Alzheimer's Disease Detection using Neighborhood Components Analysis and Feature Selection

Mohamed Maher Ben Ismail[1], Reema Alabdullatif[2], Ouiem Bchir[3]

College of Computer and Information Sciences, King Saud University, Riyadh, KSA

Abstract—In this paper, we propose a Computer Aided Diagnosis (CAD) system in order to assist the physicians in the early detection of Alzheimer's Disease (AD) and ensure an effective diagnosis. The proposed framework is designed to be fully-automated upon the capture of the brain structure using Magnetic Resonance Imaging (MRI) scanners. The Voxel-Based Morphometry (VBM) analysis is a key element in the proposed detection process as it is intended to investigate the Gray Matter (GM) tissues in the captured MRI images. In other words, the feature extraction phase consists in encoding the voxel properties in the MRI images into numerical vectors. The resulting feature vectors are then fed into a Neighborhood Component Analysis and Feature Selection (NCFS) algorithm coupled with K-Nearest Neighbor (KNN) algorithm in order to learn a classification model able to recognize AD cases. The feature selection based on NCFS algorithm improved the overall classification performance.

Keywords—Alzheimer detection; classification; feature selection

I. INTRODUCTION

Alzheimer's Disease (AD) is an irreversible and progressive disorder that affects the brain of some elderly people. It destroys the memory and thinking skills and, eventually, the ability to perform simple tasks. Among older adults, it is the most common cause of dementia accounting for 60% to 80% of the cases. AD is ranked as the sixth cause of death in the United States, and moving to be in the third position based on recent statistic of the Alzheimer's Association [3]. In fact, the number of cases has drastically increased during the last 15 years. Currently, around 5 million Americans are diagnosed with AD and their number is expected to rise almost three-folds this number by 2050 as reported by the Alzheimer's Association [4]. According to the Saudi Alzheimer's Disease Association [3], there are 130,000 patients diagnosed with the disease in the kingdom. However, given the social norms and the lack of care systems in rural areas, these numbers represent probably a proportion of the actual cases.

Typically, dementia patients lose their cognitive functioning abilities such as thinking and remembering as well as their behavioral abilities. For the severe cases, AD patients become completely dependent on others in their basic daily activities. Besides, the serious physical and mental responsibilities that need to be taken by a caregiver, who is usually a family member, financial obligations are induced as well. In fact, The Alzheimer's Association [3] claims that AD's cost exceeds quarter of a trillion dollars nationwide.

Although Alzheimer's has no cure at the time being, clinical trials are still undergoing to develop medicines able to cure this disease. In fact, there are treatments which slow or delay the symptoms and maintain the mental function of the patients as reported by The National Institute on Aging [29]. For that, early diagnosis benefits the patients to sort out their life plans, financial situations as well as any legal issues while they are capable to do so mentally. Furthermore, on a nation level, upon early diagnosis, trillions of dollars could be saved on medical and care costs. However, AD cannot be definitely diagnosed except through autopsy using a microscope to examine the brain tissues. On the other hand, clinical assessment is performed to rule out other causes of dementia through appropriate tests as shown in Web MD [35]. One should note that more accurate diagnosis is made possible through the identification of AD biomarkers which can help in detecting it even before clinical symptoms are reported on patients. Namely, neuroimaging is an example of biomarkers that can capture changes in the brain without any invasive procedure as introduced in [11].

Several researchers around the globe took the interest in designing Computer Aided Diagnosis (CAD) systems to assist in the early detection of this serious disease and assure an effective diagnosis [1]. In particular, image based approaches have been coupled with machine learning techniques to address the AD detection challenge. These efforts have yielded promising results and have exhibited noticeable margin for improvement. Despite these efforts, the choice of the appropriate features and ML techniques remains a challenging open problem with a considerable room for improvement. In other words, the selection of *(i)* the appropriate visual descriptors used to encode the visual properties of the dataset and *(ii)* the supervised learning algorithm to map these resulting feature vectors into the positive or negative classes, can be investigated to improve the overall CAD performance. Such CAD system can be used in hospitals to assist doctors and radiologists in the interpretation of the relevant medical images. This would increase the accuracy of the clinical diagnosis.

