efficiency. This is done for three binary classification problems of AD vs. NC, MCI vs. NC, and AD vs. MCI. 10-fold cross validation method is used for evaluation purposes. In the 10-fold cross validation, dataset is partitioned to 10 subsets, randomly, each included one tenth of the total dataset. Nine of subsets are used for training goals and the remaining one for testing. We have to do this for all the subsets and none of them must remain not tested.

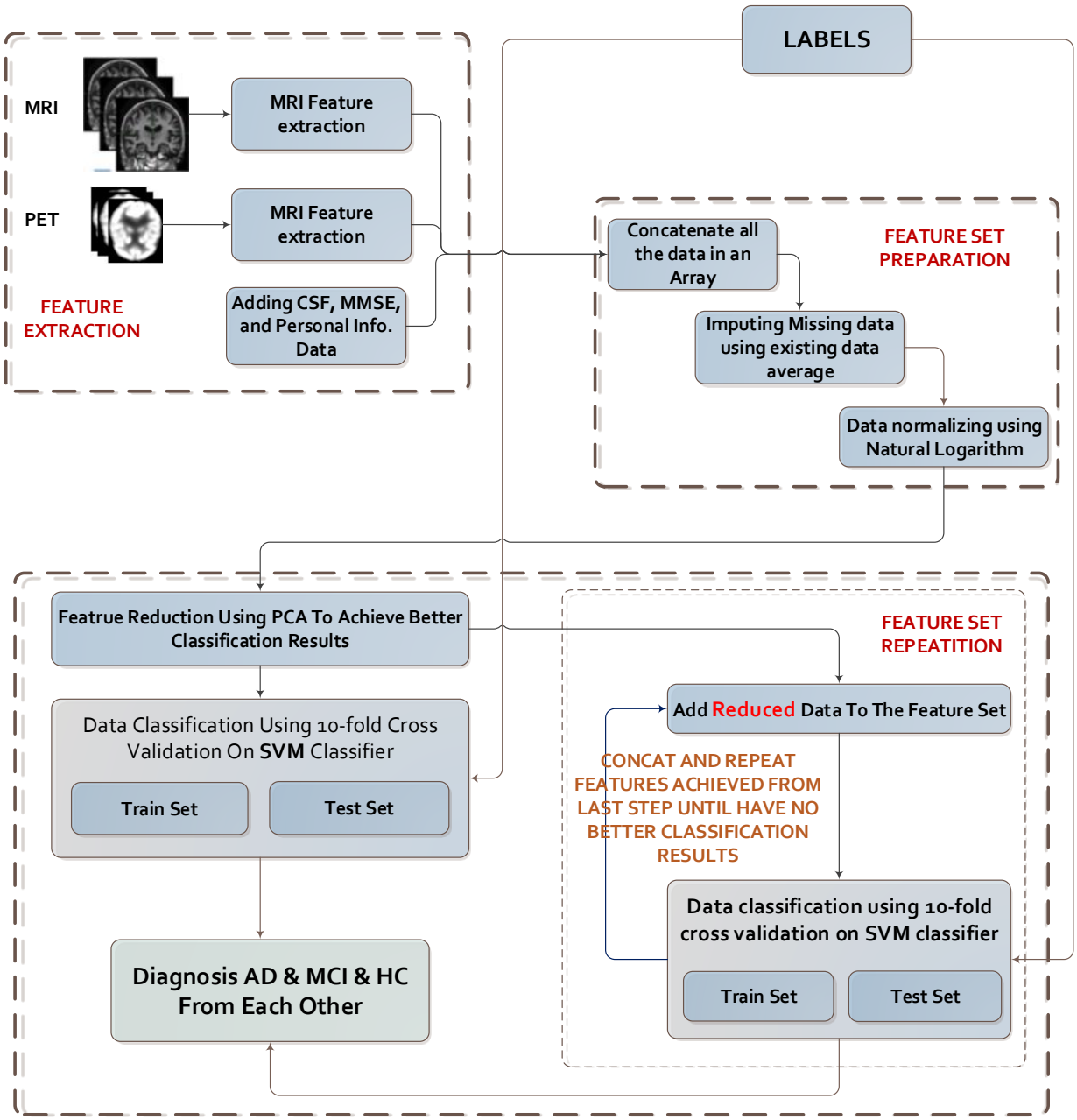


Figure 2 Diagram of steps of the proposed method (Emphasis Learning).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

Peer-reviewed version available at Sensors 2020, 20, 941; doi:10.3390/s20030941

Classification results

In order to represent the performance of the proposed method, we present the classification results obtained from the SVM classification algorithm, by 10-fold cross validation method. Table 2 shows the mean Accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, and Area Under Curve for different numbers of the repeated features tested on the proposed method on three binary classes. As can be seen, the proposed method by repeating features for 5 times showed the best accuracy rates of 98.81%, to classify NC and AD data firstly, then, NC and MCI data, and finally, MCI and AD, respectively.

