

Article

Emphasis Learning, Features Repetition in Width Instead of Length to Improve Classification Performance: Case Study—Alzheimer’s Disease Diagnosis

Hamid Akramifard ¹, MohammadAli Balafar ^{1,*}, SeyedNaser Razavi ¹ and Abd Rahman Ramli ²

¹ . Faculty of Electrical and Computer Engineering, University of Tabriz, East Azerbaijan, Tabriz 51666-16471, Iran; hamid.akramifard@tabrizu.ac.ir (H.A.); n.razavi@tabrizu.ac.ir (S.R.)

² . Department of Computer and Communication Systems Engineering, University Putra Malaysia, UPM-Serdang 43400, Malaysia; arr@eng.upm.edu.my

* Correspondence: balafarila@tabrizu.ac.ir

Received: 26 September 2019; Accepted: 28 October 2019; Published: 10 February 2020



Abstract: In the past decade, many studies have been conducted to advance computer-aided systems for Alzheimer’s disease (AD) diagnosis. Most of them have recently developed systems concentrated on extracting and combining features from MRI, PET, and CSF. For the most part, they have obtained very high performance. However, improving the performance of a classification problem is complicated, specifically when the model’s accuracy or other performance measurements are higher than 90%. In this study, a novel methodology is proposed to address this problem, specifically in Alzheimer’s disease diagnosis classification. This methodology is the first of its kind in the literature, based on the notion of replication on the feature space instead of the traditional sample space. Briefly, the main steps of the proposed method include extracting, embedding, and exploring the best subset of features. For feature extraction, we adopt VBM-SPM; for embedding features, a concatenation strategy is used on the features to ultimately create one feature vector for each subject. Principal component analysis is applied to extract new features, forming a low-dimensional compact space. A novel process is applied by replicating selected components, assessing the classification model, and repeating the replication until performance divergence or convergence. The proposed method aims to explore most significant features and highest-performing model at the same time, to classify normal subjects from AD and mild cognitive impairment (MCI) patients. In each epoch, a small subset of candidate features is assessed by support vector machine (SVM) classifier. This repeating procedure is continued until the highest performance is achieved. Experimental results reveal the highest performance reported in the literature for this specific classification problem. We obtained a model with accuracies of 98.81%, 81.61%, and 81.40% for AD vs. normal control (NC), MCI vs. NC, and AD vs. MCI classification, respectively.

Keywords: Alzheimer’s disease; Emphasis Learning; multi-modal classification; SVM; PCA

1. Introduction

Alzheimer’s disease (AD) can be described by cognitive and memory dysfunctions. This in turn is actually the major cause of dementia in older adults. Moreover, AD has been identified as one of the main causes of death in the United States [1]. The early diagnosis and prognosis of AD are important because of limitations in treatment time.

In this area, many biomedical imaging techniques for the early detection of AD are well developed and employed by researchers, including MRI [2–4], PET [5,6], and other features like CSF [7], and the Mini-Mental State Examination (MMSE) [8].

To develop automated procedures based on the above techniques to detect brain atrophy, in the initial stages of AD, regions including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, and anterior and posterior cingulate cortex [9–12] have been reported. Previous works have dealt with the construction of computer-aided diagnosis (CAD) systems. Almost all of these CAD systems are based on machine learning techniques and have three main steps: data pre-processing, feature extraction, and classification. The pre-processing procedure sets different images from different subjects, with brains of different sizes and shapes, at a comparable condition and cleans and imputes missing data from the obtained data (if any). In the second step, a feature extraction algorithm converts the input data into small vectors [13]. The classifier determines if the vectors are more similar to mild cognitive impairment (MCI) patient vectors, to AD patient vectors, or to normal control (NC) vectors.

To use these CAD systems, the metrics of entorhinal cortex have been used in AD diagnosis [14]. Automatic hippocampal volume measurement methods have almost equal results [15,16]. Hippocampal volumes and entorhinal cortex metrics seem to be equally accurate in distinguishing between AD patients and NC subjects [17]. Different techniques, such as principal component analysis (PCA), artificial neural networks (ANNs), fuzzy neural networks (FNNs), partial least square (PLS), and support vector machine (SVM), have been used in the development of these CADs.

These brain-observing techniques using machine learning can provide tools to overcome brain dysfunction problems. These combined techniques can use different modalities including MRI, PET, and other neurological data to diagnose AD/MCI patients from healthy people [18–21]. In [22] 50 MRI images from the OASIS dataset were used for characterization of MRIs of brains affected with Alzheimer's disease by fractal descriptors. Additionally, [23] used MRI images from the Alzheimer's Diseases Neuroimaging Initiative (ADNI) dataset for distinguishing AD from NC. They reported complete performance (100% accuracy) in distinguishing between the two groups. [18] reported a multiple classification using transfer learning on AD, while [24] classified AD vs. NC with a great rate of accuracy using only MRI data. [25] classified progressive MCI vs. static MCI using combined MRI, APOe4 genetic, and cognitive measures include and APOe4 genotyping.

In this area, the feature extraction and feature combination are often performed independently. As investigated in the previous studies, there are inherent relations between the modalities of MRI and PET [26]. Thus, finding the shared feature representation that combines the complementary information from different modalities (e.g., PET, MRI, and CSF) is useful to enhance the discrimination of AD and MCI patients from NC subjects.

There are some features among the described data that can help us to better diagnose AD. We use PCA for dimensionality reduction and recognition of the possibly most efficient features of the data to enhance classification potential. Feature representation using PCA reduces processing resources usage, in addition to enhancing the classification accuracy. The steps of the proposed approach can be summarized as follows:

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

Briefly, the main contributions of this study are as follows:

- A novel method named Emphasis Learning is proposed for improving classification performance.
- The proposed method is successfully adapted for the diagnosis and prognosis of AD patients and distinguishing them from normal subjects.

This methodology is the first of its kind in the literature, and works based on the notion of replication on the feature space instead of the traditional sample space. PCA was employed only as an expert feature engineer to extract high-variance features. The proposed method achieved about 99% accuracy in classifying normal subjects from AD patients. This result outperformed all of the current literature results in terms of accuracy in classifying AD patients from normal subjects.

In the following sections, the data extraction and data source are presented. Then, pre-processing, feature reduction and increment, and classification methods are explained. In the following section, the experimental results are provided. A discussion of the results and conclusions are presented in the final sections.

2. Materials and Methods

In this method, we emphasized the strongest and 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 improves as well. In other words, these features improve themselves; and the better and more precise these features, the more effective they can be. To put this another way, learning can result either when good features in small quantities repeat several times, or when such positive features repeat few times. For instance, which of the two following ways would result in higher learning rates? When a teacher solves a problem with two solutions and repeats these many times? Or, when she uses many solutions while repeating each just a few times? The answer is "both". However, for some problems the former works better, and for some, the latter. The same is true in computer applications, and deeper and broader learning takes place when there is the possibility of repeating both approaches.

It is obvious that trying to increase the accuracy of a classifier that is fine-tuned is a challenge. For example, usually, increasing the accuracy from 80% to 85% is less complex and needs lower costs and computational burden than increasing it from 95% to 96% (if possible). In this study, a method is put forward in order to make this possible and to reduce the costs and computational burden to a great extent. However, when the computed performance using the main data is low, or when utilizing the extracted features from the dimension reduction does not cause a change in the model's performance (i.e., if no outstanding feature is achieved), repeating these features may not be that influential in the tuning precision.

2.1. Characteristics of Subjects

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

Table 1. Summary of demographic data of patients and subjects. AD: Alzheimer's disease; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; NC: normal control.