In this paper, we propose to use standard features used to encode the biomarkers along with the Neighborhood Component Analysis and Feature Selection (NCFS) introduced in [15] as classification algorithm to improve the AD detection rate. More precisely, we intend to extract whole-brain atrophy features from structural MRI obtained from ADNI then apply the supervised NCAFS classifier to select a relevant subset of the highly-dimensional features to use it in k-nearest neighbor classification.

The rest the paper is organized as follows: Section II introduces the background of this work, Section III discusses an overview of the CAD researches in AD, Section IV describes our classification framework, and a summary and conclusion are presented in Section V.

II. RELATED WORKS

Different approaches and different classification methods are being investigated to introduce a reliable classification system that can detect AD in an automatic manner. Moreover, as the features that are extracted from such medical images are usually high-dimensional, suitable selection and reduction methods have been adopted to enhance the classification performance, address the curse of dimensionality, and reduce the time complexity.

In particular, the features extracted from structural images have been exploited in the design AD detections systems. Namely, the voxel-wise measures were commonly used when the intensities of the voxels were used as features. The voxels of choice are usually selected from the regions that exhibit differences between groups either found automatically by VBM or by prior knowledge of the anatomical regions affected by AD. For the later, pre-defined anatomically labeled atlases are used to locate those regions in the MR images. The authors in [9] relied on prior knowledge acquired by selecting 9 regions for which the voxels intensities were provided. Most of the selected regions were located in medial temporal lobe as its atrophy is considered a key biomarker. The number of features was then significantly reduced by pruning a Random Forest. To minimize the overfitting caused by the fact that the data includes more features than samples, multiple SVM classifiers were trained with random subsets from the pruned features and their dichotomic outcomes were averaged to create a classification index. On the other hand, the ROIs were specified on the basis of VBM analysis as outlined by the researchers in [5]. Accordingly, voxel values from the regions with decreased Gray Matter (GM) volume were used as raw features. The number of features was then reduced statistically using a probability distribution function that created a histogram of the intensity values. The optimal number of bins in the histograms was selected based on Fisher criterion maximization. As in [9], an SVM with linear kernel was used for the classification.

A feature extraction approach, similar to the one proposed in [6], was outlined in [7] along with a feature ranking method as feature selection approach. The approach aims to measure the relevance and the discriminative power of the visual features using a statistical t-test. A subset containing the top ranked features was then fed into a linear SVM learner for the classification task. The framework outlined in [23] relied also on the voxel values of GM map. However, the voxels of the whole GM tissue were acquired instead of specific ROIs. For those features, bottom-up hierarchical clustering was conducted to build a tree that illustrates the structural relationships among them. Then, the relationships captured by the tree were imposed on a sparse learning to determine the informative features. More than one brain tissue was incorporated in the development of the system introduced in [21]. Specifically, voxel intensities in both GM and white matter (WM) maps were saved as raw features followed by reduction by means of the statistical model Partial Least Square (PLS). PLS is similar to the famous Principal Component Analysis (PCA). However, it takes the class labels into account. Combining the features resulted in a slightly better performance when coupled with a linear SVM compared to GM based performance. The Regularized Logistic Regression (RLR) was adopted by the researchers in [8] to directly operate on the density map of GM instead of performing the reduction and classification at two steps in the voxel space. Scores were assigned by the RLR classifier based on conditional probabilities metrics that captures the similarity of the anatomical patterns found in a given individual to those in AD patients.