Table 2 Comparing performance metrics. Classification accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and area under curve (AUC) for all features and 25 PCA elements.

Data	Classes	ACC	sensitivity	Specificity	PPV	NPV	AUC
All Data	AD-NC	95.54	93.74	98.32	98.84	91.09	0.9577
	AD-MCI	81.41	89.02	68.09	82.99	78.00	0.7835
	MCI-NC	79.41	67.48	92.37	90.56	72.34	0.7993
Reduced Data using PCA	AD-NC	97.20	95.46	**99.86**	**99.90**	93.53	0.9768
	AD-MCI	**81.61**	88.45	**69.02**	**84.03**	76.39	**0.7846**
	MCI-NC	79.45	67.20	**92.96**	**91.33**	71.97	0.8011
2 Time Repeated Reduced Data	AD-NC	98.03	97.18	99.26	99.47	96.09	0.9831
	AD-MCI	80.37	88.57	66.38	81.80	77.28	0.7766
	MCI-NC	79.94	68.49	91.64	89.31	74.03	0.7991
3 Time Repeated Reduced Data	AD-NC	98.61	98.15	99.27	99.47	97.46	0.9863
	AD-MCI	80.47	88.90	66.26	81.62	77.98	0.7767
	MCI-NC	79.93	68.70	90.80	87.85	74.98	0.7980
4 Time Repeated Reduced Data	AD-NC	98.67	98.24	99.27	99.47	97.59	**0.9876**
	AD-MCI	80.61	88.92	66.59	81.81	78.02	0.7784
	MCI-NC	80.55	69.84	90.62	87.47	76.20	0.7998
5 Time Repeated Reduced Data	AD-NC	**98.81**	98.52	99.21	99.42	97.98	**0.9875**
	AD-MCI	80.69	89.46	66.37	81.29	79.39	0.7803
	MCI-NC	80.92	70.64	90.17	86.60	77.36	0.8016
6 Time Repeated Reduced Data	AD-NC	98.59	98.51	98.69	99.03	97.98	0.9866
	AD-MCI	80.81	89.47	66.65	81.46	79.43	0.7793
	MCI-NC	80.67	70.26	90.05	86.43	77.05	0.8045
7 Time Repeated Reduced Data	AD-NC	98.50	**98.61**	98.37	98.80	**98.11**	0.9852
	AD-MCI	80.71	89.07	66.66	81.82	78.32	0.7778
	MCI-NC	**81.44**	71.28	90.54	87.09	77.89	0.8056
8 Time Repeated Reduced Data	AD-NC	98.34	98.56	98.04	98.56	98.04	0.9835
	AD-MCI	80.84	**89.55**	66.57	81.45	**79.52**	0.7789
	MCI-NC	81.42	**71.51**	90.30	86.84	**77.98**	0.8075
9 Time Repeated Reduced Data	AD-NC	98.31	98.41	98.18	98.65	97.85	0.9822
	AD-MCI	80.51	88.91	66.43	81.62	78.13	0.7767
	MCI-NC	81.40	71.22	90.54	87.09	77.82	**0.8079**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

Peer-reviewed version available at Sensors 2020, 20, 941; doi:10.3390/s20030941

Table 3 Comparison of the proposed method with other methods.

	AD vs NC				AD vs MCI				MCI vs NC			
	Acc%	Sen%	Spec%	AUC	Acc%	Sen%	Spec%	AUC	Acc%	Sen%	Spec%	AUC
(D. Zhang et al., 2011)	93.20	93.00	93.30	0.98	-	-	-	-	76.40	81.80	66.00	0.81
(Dai et al., 2013)	90.81	92.59	90.33	0.94	85.92	82.46	87.59	0.87	81.92	78.51	88.34	0.81
(Suk et al., 2014)	93.05	90.86	94.57	0.95	**88.98**	82.11	**90.65**	0.90	83.67	**96.79**	57.28	0.82
(J. Liu et al., 2016)	94.65	95.03	91.76	0.95	88.63	**91.55**	86.25	**0.91**	**84.79**	88.91	80.34	**0.83**
(Mishra et al., 2018)	89.15	85.06	92.53	0.93	-	-	-	-	-	-	-	-
(Khedher et al., 2015)	88.96	92.35	86.24	0.93	84.59	88.75	83.07	0.89	82.41	84.12	80.48	0.81
(Lian et al., 2019)	90.00	82.00	97.00	0.95	-	-	-	-	-	-	-	-
(Ben Ahmed et al., 2014)	87.00	75.50	**100**	0.85	72.23	75.00	70.00	0.76	78.22	70.73	83.34	0.77
(Zhou et al., 2018)	93.75	87.5	**100**	-	-	-	-	-	-	-	-	-
Proposed Method (EL)	**98.81**	**98.52**	99.21	**0.987**	81.61	88.45	69.02	0.785	81.40	71.22	**90.54**	0.81