	Count	Male	Female	Married	Widowed	Divorced	Never Married	Average Age	Average 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

2.2. 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 after an overnight use of 20- or 24-gauge spinal needle. The FluoroDeoxyGlucose-Positron Emission Tomography (FDG-PET) images were average values obtained from 30–60 min post injection; their voxel size was converted to standard and smoothed to 8 mm full width at half of maximum resolution.

2.2.1. MRI Acquisition Parameters

In the multiple ADNI sites, multiple machines (e.g., Siemens, Philips, and GE Medical scanners) are used. Standard protocol was developed to evaluate 3D T1-weighted sequences for morphometric analyses [27]. Structural brain MRI scans were acquired using 1.5 T and 3 T MRI scanners. Most of the 1.5 T MRIs were obtained from GE Medical scanners, and most of the 3 T MRIs were acquired from Siemens machines.

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

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

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

Figure 1 shows a sample MRI imaging of an NC subject and an AD patient. The figure demonstrates decreased gray matter (GM) volume in the AD patient compared to the normal control.

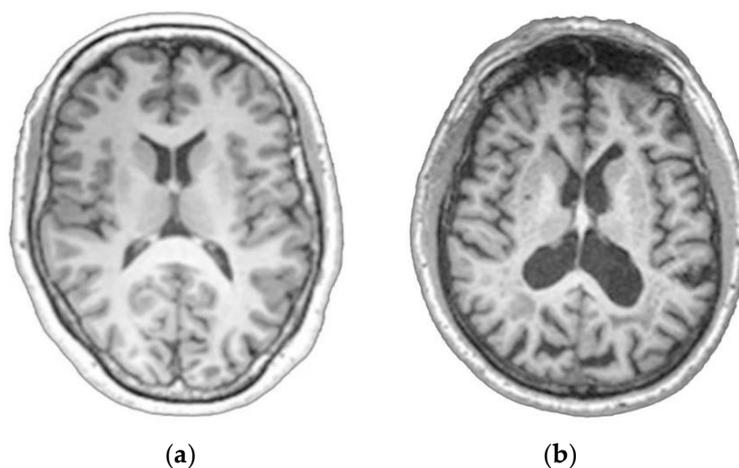


Figure 1. MRI imaging sample: (a) NC subject; (b) AD patient.

2.2.2. Pre-Processing of MRI Images

Spatial parametric mapping (SPM) software was used for pre-processing [28]. SPM was used for realignment, smoothing, spatial normalization, and feature extraction from MRI regions of interest (ROIs). The pre-processing steps using VBM8 tools were as below:

1. Check that image format is in a suitable condition using SPM tools.

2. Segment the images to identify gray matter and white matter (WM) and wrap GM to the segmented image to Montreal Neurological Institute (MNI) space using the SPM tools.
3. Estimate deformations 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. Generate spatially normalized and smoothed GM images normalized to MNI space. Using the estimated deformations by the DARTEL tools of SPM, generate smoothed/modulated wrapped GM and WM images.

Note that using CAT12 for MRI segmentation and feature extraction can also obtain promising results, as reported by Farokhian et al. [29]. This study demonstrated better performance using CAT12 over VBM8 tools.

Data cleansing and selection were done in the pre-processing step. In the second step (feature extraction), the input data were converted into small vectors [13]. The classification algorithm determined whether the vectors are more similar to an MCI/AD patient or to a normal control (NC) subject.

2.3. 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 pre-processing and preparing. MIPAV software was used for AC–PC correction. We segmented structural MRI images into WM, GM, and CSF images (in the literature, GM matter has been widely used in AD diagnosis, and most of the brain structural MRI studies in AD focused on only gray matter abnormalities [30]). Then, to extract ROI features of all the images, we used Voxel Based Morphometry tools of Spatial Parametric Mapping (VBM-SPM) to extract GM features as well. After that, volumetric changes in specific regions such as entorhinal cortex, hippocampus, and temporal and parietal lobes were used. For each ROI, a mask was made using WFU Pick Atlas tools (https://www.nitrc.org/projects/wfu_pickatlas/).

The PET images were aligned to the corresponding MR images strictly in ADNI. The recognized areas including MRI gray matter tissue volume, average voxel values, and average of PET voxel values (that were downloaded from the ADNI database) were used as features. In the literature, these features have been used for AD/MCI diagnosis [31–34]. Three CSF biomarkers (i.e., Ab42, t-tau, and p-tau) were also used in making the feature set. Therefore, 144 features were used to form the final feature set consisting of 132 Voxel values and Volume of MRIs, 1 MMSE score, 4 pieces of personal information, 3 CSF biomarkers, and 4 PET voxel values (since we could not access PET images in the ADNI database, we used only four PET voxel values that were extracted and uploaded to the ADNI database). Finally, the vectors of the extracted features were normalized by applying natural logarithm. K-fold cross validation method was used for testing and evaluation.

To parcellate the brain, Automatic Anatomical Labeling (AAL) atlas (<http://www.gin.cnrs.fr/en/tools/aal/>) (Figure 2) was used, as proposed by [35]. In this atlas, the brain is parcellated into 90 cerebral regions and 26 cerebellar regions. The sixteen most effective regions of the brain for this work are respectively the left amygdala, inferior temporal gyrus, left middle frontal gyrus, left inferior temporal gyrus, right amygdala, left middle temporal gyrus, left middle temporal gyrus, left supramarginal gyrus, left middle frontal gyrus, left inferior frontal gyrus, left hippocampus, left angular gyrus, left superior frontal gyrus, right supramarginal gyrus, right hippocampus, and right parahippocampal gyrus. We used both voxel values and volumes of these regions in this study.

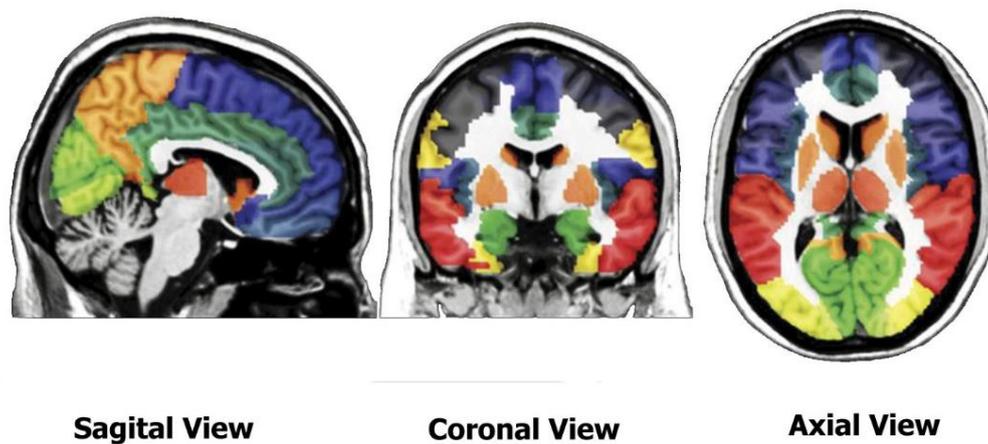


Figure 2. Automatic Anatomical Labeling (AAL) atlas. Different regions of brain are in different colors.

3. Classification Methods

For diagnosis and prognosis of AD and MCI, some classification algorithms are common and some algorithms play supportive roles. Among them, SVM and PCA are the most used.

3.1. Feature Reduction Method

One of the most common linear techniques for data dimension reduction is PCA. PCA was introduced by Karl Pearson in [36]. It maps the data to a lower dimension while maintaining the data's variance. To use this method, the covariance matrix of the data and the eigenvectors on this matrix must be computed. The eigenvectors from the largest eigenvalues (i.e., 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, the process yields a smaller number of eigenvectors, and there may be some associated data loss. However, the most important variances should be retained by the remaining eigenvectors. Figure 3 shows the eigenvectors of a dataset.