Instead of handling voxel values, the authors in [12] processed features further to obtain better visual data representation. Specifically, after VBM analysis, GM map generation and ROIs selection respectively, the texture features were extracted. Namely, the Gray Level Co-occurrence Matrix (GLCM) and Gabor filters were used for the texture feature extraction. On the other hand, an SVM Recursive Feature Elimination (SVM-RFE) was used along with a covariance metric to remove redundant data and consider the relevant features only. Thu SVM-RFE discarded the least significant features for classification in a backward sequential selection manner while the covariance was used to measure the correlation between the features. The measure of the cortical thickness was adopted in [10]. It was obtained by measuring the distances between the vertices on the meshes of the inner and the outer cortical surfaces. The feature vector of each subject encoding the thicknesses information was then transformed to the frequency space to remove noise. To accomplish that, a Manifold Harmonic Transform with eigen functions of the Laplace–Beltrami operator as the basis function of the transform was used. A cut-off to a certain number of eigenvectors was made which filtered the high frequency components while preserving the discriminating low frequency data. Principal Component Analysis (PCA) was then used followed by Linear Discriminant Analysis to transform the feature vectors into points in LDA space then find the axis that best separates the groups. In [13], the features were constructed as an ensemble of the following: the average cortical thickness values, the standard deviation in cortical thickness, the total surface area of the cortex, the volumes of cortical and WM ROIs. Note that a Logistic Regression with stability selection introduced in [24] yielded the best results when couple with a random forest classifier. Beside voxel values and cortical thickness, other methods have been explored in this area as well. Namely, the researchers in [17] experimented the Wavelet coefficients obtained using a Two-dimensional Discrete Wavelet Transform (2D-DWT) along with Haar wavelet function of level 3. On the other hand, the Principal Component Analysis (PCA) was adopted as a dimensionality reduction technique and the Normalized Mutual Information Feature Selection (NMIFS) that is derived from the minimum Redundancy Maximum Relevance (mRMR) was used as feature selection. Circular Harmonic Functions (CHFs) were utilized in [2] to extract local features of hippocampus and posterior cingulate cortex (PCC) from each slice of the 3D MRI. Those two ROIs

were selected for the noticeable harm the disease imposes on them. As depicted in [30], the features obtained from each ROI were then quantized using the bag-of-visual words approach by representing them as a histogram of occurrence of quantized visual features. The histograms of both ROIs were then combined in a single vector creating a signature for each subject. PCA was used for dimensionality reduction and the resulting features were finally fed into an SVM with the Radial Basis Function (RBF) as a kernel. Another approach introduced by the researchers consisted in fusing the features extracted from different modalities prior to learning the classification models. This was intended to exploit the complementary information provided by the different features to better capture the classes properties. Recently, the performance of deep learning on MRI and PET images was investigated in [22]. Automatic extraction of multi-level features was conducted by cascaded convolutional neural networks (CNN) on images from both modalities. The fully connected layer of the network was then followed by a softmax layer to perform the classification.

In [18], the authors investigated the connectivity of the hippocampus, precuneus and primary visual cortex and correlated each with the voxels of the brain. Sixteen AD patients and sixteen healthy controls were considered for this research. It was revealed that AD patients showed greater FC in left hippocampus with right insula. In [34], four seeds, right and left hippocampi and isthmus of the cingulate cortices (ICCs), which are parts within DMN were selected in a multi-class classification among them 10 AD, and 12 HC subjects. Pearson correlation coefficients were calculated between all possible pairs of the ROIs resulting in a 6-dimentional feature vectors. Then, to maximize the group differences and reduce noise, regularized LDA was applied to map the features into a one-dimensional sub-space. Finally, a decision tree was constructed and the classification was performed using AdaBoost ensemble learning.

For ICA, the researchers in [31-32] experimented 8 different types of FC measures and their variations to define the connections mostly related to the disease and might deliver better classification performance. Among these measures, matrices that record the connections between components obtained from ICA, the dynamics of these matrices found via a sliding window, and the graph properties of the matrices were used. For classification, Elastic net logistic regression was deployed to evaluate those measures individually. The FC dynamics variation outperformed the other measures. On the other hand, a new metric derived directly from the rs-fmri signal instead of FC measure was depicted in [20]. ICA was conducted to decompose the signals acquired from 15 AD and 15 healthy elderlies into their spatial components and their weights of time-courses. Then, a goodness-of-fit (GoF) calculation, template matching and SVM classification were applied to identify the neuronal components among the decomposed ones. Next, a brain activity map was constructed based on these components by computations involving the BOLD signals amplitudes and their standard deviations. Hippocampus and accumbens were selected separately as part of several experiments to train a linear SVM model. The rs-FMRI data obtained from ADNI as used to classify 20 healthy

subjects and 20 patients with AD in [19] using a linear SVM model. The automated anatomical labeling (AAL) atlas, a software package and digital atlas of the human brain, was used to divide the whole brain into 90 distinct regions and construct a graph with the regions as nodes. The signal of each node was computed by averaging the time series of voxels in each region to represent them and Pearson's correlation coefficients were employed to define the edges of the functional connectivity network. Graph metrics were then computed and used as discriminative features after selecting the optimal subset via Fisher score feature selection algorithm.