Here, experimental results have been presented. SVM classification algorithm with linear kernel was used for Alzheimer's Disease diagnosis. The evaluations were done using only one set of reduced data, and then using different number repetition of the reduced data. For this part, 144 selected features are used (including 132 MRI, 1 MMSE, 4 personal information, 3 CerebroSpinal Fluid, and 4 PET images). From Table 2 and Figures 3 to 5, the ROCs and Box-plots, AUC values increased after repeating reduced features, each time until the 5th repetition. From figure 3, using repeated features in width (Emphsised features) compresed Box-plots more, and it demonstrates higher stability in the classification. Bold numbers in Table 2 show the highest accuracy and other performance measures for the proposed method.

Discussion

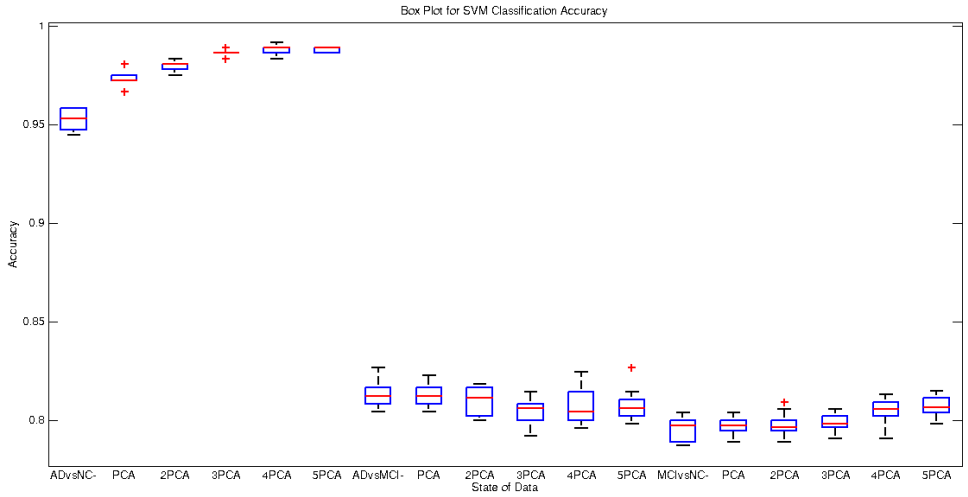
As mentioned in the main idea of this paper and as it is clear in Table 3, this metod has lower effects on data with low rates of precision and fairly weak models. In other words, this method emphasizes strong features. As can be seen in Table 2, regarding the models in which there is a reduction in precision after dimension reduction, or there is not much positive change, repetition in dimensions cannot cause a considerable increase in the model precision, as well. This also is predictable considering the main idea of the method; because the model emphasizes valuable features and when the extracted features do not have any considerable effect on model precision, repeating them cannot be very helpful in increasing the model precision. Issues regarding the main idea and the results of incorporating it are discussed in the following.

Feature Representation

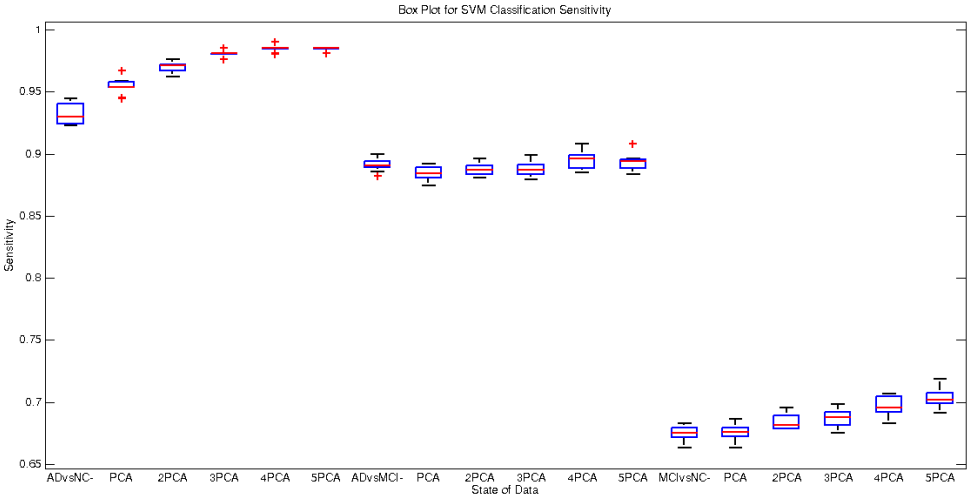
Across classification tasks, different numbers of input features can affect diagnosis of AD in supervised learning. In the literature, the effects of considering different input sizes for different classification problems are discussed a lot. The original features are informative for brain disease diagnosis, but this increase in the feature vector size will result in better and more calculable diagnosis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