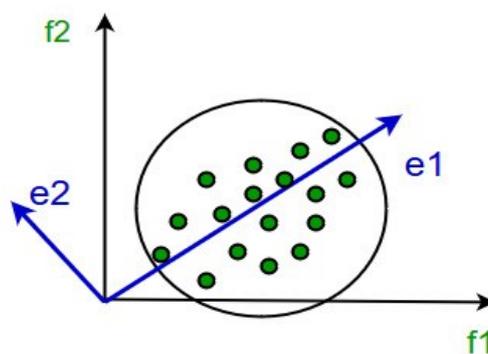


Figure 3. Schematic representation of dataset eigenvectors.

PCA has the advantages of reducing the required storage space and computation time eliminating redundant features. Some associated disadvantages include a reduction in some of the original data's information, its failure when mean and covariance are not sufficient to define the data, and uncertainty in the number of principal components required to retain the data's information.

3.2. Increasing Dimensions of Data to Achieve Better Classification Results

The main idea of this paper is to repeat the most efficient features in classification. Theoretically, increasing the dimensions of data can sometimes yield better and sometimes worse classification results; but what if we find and repeat good data features to make a classification model? Our experiment

shows that this theory worked excellently! For the diagnosis of Alzheimer’s Disease, we tested this method after extracting the best features of the data set using PCA, conducted a dimension reduction, and repeated these new features. We repeated these features as input data for the classification algorithm (here SVM) until this action had no further positive effect on classification performance or did not reduce it. Figure 4 shows the proposed method diagram. The performance of the proposed model increased in some cases after applying PCA, but in many other cases it might lead to performance loss. Because using more PCs can lead to performance loss, we experimentally found that using only 25 gave the best results.

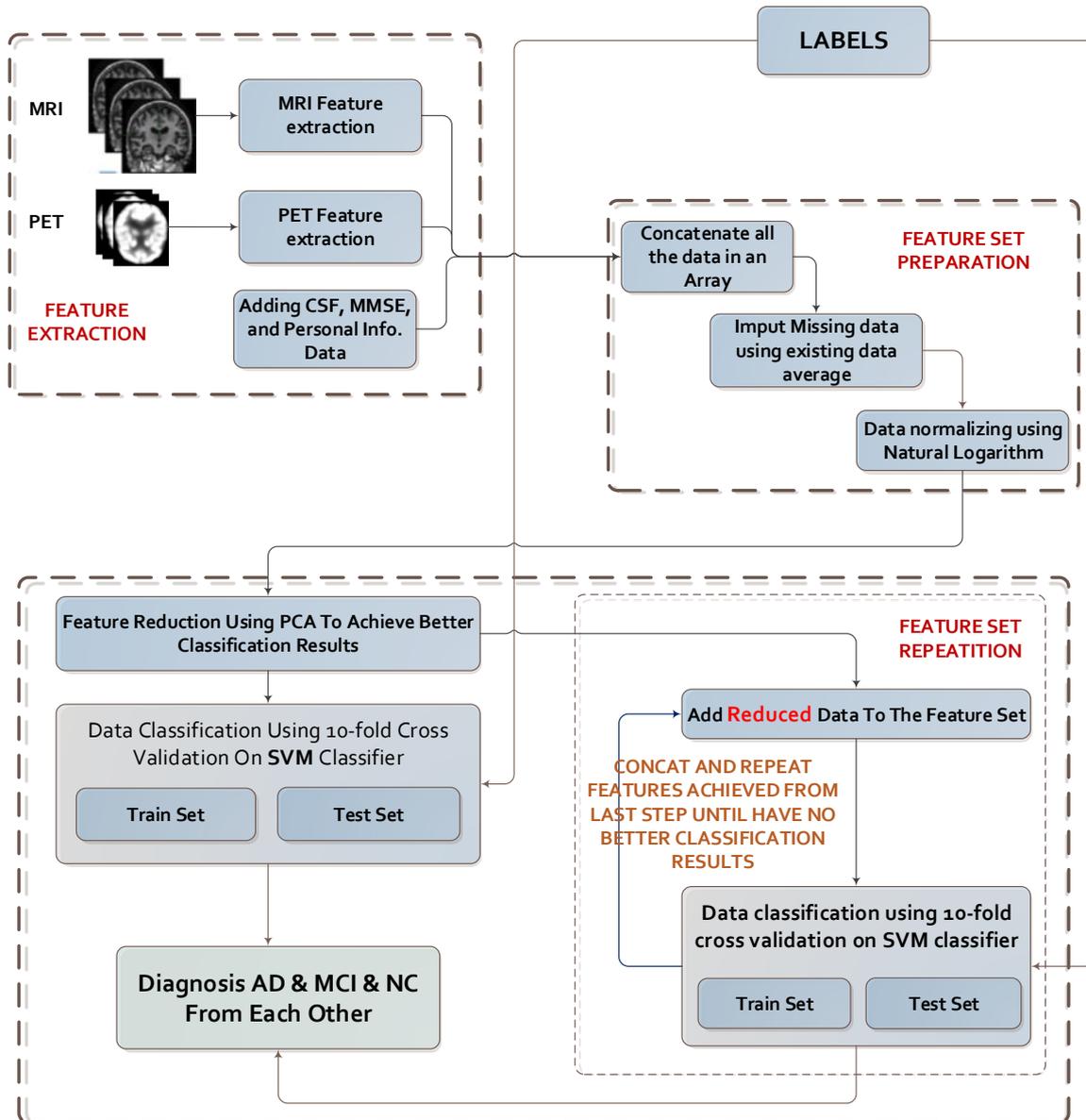


Figure 4. Diagram of steps of the proposed method (Emphasis Learning). CSF: cerebro-spinal fluid.

3.3. SVM

One binary classification method that is successfully used in many domains is SVM [3,37–39]. The classification efficiency of SVM in training very-high-dimensional data has been proven [40,41]. Moreover, SVM has been applied to voice activity detection, pattern recognition, classification, and regression analysis [42,43]. 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 subjects, pairwise. We employed only the training samples to adjust the hyperparameters of the SVM, based on which the best model was selected. Using the proposed method, we could automatically select the best model among other trained models due to the achieved performance rates.

3.4. Data Normalization vs. Data Standardization

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

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (1)$$

where X_{norm} is the normalized input data. The problem with normalization is outlier elimination.

On the other hand, standardization maps original data to have a mean of 0, and is recommended in some cases. Equation (2) demonstrates standardization:

$$X_{std} = \frac{X - X_{mean}}{X_{dev}}, \quad (2)$$

where X_{std} implies standardized data, X_{mean} is the mean of the original data, and X_{dev} is the standard deviation of the data.

3.5. Evaluation Criteria

Accuracy is a well-known evaluation measure for classification methods. Using accuracy, we computed the correctly classified samples and all samples' ratio. Two other common evaluation metrics are sensitivity and specificity. The receiver operating characteristic (ROC) and the area under the curve (AUC) are other performance parameters for diagnosis procedures. The positive predictive value (PPV) and negative predictive value (NPV) are widely used measures to describe 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 the number of true positives (number correctly classified as patients); TN is the number of true negatives (number correctly classified as non-patients); FP is the number of false positives (number of non-patients wrongly classified as patients); FN is the number of false negatives (number of patients wrongly classified as non-patients).

We used the sensitivity and specificity to evaluate the rate of true positives or negatives (i.e., the ratio of correctly classified AD or MCI patients or NC subjects and total subjects). These measures show the method's 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 were sampled randomly—one set for testing and nine for training the classifier. This was done for all 10 sets, and the average of the evaluation parameters was used to show the performance of the

classification method. In this article, we repeated the K-fold method 100 times and the average of averages was used to represent the method's performance.