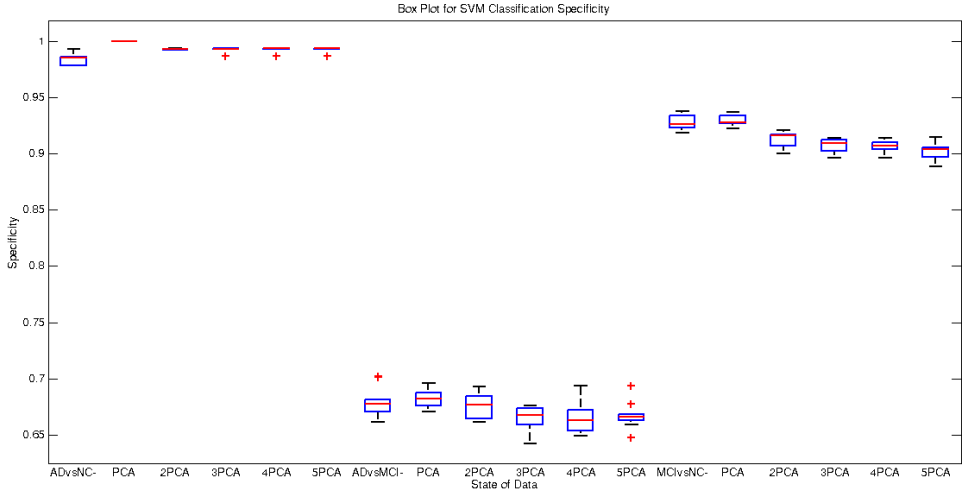
Peer-reviewed version available at *Sensors* **2020**, *20*, 941; doi:10.3390/s20030941



a)



b)



c)

Figure 3 Boxplots for recognition of AD, MCI, and NC subjects: a) Accuracy, b) Sensitivity, and c) Specificity of all featurs and 25 PCA elements for 1 to 9 times repeated features.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

Peer-reviewed version available at *Sensors* **2020**, *20*, 941; doi:10.3390/s20030941