The steps of the algorithm are as follows (Algorithm 1):

Algorithm 1. The steps of the algorithm of the method.

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. Imputed data <- Imputation (Concatenated Features)
 4. data <- ln(Imputed data) 'data normalization using Natural Logarithm'
 5. Normalized data <- (data-min(data))/(max(data)-min(data)) 'data normalization (Xnorm)'
 6. Reduced data <- PCA (Normalized data)
 7. Data <- [Reduced data]
 8. Diagnosis <- SVM(Data) 'classification using SVM'
 9. Data <- [Data, Reduced data]
 10. Go to (8) until no further performance improvements are obtained
 11. Select the best performance and Finish
-

The third step was concerned with missing values, because there are many missing values in PET and CSF data. The missing values were replaced by the average (mean) of the existing items for all subjects as suggested by [44]. This method was successfully applied to fill the missing values and we saw the performance gain after using that. Note that PET and CSF examinations have difficult processes. For this reason, some of the ADNI subjects were avoid from these experiments. It should be said that none of MMSE, MRI, or demographic data had missing values so all of them were used in our experimentations.

4. Experimental Results

Here, we evaluate our proposed method's efficiency. This was done for three binary classification problems: AD vs. NC, MCI vs. NC, and AD vs. MCI, and a 10-fold cross validation method was used for evaluation purposes. In the 10-fold cross validation, the dataset was randomly partitioned into 10 subsets, each including one-tenth of the total dataset. Nine subsets were used for training goals and the remaining one for testing. We did this for all subsets.

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. Table 2 shows the mean accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve for different numbers of the repeated features tested on the proposed method on three binary classes. As can be seen, by repeating features five times, the proposed method showed the best accuracy rates of 98.81%, 81.61%, and 81.40% in classifying NC vs. AD, MCI vs. AD, and NC vs. MCI data, respectively.

Here, experimental results are presented. An SVM classification algorithm with a linear kernel was used for Alzheimer's disease diagnosis. The evaluations were done using only one set of reduced data, and then using different numbers of repetitions of the reduced data. Toward this aim, 144 selected features were used (including 132 MRI voxel and volume values, 1 MMSE, 4 types personal

information, 3 cerebro-spinal fluid biomarkers, and 4 PET image voxel sizes). Table 2 shows the accuracy of discriminating AD, MCI, and NC from each other using each group of data, alone.

As deduced from Table 2 and Figures 5–7, the AUC values increased after repeating reduced features each time until the fifth repetition. As can be seen in Figure 5, using repeated features in width (emphasized features) compressed box-plots further, demonstrating higher stability in the classification. The bold numbers in Table 2 show the highest values for accuracy and other performance measures for the proposed method.

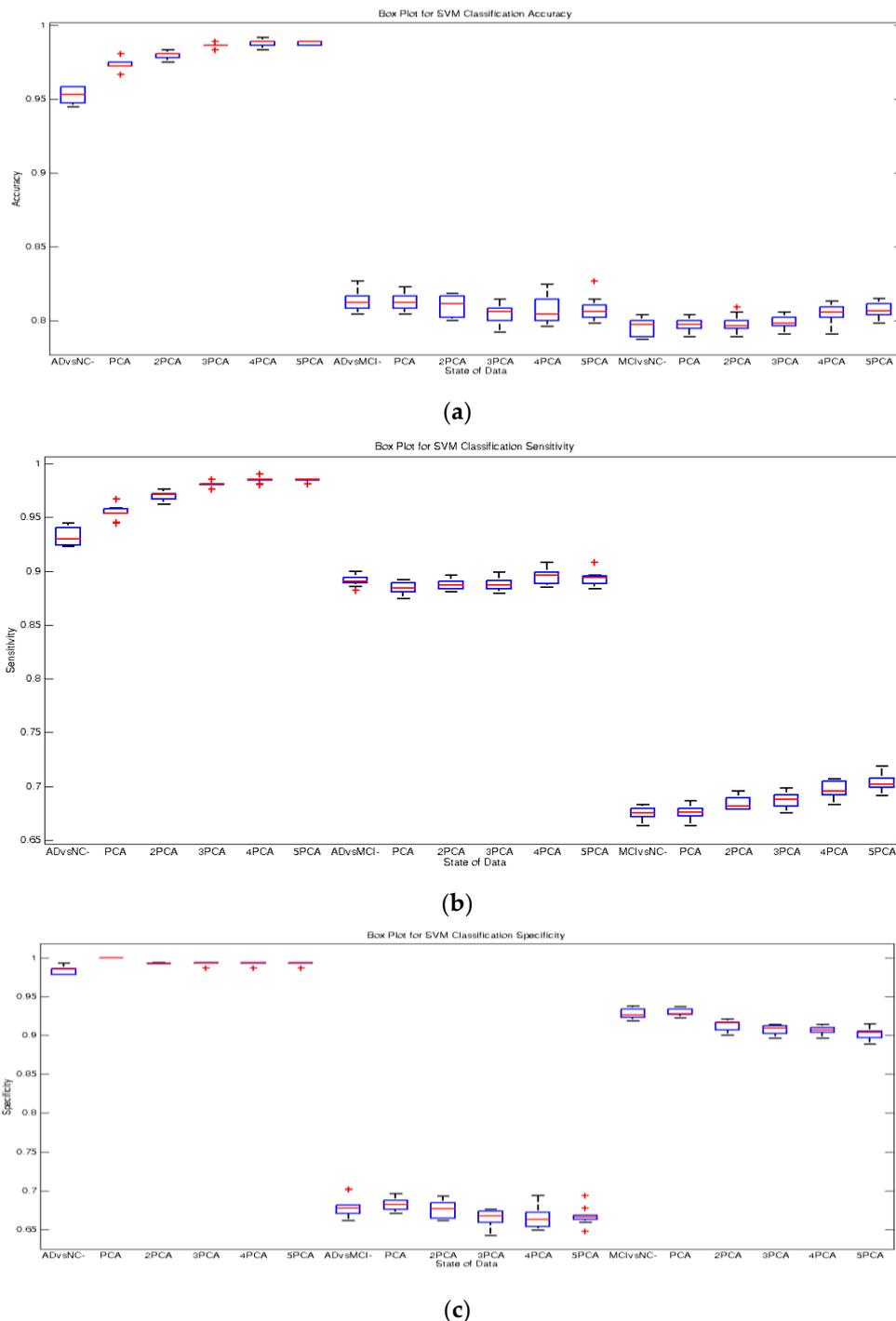


Figure 5. Boxplots for the recognition of AD, MCI, and NC subjects: (a) accuracy, (b) sensitivity, and (c) specificity of all features and 25 PCA elements for 1–9× repeated feature reduction.