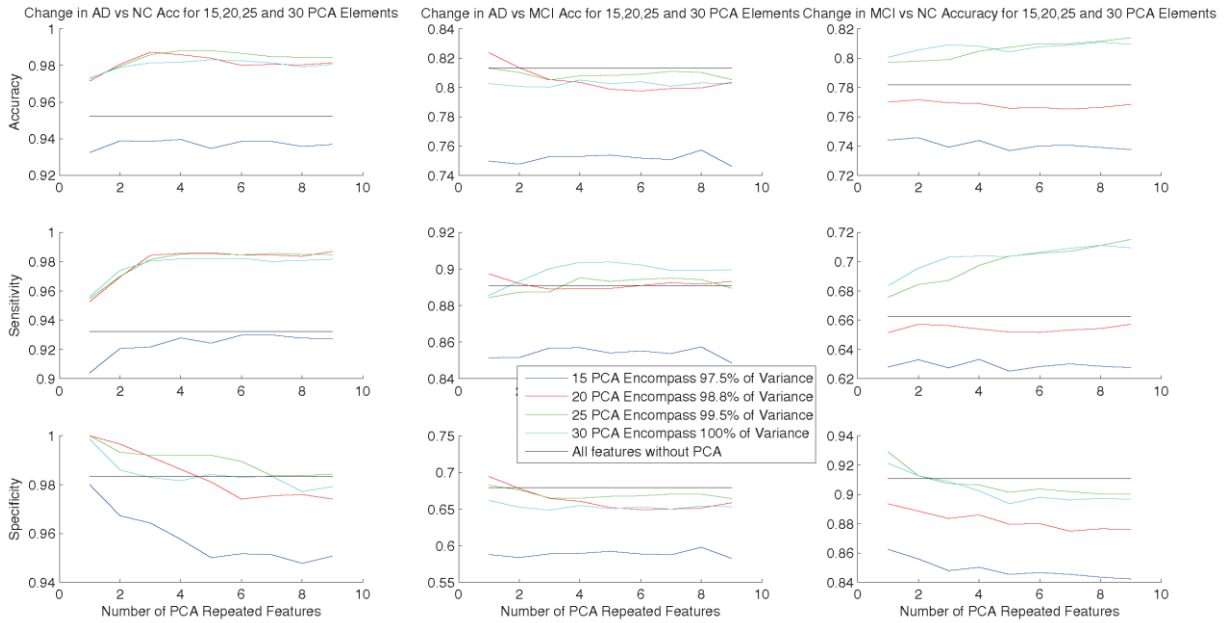
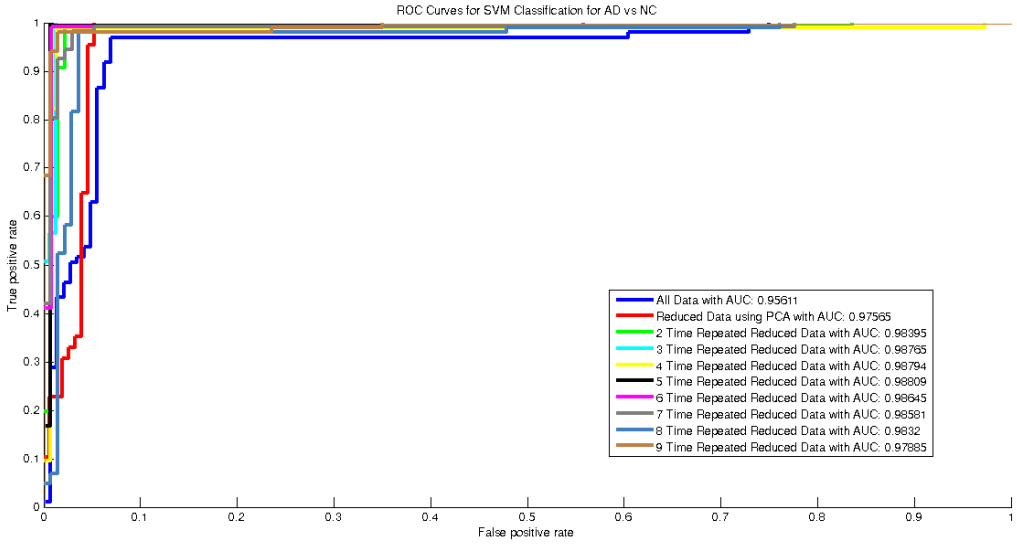


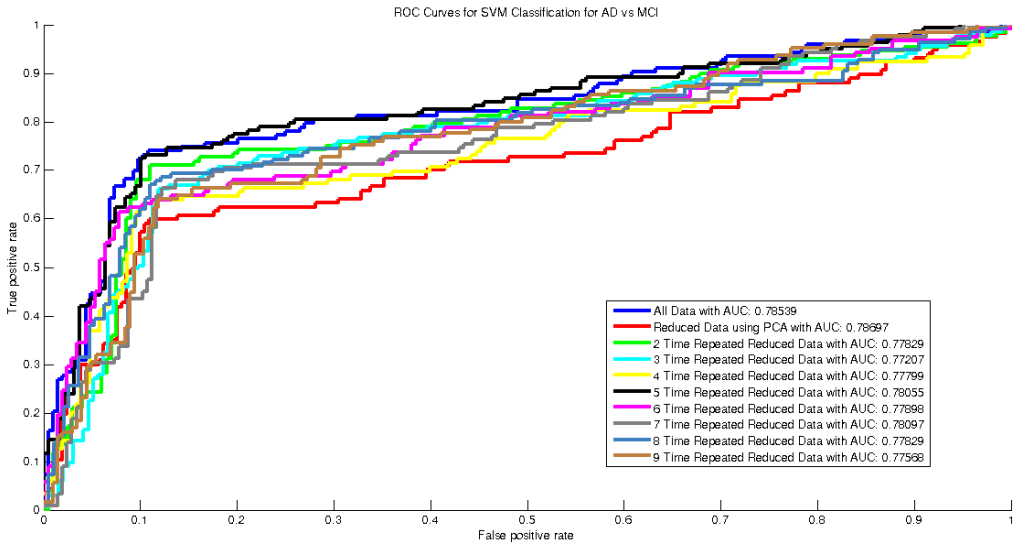
Figure 4. Comparison of changes in averaged accuracy, sensitivity and specificity for 15, 20, 25 , and 30 of PCA elements in 1 to 9 times repeated features, and all the features (black line).



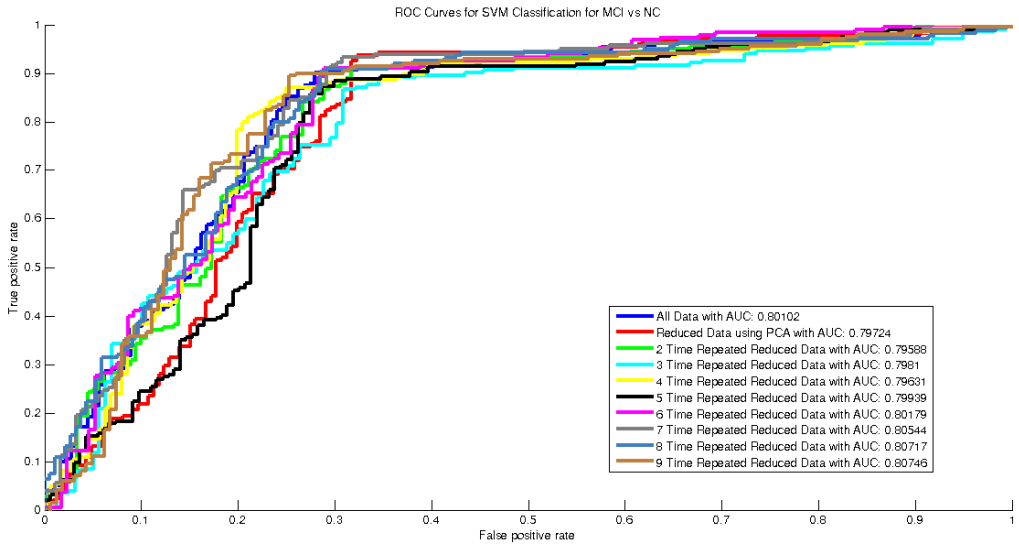
a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019 doi:10.20944/preprints201909.0308.v1

Peer-reviewed version available at *Sensors* **2020**, *20*, 941; doi:10.3390/s20030941



b)



c)

Figure 5. **ROC Curves for recognition of: a) AD vs NC, b) AD vs MCI, , and c) MCI vs NC, of all featurs and 25 PCA elements for 1 to 9 times repeated features.**

It can be said that, in comparison with the original feature, the proposed method improved the diagnostic accuracy for all the considered classification problems, altogether. The proposed method outperformed the other methods in three binary classification problems, mostly. Using this method, notwithstanding the limited number of samples, helps reduce errors for classification problems and hence enhance the classification accuracy. Previous methods can only use limited number of features in learning, but here, this limitation is limited.

There will not be an interpretation of the trained model and also the feature representations. Each added unit in the input represents linear combination of the high-level input features. That is, by repeating each high-level feature, e.g., mean intensity from FDG-PET or GM volume from MRI,

the model can cover different relations for low-level features. Using this method, and from a neuroscientific perspective, the relations from MRI features and from FDG-PET features could be bold. This way, new and increased inputs of the high-level features represent their helpfulness for classifying patients and their healthy normal controls over each other. Using this method cannot helps us to interpret or visualize the model's outputs and it still remains one of the unsolved pattern recognition and machine learning problems. On the contrary, it is clear that this combined information will be useful in AD/MCI diagnosis.

Feature Reduction and Increase

Here, we compare the results of the proposed method to PCA and before PCA results. In the feature set, we considered the clinical labels and clinical scores of MMSE. In conclusion, we observed that the method using the increase in a specific packet number of features outperformed others as presented in Table 3. Here, we selected 15 to 30 PCA components in approximately 94.5 to 100 percent of variance to test the proposed method.

The reason of the high performance of the proposed method can be explained as below:

Considering learning by a perfect object, when repeating it in the training process, the model can learn that object's features better and better; it is because, the richer information of the object can be learned by the model. Similarly, when we repeat perfect features (features obtained after applying PCA that gave us rich features of the objects), the model can learn richer information about the objects. Similar to repeating objects after specific number of repetition, the model can be over-trained, and specifying the number of the repetition will be a precise action. Therefore, the method can make features that can accurately model the target values, i.e., labels and imaginary symptoms.

It is definition of repeating features in width (feature repetition or Emphasized features) instead of length (sample repetition) in algorithm. We can say the repeated features for the labels can discriminate AD and MCI patients from NC subjects.

Comparison with the state-of-the-art methods

To validate the performance of the proposed method, we presented performance measures of the proposed method in Table 2. The performance of our proposed method is compared to other state of the art methods in Table 3. For comparison, we used all the 705 ADNI samples. The accuracy rates of our proposed method were 98.81%, 81.61%, and 81.40% for AD vs. NC, MCI vs. NC, and AD vs. MCI, classification, respectively.