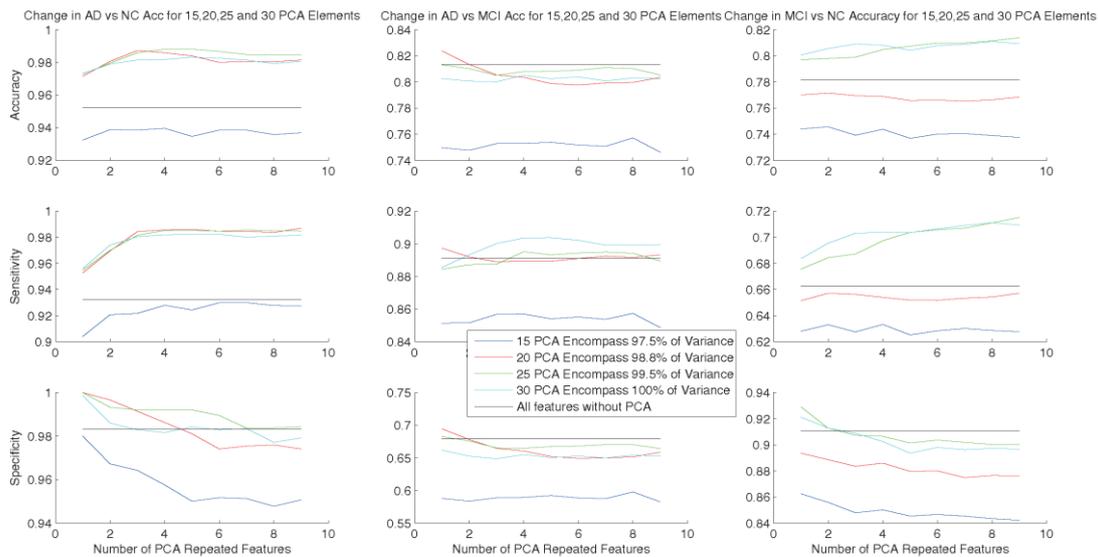
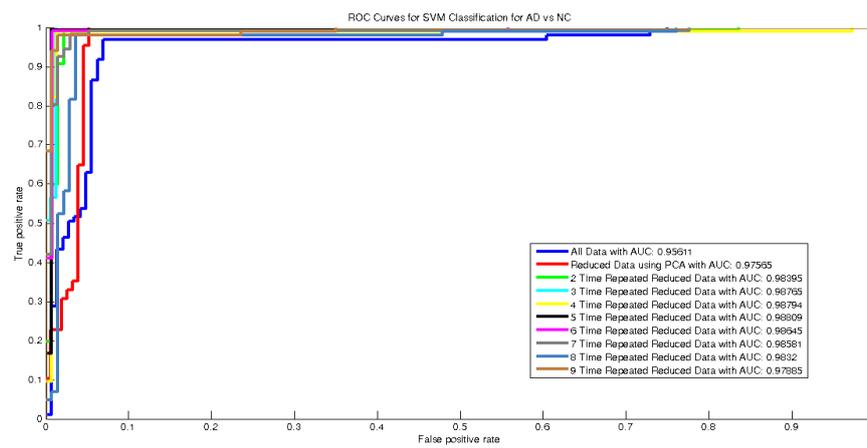


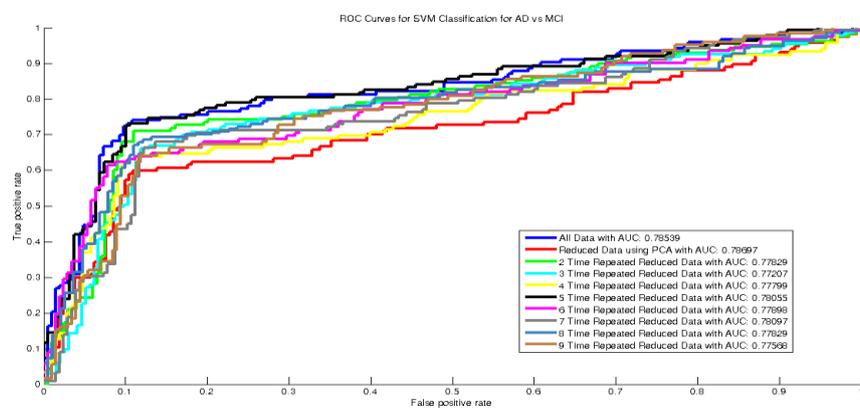
Figure 6. Comparison of changes in average accuracy, sensitivity and specificity for 15, 20, 25, and 30 PCA elements in 1–9× repeated feature reduction, and all the features (black line).

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

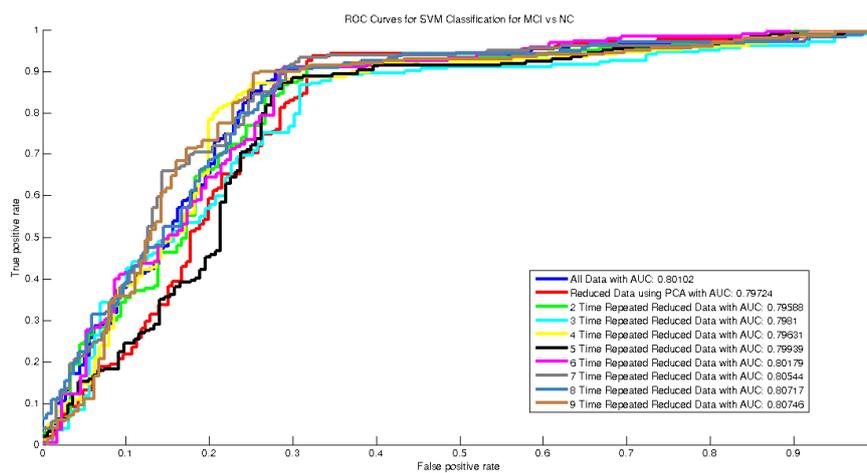
Data	Classes	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AUC
All Data	AD–NC	95.54	93.74	98.32	98.84	91.09	0.9577
	AD–MCI	81.41	89.02	68.09	82.99	78.00	0.7835
	MCI–NC	79.41	67.48	92.37	90.56	72.34	0.7993
Reduced Data Using PCA	AD–NC	97.20	95.46	99.86	99.90	93.53	0.9768
	AD–MCI	81.61	88.45	69.02	84.03	76.39	0.7846
	MCI–NC	79.45	67.20	92.96	91.33	71.97	0.8011
2 × 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 × 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 × 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 × 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 × 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 × 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 × 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 × 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	0.808



(a)



(b)



(c)

Figure 7. Receiver operating characteristic (ROC) curves for recognition of: (a) AD vs. NC, (b) AD vs. MCI, and (c) MCI vs. NC, of all features and 25 PCA elements for 1–9× repeated feature reduction.

In addition, we have added more explanatory experiments as Supplementary Material. There, we compare Homogeneous and Heterogeneous Emphasis Learning methods. The Homogeneous Emphasis Learning repeats all the selected features equally and homogeneously, and the Heterogeneous Emphasis Learning repeats selected features unequally and heterogeneously.

5. Discussion

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

Table 3. Comparison of the proposed method with other methods; including dataset and indicators. EL: Emphasis Learning. ADNI: Alzheimer’s Diseases Neuroimaging Initiative.

Method	Data type(s) (n, Dataset)	AD vs. NC				AD vs. MCI				MCI vs. NC			
		Acc%	Sen%	Spec%	AUC	Acc%	Sen%	Spec%	AUC	Acc%	Sen%	Spec%	AUC
Zhang et al., 2011 [45]	MRI, PET, CSF, MMSE, ADAS-Cog (202, ADNI)	93.20	93.00	93.30	0.98	-	-	-	-	76.40	81.80	66.00	0.81
Dai et al., 2013 [46]	MRI (83, OASIS)	90.81	92.59	90.33	0.94	85.92	82.46	87.59	0.87	81.92	78.51	88.34	0.81
J. Liu et al., 2016 [47]	MRI, PET (710, ADNI)	94.65	95.03	91.76	0.95	88.63	91.55	86.25	0.91	84.79	88.91	80.34	0.83
Beheshti et al., 2017 [24]	MRI (186, ADNI)	93.01	89.13	96.80	0.935	-	-	-	-	-	-	-	-
Mishra et al., 2018 [48]	MRI (417, ADNI)	89.15	85.06	92.53	0.93	-	-	-	-	-	-	-	-
Khedher et al., 2015 [49]	MRI (818, ADNI)	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 [50]	MRI (1457, ADNI)	90.00	82.00	97.00	0.95	-	-	-	-	-	-	-	-
Ben Ahmed et al., 2014 [51]	MRI (218, ADNI)	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 [52]	MRI (507, ADNI)	93.75	87.5	100	-	-	-	-	-	-	-	-	-
Suk et al., 2014 [53]	MRI, PET, CSF, MMSE, ADAS-Cog (202, ADNI)	93.05	90.86	94.57	0.95	88.98	82.11	90.65	0.90	83.67	96.79	57.28	0.82
Maqsood et al., [18]	MRI (392, OASIS)	89.66	100	82	-	-	-	-	-	-	-	-	-
Proposed Method (EL)	MRI, PET, CSF, MMSE (705, ADNI)	98.81	98.52	99.21	0.987	81.61	88.45	69.02	0.785	81.40	71.22	90.54	0.81