Most of the listed methods in Table 3 have used ADNI database, and we used all the images and data in ADNI, consisting of 705 participants. As some of the reported methods used a portion of the ADNI samples, we have rights to compare our proposed method to them. As can be seen in the Table 3, the proposed method obtained the highest accuracy (98.81%) in AD vs. NC diagnosis, and due to the balanced dataset, the accuracy is a suitable performance measure.

Additionally, sensitivity (recall rate) of the proposed method (98.52) is the highest sensitivity rate as can be seen in Table 3. The proposed method stood at the second place with specificity of 99.21% against perfect detection rate of Ben Ahmed et al.'s (2014). However, Ben Ahmed et al.'s method had a high false alarm rate while the proposed method succeeded to achieve a trade-off between the two! It is worth mentioning that the lower the specificity, the higher the sensitivity (94.83 vs. 87 for example in case of Ben Ahmed et al.'s method). Nevertheless, combined methods can be used to achieve better performance.

We can redress proposed method's some imperfections using other method that has better performance. According to the Table 3, the proposed method in most of the performance measures

was in the first or second place and from that we can realize that the proposed method is the most dominant one besides the highest accuracy compared to other methods.

Limitations of the proposed method

The proposed method has some limitations and handicaps. In PET imaging, the partial volume effect induced by a combination of image sampling and restricted resolution of PET, in the reconstructed images, can bring under- or over valuation for regional radioactivity condensation; and thereupon, more errors in statistical parametric images may happen (Aston et al., 2002). Combination of multiple tissue values will likely effect the differences between voxels of Gray and White Matters. Since part of proposed method is ROI-based feature selection, this partial voxel quality reduction will have inconsiderable effect on the performance of the method.

We can say, the proposed structure used to form the feature sets in this experiment cannot be optimal for other datasets. We need studies such as learning the optimal and strong feature sets for repetition and practical use of the proposed method. The NC group in the dataset could include both healthy controls and subjective cognitive complaints because there is no supplementary information about this group. The features concatenation from MRI, FDGPET, MMSE scores, and CSF modalities into a single vector and repetition of the features after the feature reduction can efficiently distinguish between AD and MCI patients and NC subjects.

Conclusions

In this study, we proposed a simple but practical and drastic method for classification issue, and tested it for Alzheimer's disease diagnosis. Our proposed method was aimed at finding the best features and repeating them until we do not achieve better classification perfomance. We examined our method on the ADNI database of AD. The experiments showed that we can achieve much better perfomance using combined features of MRI, PET, CSF, MMSE, and Personal Infrormation, especially when we repeated the reduced features on all the three binary classification problems. Experiments indicated performance and effectiveness of the proposed method, that are accuracy rates of 98.81%, 81.61%, and 81.40% for AD vs. NC, MCI vs. NC, and AD vs. MCI classification problems, respectively. As can be seen, using this method increased performance of the three binary problems, one incredibly. The results showed that the classification accuracy is improved with the optimized feature selection and repeatition. Some other feature selection and reduction such as information gain method can be used to select the more sensitive features in AD and MCI diagnosis. Using other feature reduction or selection methods and repeating reducd data can be a future work. Combining results of other feature reduction and selection methods and establishing a classification framework, using these, can be another future work too. As another future work, applying this method to clustering field can be recommended.

Conflict of Interests: There are no conflicts of interest associated with this publication and there has not been financial support for this research work that could have influenced its outcome.

Acknowledgements: The data used in this paper were obtained from the Alzheimer's Diseases Neuroimaging Initiative (ADNI) database. 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). The primary goal of ADNI has been to test whether magnetic resonance imaging, other biomarkers, and clinical and neuropsychological estimations can be combined to measure the progression of early stage of Alzheimer's disease.

The purpose of determination of sensitive and specific biomarkers of early AD progression is to develop new treatments and monitor their effectiveness, and reducing the time and cost of clinical experiments.[3]

References

Aston, J. A., Cunningham, V. J., Asselin, M. C., Hammers, A., Evans, A. C., & Gunn, R. N. (2002). Positron emission tomography partial volume correction: estimation and algorithms. *J Cereb Blood Flow Metab, 22*(8), 1019-1034. doi: 10.1097/00004647-200208000-00014

Ben Ahmed, O., Benois-Pineau, J., Allard, M., Ben Amar, C., & Catheline, G. (2014). Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimedia Tools and Applications, 74*(4), 1249-1266. doi: 10.1007/s11042-014-2123-y

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov., 2*(2), 121–167.

Dai, D., He, H., Vogelstein, J. T., & Hou, Z. (2013). Accurate prediction of AD patients using cortical thickness networks. *Machine vision and applications, 24*(7), 1445-1457.

Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging, 32*(12), 19–27.

Duin, R. P. W. (2000). *Classifiers in almost empty spaces*. Paper presented at the Proceedings 15th International Conference on Pattern Recognition.

Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage, 41*(2), 277-285. doi: 10.1016/j.neuroimage.2008.02.043

Garyfallou, G. Z., Ketebu, O., Sahin, S., Mukaetova-Ladinska, E. B., Catt, M., & Yu, E. H. (2017). Electrochemical Detection of Plasma Immunoglobulin as a Biomarker for Alzheimer's Disease. *Sensors (Basel), 17*(11). doi: 10.3390/s17112464

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for AD in a multimodality framework: an analysis of MCI progression in the ADNI population. *Neuroimage, 55*(2), 574–589.

Islam, K., Damiati, S., Sethi, J., Suhail, A., & Pan, G. (2018). Development of a Label-Free Immunosensor for Clusterin Detection as an Alzheimer's Biomarker. *Sensors (Basel), 18*(1). doi: 10.3390/s18010308

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., & Segovia, F. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing, 151, Part 1*, 139-150. doi: http://dx.doi.org/10.1016/j.neucom.2014.09.072

Kloppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., . . . Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain, 131*(Pt 3), 681-689. doi: 10.1093/brain/awm319

La Foresta, F., Morabito, C. F., Marino, S., & Dattola, S. (2019). High-Density EEG Signal Processing Based on Active-Source Reconstruction for Brain Network Analysis in Alzheimer's Disease. *Electronics, 8*(9). doi: 10.3390/electronics8091031

Lian, C., Liu, M., Zhang, J., & Shen, D. (2019). Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis using Structural MRI. *IEEE Trans Pattern Anal Mach Intell*, 1-1. doi: 10.1109/TPAMI.2018.2889096

[3]This section has been added because of ADNI request.

Liu, F., Suk, H. I., Wee, C. Y., Chen, H., & Shen, D. (2013). High-order graph matching based feature selection for Alzheimer's disease identification. *Med Image Comput Comput Assist Interv, 16*(Pt 2), 311-318.

Liu, J., Li, M., Lan, W., Wu, F.-X., Pan, Y., & Wang, J. (2016). Classification of alzheimer's disease using whole brain hierarchical network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15*(2), 624-632.

Maqsood, M., Nazir, F., Khan, U., & Aadil, F. (2019). Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans. *Sensors (Basel), 19*(11). doi: 10.3390/s19112645

Mishra, S., Beheshti, I., Khanna, P., & Initiative, f. t. A. s. D. N. (2018). A statistical region selection and randomized volumetric features selection framework for early detection of Alzheimer's disease. *International Journal of Imaging Systems and Technology, 28*(4), 302-314. doi: 10.1002/ima.22290

Mourao-Miranda, J., Bokde, A. L., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage, 28*(4), 980-995. doi: 10.1016/j.neuroimage.2005.06.070

Nettiksimmons, J., Harvey, D., Brewer, J., Carmichael, O., DeCarli, C., Jack Jr, C. R., . . . Weiner, M. W. (2010). Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. *Neurobiology of Aging, 31*(8), 1419-1428.

Nordberg, A., Rinne, J. O., Kadir, A., & Langstrom, B. (2010). The use of PET in Alzheimer disease. *Nat. Rev. Neurol., 6*, 78–87.

Shawe-Taylor, J., & Cristianini, N. (2000). *Support Vector Machines and Other Kernel-Based, Learning Methods*. UK: Cambridge University Press.

Suk, H.-I., Lee, S.-W., Shen, D., & Initiative, A. s. D. N. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage, 101*, 569-582.

Thies, W., & Bleiler, L. (2012). 2012 Alzheimer’s disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*.

Toro, C. A. O., & Gonzalo Martin, C. (2018). Supervoxels-Based Histon as a New Alzheimer's Disease Imaging Biomarker. *Sensors (Basel), 18*(6). doi: 10.3390/s18061752

Wee, C.-Y., Yap, P.-T., Li, W., Denny, K., Browndyke, J. N., Potter, G. G., . . . Shen, D. (2011). Enriched white matter connectivity networks for accurate identification of MCI patients. *Neuroimage, 54*(3), 1812-1822.

Zhang, D., & Shen, D. (2012). Multi-modalmulti-task learning for joint prediction ofmultiple regression and classification variables in Alzheimer's disease. *Neuroimage, 59*, 895–907.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage, 55*, 856–867.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Trans. Med. Imaging, 20*, 45–57.

Zhou, K., He, W., Xu, Y., Xiong, G., & Cai, J. (2018). Feature Selection and Transfer Learning for Alzheimer’s Disease Clinical Diagnosis. *Applied Sciences, 8*(8), 1372.