5.1. Feature Representation

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

In comparison with the original features, the proposed method greatly improved the diagnostic accuracy for all the considered classification problems. The proposed method outperformed the other methods in three binary classification problems. Using this method, notwithstanding the limited number of samples, helped reduce errors for classification problems and hence enhanced the classification accuracy. Previous methods can only use limited number of features in learning, but this limitation is overcome here.

There will not be an interpretation of the trained model or the feature representations. Each added unit in the input represents a 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 neuro-scientific perspective, the relations from MRI features and from FDG-PET features could be enhanced. This way, new and increased inputs of the high-level features represent their helpfulness in classifying patients and healthy normal controls. Using this method cannot help us to interpret or visualize the model's outputs, and this remains one of the unsolved pattern recognition and machine learning problems. In contrast, it is clear that this combined information will be useful in AD/MCI diagnosis.

5.2. Feature Reduction and Increasing—Feasibility of the Proposed Method

Here, we compare the results of the proposed method with PCA and the before-PCA results. In the feature set, we considered the clinical labels and clinical scores of MMSE. We observed that the method using increased specific numbers of feature packets outperformed others (Table 3). Here, we selected 15 to 30 PCA components explaining approximately 94.5% to 100% of the variance to test the proposed method.

The reason for the high performance of the proposed method can be explained as follows.

Consider learning perfect objects; when repeating them in the training process, the model can learn the object's features better and better. This 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 object features), the model can learn richer information about the objects. Similar to repeating objects after a specific number of repetitions, the model can be over-trained, and specifying the number of the repetitions is a precise action. Therefore, the method can make features that can accurately model the target values (i.e., labels and clinical scores and imaginary symptoms). This is the definition of repeating features in width (feature repetition or emphasized features) instead of length (sample repetition) in the algorithm. We can say the repeated features for the labels could discriminate AD and MCI patients from NC subjects.

As we said, the main idea of this paper is to repeat the most efficient features in classification. Our experiment showed that the theory of finding and repeating the good features of the data for classification problems will work excellently (Tables 2 and 3).

5.3. Classification Algorithm

SVM is a widely used algorithm in the area of Alzheimer's disease. We selected this algorithm with a linear kernel. Non-linear SVMs usually achieve better performance, but here in our tests, performance differences between the two types of kernel were not significant, and linear SVMs were faster to train, as shown in Table 4.

Table 4. Time complexity of SVM with linear kernel vs. Radial Basis Function (RBF) kernel.

Data	Classes	SVM Training Time—Linear Kernel (s)	SVM Training Time—RBF Kernel (s)	Linear—RBF (S)
All Data	AD–NC	5.5170	5.7599	−0.2429
	AD–MCI	6.7963	9.6903	−2.8941
	MCI–NC	6.8982	9.0480	−2.1498
Reduced Data Using PCA	AD–NC	4.5012	5.2301	−0.7289
	AD–MCI	5.7727	6.6944	−0.9217
	MCI–NC	7.3574	9.5005	−2.1432
3 × Reduced Data	AD–NC	5.0067	7.8505	−2.8437
	AD–MCI	6.3709	9.9687	−3.5978
	MCI–NC	8.1439	12.9038	−4.7599
5 × Reduced Data	AD–NC	5.9490	9.6567	−3.7077
	AD–MCI	7.3554	12.8005	−5.4451
	MCI–NC	11.9281	18.2044	−6.2763
9 × Reduced Data	AD–NC	8.7938	12.0661	−3.2722
	AD–MCI	10.0609	16.9106	−6.8497
	MCI–NC	14.0926	23.2651	−9.1726

As can be seen in Table 5, using all the data for the AD vs. NC task took 5.5170 seconds and after PCA it took 4.5012 s, increasing accuracy (from Table 2: accuracy increased from 95.54% to 97.20%) while reducing time. For this task, with five-fold repetition, the time was 5.9490 s while the accuracy was 98.81% (i.e., slight increase in the time complexity and a fair increase in the performance).

Table 5. Accuracy of AD, MCI, and NC classification, using each group of data, alone.

CLASSES	AD–NC ACC%	AD–MCI ACC%	MCI–NC ACC%
Personal Information	0.609	0.595	0.553
MMSE Data	0.919	0.785	0.703
MRI Data	0.868	0.684	0.697
CSF Data	0.594	0.524	0.643
PET Data	0.625	0.667	0.574

5.4. Comparison with the State-of-the-Art Methods

To validate the performance of the proposed method, we present the significance of the results 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 705 ADNI samples. The accuracy rates of our proposed method were 98.81%, 81.61%, and 81.40% for AD vs. NC, AD vs. MCI, and MCI vs. NC classification, respectively.

Most of the listed methods in Table 3 used the ADNI database, and we used all the images and data in ADNI, consisting of 705 participants. This is in contrast to some of the reported methods that used a portion of the ADNI samples. When the proportion of the used sample was sufficient compared to the entire sample, we compared our proposed method to theirs. As can be seen in 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, the sensitivity (recall rate) of the proposed method (98.52%) was the highest, as can be seen in Table 3. The proposed method stood in second place with a specificity of 99.21% against the perfect detection rate of Ben Ahmed et al. [51]. However, Ben Ahmed et al.'s method had a high false alarm rate while the proposed method succeeded in achieving a trade-off between the two. Note that lower specificity of our method can be cover by higher sensitivity, e.g., 94.83% of our method vs. 87%

in the case of Ben Ahmed et al.'s method. Nevertheless, combined methods can be used to achieve better performance.

In [23], 70 MRI images from the ADNI dataset were used for AD vs. NC classification—35 images of AD patients and 35 images of NC subjects. They achieved 100% accuracy in distinguishing between the two groups. Because the number of subjects used in our study was 10 times greater than this last study, it is not fair to compare this study with our proposed method. [18] achieved 89.66% accuracy in binary AD classification (recognizing AD from subjects) but they achieved 92.85% accuracy in multiple classification using transfer learning. Using only MRI data [24] could classify AD vs. NC with an accuracy rate of 93.01%. They achieved this performance using only 186 MR images. Using 785 MRI data [25] classified progressive MCI vs. static MCI with an accuracy of 86%, using combined MRI, APOe4 genetic data, and available clinical practice variables and cognitive measures including neuropsychological cognitive assessment tests like the Dementia Rating Scale (CDRSB), the Alzheimer's Disease Assessment Scale (ADAS11, ADAS13), episodic memory evaluations in the Rey Auditory Verbal Learning Test (RAVLT).

We could redress some imperfections of the proposed method using another method that has better performance. According to Table 3, in most of the performance measures the proposed method was in the first or second place, and we can see that the proposed method was dominant, and had the highest accuracy compared to other methods.

5.5. Limitations of the Work

The proposed method has some limitations. In PET imaging, the partial volume effect induced by a combination of image sampling and the restricted resolution of PET in the reconstructed images can bring under- or over valuation for regional radioactivity condensation. Therefore, more errors in statistical parametric images may occur [54]. Here, we must say that we could not download relevant PET images from ADNI and we only downloaded the extracted values of PETs from the ADNI database.

Combination of multiple tissue values would likely affect the differences between voxels of gray and white matters. Since our method is ROI-based feature selection, this partial voxel quality reduction would have an inconsiderable effect on the performance of the method.

We can say the proposed structure used to form the feature sets in this experiment could be non-optimal for other datasets. We need studies such as those learning 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, FDG-PET, MMSE scores, and CSF modalities into a single vector and repetition of the features after the feature reduction could efficiently distinguish between AD and MCI patients and NC subjects.

6. Conclusions

In this study, we proposed a simple but practical and effective method for classification, and tested it for Alzheimer's disease diagnosis. Our proposed method found the best features and repeated them until no further improvements to classification performance were obtained. We examined our method on the ADNI database of AD. The experiments showed that we could achieve much better performance using the combined features of MRI, MMSE, and personal information, especially when we repeated the reduced features on all three binary classification problems (i.e., AD vs. NC, AD vs. MCI, and MCI vs. NC). Experiments indicated the performance and effectiveness of the proposed method: accuracy rates of 98.81%, 81.61%, and 81.40% for AD vs. NC, AD vs. MCI, and MCI vs. NC classification problems, respectively. As can be seen, using this method increased the performance of the three binary problems incredibly. The results showed that the classification accuracy was improved with the optimized feature selection, which indicates that the information gain method can be used to select the more sensitive anatomical regions in AD and MCI diagnosis. Using other feature reduction

or selection methods and repeating reduced data could be the subject of a future work. Combining the results of other feature reduction and selection methods and establishing a classification framework, while using them, could be another future work. This study employed VBM8 tools, which yielded promising results. However, there are robust segmentation tools to explore (e.g., CAT12) that could be used to improve the diagnosis results. As another future work, applying this method to clustering can be recommended. Finally, features recommended by experts could be put through the proposed model in order to achieve better performance as a future work.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-8220/20/3/941/s1>.

Author Contributions: Conceptualization, H.A.; methodology, H.A.; software, H.A.; validation, H.A.; investigation, H.A.; resources, H.A.; writing—original draft preparation, H.A.; writing—review and editing, H.A.; visualization, H.A.; supervision, M.B., S.R., and A.R.R.; project administration, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to specifically thank to M Dashtban (Informatics, University of Reading, UK) for his insightful comments and continuous support to make this research accomplished. We are so appreciated for the precious and critical reviews of accredited reviewers and the editorial board who made our work much better. 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), and 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 stages of Alzheimer’s disease. The purpose of the determination of sensitive and specific biomarkers of early AD progression is to develop new treatments and monitor their effectiveness, and to reduce the time and cost of clinical experiments.

Conflicts of Interest: 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.

References

1. Thies, W.; Bleiler, L. 2012 Alzheimer’s disease facts and figures. *Alzheimer’s Dement. J. Alzheimer’s Assoc.* **2012**, *8*, 131–168. [[CrossRef](#)]
2. Davatzikos, C.; Fan, Y.; Wu, X.; Shen, D.; Resnick, S.M. Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging* **2008**, *29*, 514–523.
3. Kloppel, S.; Stonnington, C.M.; Chu, C.; Draganski, B.; Scahill, R.I.; Rohrer, J.D.; Fox, N.C.; Jack, C.R., Jr.; Ashburner, J.; Frackowiak, R.S. Automatic classification of mr scans in Alzheimer’s disease. *Brain* **2008**, *131*, 681–689. [[CrossRef](#)] [[PubMed](#)]
4. Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehericy, S.; Habert, M.O.; Chupin, M.; Benali, H.; Colliot, O. Automatic classification of patients with alzheimer’s disease from structural mri: A comparison of ten methods using the adni database. *Neuroimage* **2011**, *56*, 766–781. [[CrossRef](#)] [[PubMed](#)]
5. Foster, N.L.; Heidebrink, J.L.; Clark, C.M.; Jagust, W.J.; Arnold, S.E.; Barbas, N.R.; DeCarli, C.S.; Turner, R.S.; Koeppe, R.A.; Higdon, R.; et al. Fdg-pet improves accuracy in distinguishing frontotemporal dementia and alzheimer’s disease. *Brain* **2007**, *130*, 2616–2635. [[CrossRef](#)]
6. Nordberg, A.; Rinne, J.O.; Kadir, A.; Langstrom, B. The use of pet in Alzheimer disease. *Nat. Rev. Neurol.* **2010**, *6*, 78–87. [[CrossRef](#)]
7. Dubois, B.; Feldman, H.H.; Jacova, C.; DeKosky, S.T.; Barberger-Gateau, P.; Cummings, J.; Delacourte, A.; Galasko, D.; Gauthier, S.; Jicha, G.; et al. Research criteria for the diagnosis of Alzheimer’s disease: Revising the nincds–adrda criteria. *Lancet Neurol.* **2007**, *6*, 734–746. [[CrossRef](#)]
8. Hajian-Tilaki, K. Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Casp. J. Intern Med.* **2013**, *4*, 627–635.
9. Killiany, R.J.; Gomez-Isla, T.; Moss, M.; Kikinis, R.; Sandor, T.; Jolesz, F.; Tanzi, R.; Jones, K.; Hyman, B.T.; Albert, M.S. Use of structural magnetic resonance imaging to predict who will get alzheimer’s disease. *Ann. Neurol.* **2000**, *47*, 430–439. [[CrossRef](#)]
10. Dickerson, B.C.; Goncharova, I.; Sullivan, M.P.; Forchetti, C.; Wilson, R.S.; Bennett, D.A.; Beckett, L.A.; deToledo-Morrell, L. Mri-derived entorhinal and hippocampal atrophy in incipient and very mild alzheimer’s disease. *Neurobiol. Aging* **2001**, *22*, 747–754. [[CrossRef](#)]

11. Convit, A.; Asis, J.D.; Leon, M.J.D.; Tarshish, C.Y.; Santi, S.D.; Rusinek, H. Atrophy of the medial occipitotemporal, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiol. Aging* **2000**, *21*, 19–26. [[CrossRef](#)]
12. Chételat, G.; Desgranges, B.; Sayette, V.D.L.; Viader, F.; Eustache, F.; Baron, J. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport* **2002**, *13*, 1939–1943. [[CrossRef](#)] [[PubMed](#)]
13. Duin, R.P.W. Classifiers in almost empty spaces. In Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 3–7 September 2000; pp. 1–7.
14. Juottonen, K.; Laakso, M.P.; Partanen, K.; Soininen, H. Comparative mr analysis of the entorhinal cortex and hippocampus in diagnosing alzheimer disease. *Am. J. Neuroradiol.* **1999**, *20*, 139–144. [[PubMed](#)]
15. Morra, J.H.; Tu, Z.; Apostolova, L.G.; Green, A.E.; Avedissian, C.; Madsen, S.K.; Parikshak, N.; Toga, A.W.; Jack, C.R., Jr.; Schuff, N.; et al. Automated mapping of hippocampal atrophy in 1-year repeat mri data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *Neuroimage* **2009**, *45*, S3–S15. [[CrossRef](#)]
16. Colliot, O.; Chételat, G.; Chupin, M.; Desgranges, B.; Magnin, B.; Benali, H.; Dubois, B.; Garnero, L.; Eustache, F.; Lehericy, S. Discrimination between alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology* **2008**, *248*, 194–201. [[CrossRef](#)]
17. Kantarci, K. Magnetic resonance markers for early diagnosis and progression of Alzheimer's disease. *Expert Rev. Neurother.* **2005**, *5*, 663–670. [[CrossRef](#)]
18. Maqsood, M.; Nazir, F.; Khan, U.; Aadil, F. Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans. *Sensors* **2019**, *19*, 2645. [[CrossRef](#)]
19. Islam, K.; Damiati, S.; Sethi, J.; Suhail, A.; Pan, G. Development of a Label-Free Immunosensor for Clusterin Detection as an Alzheimer's Biomarker. *Sensors* **2018**, *18*, 308. [[CrossRef](#)]
20. Toro, C.A.O.; Gonzalo Martin, C. Supervoxels-Based Histon as a New Alzheimer's Disease Imaging Biomarker. *Sensors* **2018**, *18*, 1752. [[CrossRef](#)]
21. Garyfallou, G.Z.; Ketebu, O.; Sahin, S.; Mukaetova-Ladinska, E.B.; Catt, M.; Yu, E.H. Electrochemical Detection of Plasma Immunoglobulin as a Biomarker for Alzheimer's Disease. *Sensors* **2017**, *17*, 2464. [[CrossRef](#)]
22. Lahmiri, S. Image characterization by fractal descriptors in variational mode decomposition domain: Application to brain magnetic resonance. *Phys. A Stat. Mech. Its Appl.* **2016**, *456*, 235–243. [[CrossRef](#)]
23. Lahmiri, S.; Shmuel, A. Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease. *Biomed. Signal Process. Control* **2019**, *52*, 414–419. [[CrossRef](#)]
24. Beheshti, I.; Demirel, H.; Matsuda, H. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput. Biol. Med.* **2017**, *83*, 109–119. [[CrossRef](#)] [[PubMed](#)]
25. Spasov, S.; Passamonti, L.; Duggento, A.; Liò, P.; Toschi, N.; Initiative, A.S.D.N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* **2019**, *189*, 276–287. [[CrossRef](#)]
26. Catana, C.; Drzezga, A.; Heiss, W.D.; Rosen, B.R. Pet/mri for neurologic applications. *J. Nucl. Med.* **2012**, *53*, 1916–1925. [[CrossRef](#)]
27. Jack, C.R., Jr.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Weiner, M.W. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* **2008**, *27*, 685–691. [[CrossRef](#)]
28. Ashburner, J.; Ridgway, G.R. Symmetric diffeomorphic modeling of longitudinal structural MRI. *Front. Neurosci.* **2013**, *6*, 197. [[CrossRef](#)]
29. Farokhian, F.; Beheshti, I.; Sone, D.; Matsuda, H. Comparing CAT12 and VBM8 for detecting brain morphological abnormalities in temporal lobe epilepsy. *Front. Neurol.* **2017**, *8*, 428. [[CrossRef](#)]
30. Guo, X.; Wang, Z.; Li, K.; Li, Z.; Qi, Z.; Jin, Z.; Yao, L.; Chen, K. Voxel-based assessment of gray and white matter volumes in Alzheimer's disease. *Neurosci. Lett.* **2010**, *468*, 146–150. [[CrossRef](#)]

31. Davatzikos, C.; Bhatt, P.; Shaw, L.M.; Batmanghelich, K.N.; Trojanowski, J.Q. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. *Neurobiol. Aging* **2011**, *32*, 19–27. [[CrossRef](#)]
32. Hinrichs, C.; Singh, V.; Xu, G.; Johnson, S.C. Predictive markers for ad in a multimodality framework: An analysis of mci progression in the adni population. *Neuroimage* **2011**, *55*, 574–589. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, D.; Shen, D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease. *Neuroimage* **2012**, *59*, 895–907. [[CrossRef](#)] [[PubMed](#)]
34. Liu, F.; Suk, H.I.; Wee, C.Y.; Chen, H.; Shen, D. High-order graph matching based feature selection for alzheimer’s disease identification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer-Verlag: Berlin, Germany, 2013.
35. Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; Joliot, M. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* **2002**, *15*, 273–289. [[CrossRef](#)] [[PubMed](#)]
36. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
37. Fan, Y.; Resnick, S.M.; Wu, X.; Davatzikos, C. Structural and functional biomarkers of prodromal alzheimer’s disease: A high-dimensional pattern classification study. *Neuroimage* **2008**, *41*, 277–285. [[CrossRef](#)]
38. Mourao-Miranda, J.; Bokde, A.L.; Born, C.; Hampel, H.; Stetter, M. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage* **2005**, *28*, 980–995. [[CrossRef](#)]
39. Dashtban, M.; Li, W. Deep learning for predicting non-attendance in hospital outpatient appointments. In *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences (HICSS)*, Maui, HI, USA, 8–11 January 2019; pp. 3731–3740.
40. Dashtban, M.; Balafar, M. Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics* **2017**, *109*, 91–107. [[CrossRef](#)]
41. Dashtban, M.; Balafar, M.; Suravajhala, P. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. *Genomics* **2018**, *110*, 10–17. [[CrossRef](#)]
42. Burges, C.J.C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
43. Shawe-Taylor, J.; Cristianini, N. *Support Vector Machines and Other Kernel-Based, Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
44. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: New York, NY, USA, 2019.
45. Zhang, D.; Wang, Y.; Zhou, L.; Yuan, H.; Shen, D. Multimodal classification of Alzheimer’s disease and mild cognitive impairment. *Neuroimage* **2011**, *55*, 856–867. [[CrossRef](#)]
46. Dai, D.; He, H.; Vogelstein, J.T.; Hou, Z. Accurate prediction of AD patients using cortical thickness networks. *Mach. Vis. Appl.* **2013**, *24*, 1445–1457. [[CrossRef](#)]
47. Liu, J.; Li, M.; Lan, W.; Wu, F.-X.; Pan, Y.; Wang, J. Classification of Alzheimer’s disease using whole brain hierarchical network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *15*, 624–632. [[CrossRef](#)] [[PubMed](#)]
48. Mishra, S.; Beheshti, I.; Khanna, P.; Alzheimer’s Disease Neuroimaging Initiative. A statistical region selection and randomized volumetric features selection framework for early detection of Alzheimer’s disease. *Int. J. Imaging Syst. Technol.* **2018**, *28*, 302–314. [[CrossRef](#)]
49. Khedher, L.; Ramírez, J.; Górriz, J.M.; Brahim, A.; Segovia, F. Early diagnosis of Alzheimer’s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing* **2015**, *151*, 139–150. [[CrossRef](#)]
50. Lian, C.; Liu, M.; Zhang, J.; Shen, D. Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer’s Disease Diagnosis using Structural MRI. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [[CrossRef](#)] [[PubMed](#)]
51. Ben Ahmed, O.; Benois-Pineau, J.; Allard, M.; Ben Amar, C.; Catheline, G. Classification of Alzheimer’s disease subjects from MRI using hippocampal visual features. *Multimed. Tools Appl.* **2014**, *74*, 1249–1266. [[CrossRef](#)]
52. Zhou, K.; He, W.; Xu, Y.; Xiong, G.; Cai, J. Feature selection and transfer learning for alzheimer’s disease clinical diagnosis. *Appl. Sci.* **2018**, *8*, 1372. [[CrossRef](#)]

53. Suk, H.-I.; Lee, S.-W.; Shen, D.; Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage* **2014**, *101*, 569–582. [[CrossRef](#)]
54. Aston, J.A.; Cunningham, V.J.; Asselin, M.C.; Hammers, A.; Evans, A.C.; Gunn, R.N. Positron emission tomography partial volume correction: Estimation and algorithms. *J. Cereb. Blood Flow Metab.* **2002**, *22*, 1019–1034. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).