III. PROPOSED METHOD

This research is intended to design and implement a reliable CAD system for automatic detection of Alzheimer's disease (AD) based on MR images. Specifically, typical features are extracted to encode the visual properties of the patient MRI image. Then, the mapping between these features and the pre-defined class values (AD or Not-AD) is learned in a supervised manner. In other word, a supervised learning task is carried out by training a classification model using the annotated features extracted from the MR images. The resulting model is then intended to predict the class label of any unannotated MRI image features. Typically, the available data is divided into training and testing sets. Each image from the training set is then processed to extract a new representation vector that encapsulates its visual properties. Thus, a matrix containing all the training vectors is then fed into the classification algorithm along with their corresponding labels to learn a model for this specific problem. The system is evaluated with the testing set by comparing the labels predicted by the learned model with the ground truth labels. Though feature selection is not typically a key element, it is considered an important step in cases of high-dimensional vectors as in AD classification problems.

NCFS is an embedded feature selection method which aims to find a weighting vector w for the features. In fact, the optimal weight vector is the one that maximizes the leave-one-out classification of nearest neighbor algorithm based on a gradient ascent technique. Let $T = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)\}$ be a set of N training samples where x_i is a d-dimensional feature vector and $y_i \in \{1, \dots, C\}$ is its corresponding class label. The weighted distance between two samples x_i and x_j is obtained using:

$$D_w(x_i, x_j) = \sum_{l=1}^{d} w_l^2 |x_{il} - x_{jl}| \qquad (1)$$

where w_l is the weight of the lth feature in the vectors of the sample points. The Gradient ascent algorithm relies on differentiable functions to find the optimal solution. However, the function for selecting the nearest neighbors as a reference point for classification is non-differentiable, thus a probability distribution is used as an approximation to select the reference point. Equation (2) calculates the probability that x_i selects x_j as the reference point.

$$p_{ij} = \begin{cases} \frac{k(D_w(x_i, x_j))}{\sum_{m \neq i} k(D_w(x_i, x_m))} & if \ i \neq j \\ 0, & if \ i \neq j \end{cases} \qquad (2)$$

where $\kappa(z) = exp(-z/\sigma)$ is a kernel function and σ is the kernel width which is a parameter between 0 till $+\infty$ that affect the selection probability of each point in the set; i.e. if $\sigma \rightarrow 0$ only nearest neighbors can be selected and if $\sigma \rightarrow +\infty$ all the samples have the same chance.

The probability that a sample x_i is correctly classified can be calculated as:

$$p_i = \sum_j y_{ij}p_{ij} \tag{3}$$

with $y_{ij} = 1$ iff $y_i = y_j$ and $y_{ij} = 0$ otherwise.

The approximate leave-one-out classification accuracy can be found for a particular weighting vector as follows:

$$\xi(w) = \frac{1}{N}\sum_i p_i = \frac{1}{N}\sum_i \sum_j y_{ij}p_{ij} \tag{4}$$

A regularization parameter $\lambda > 0$ that can be tuned via cross validation is also included in the object function to reduce overfitting and accomplishing the feature selection by driving many of the weights in w to 0:

The regularization term formulated as:

$$\xi(w) = \sum_i \sum_j y_{ij}p_{ij} - \lambda \sum_{l=1}^d w_l^2 \tag{5}$$

The derivative of the object function $\xi(w)$ with respect to w_l is computed as:

$$\frac{\partial \xi(w)}{\partial w_l} = \sum_i \sum_j y_{ij}\left[\frac{2}{\sigma}p_{ij}\left(\sum_{m \neq i}p_{im}|x_{il} - x_{ml}| - |x_{il} - x_{jl}|\right)w_l\right] - 2\lambda w_l$$

$$= \frac{2}{\sigma}\sum_i \left(p_i\sum_{m \neq i}p_{im}|x_{il} - x_{ml}| - \sum_j y_{ij}p_{ij}|x_{il} - x_{jl}|\right)w_l - 2\lambda w_l$$

$$= 2\left(\frac{1}{\sigma}\sum_i(p_i\sum_{j \neq i}p_{ij}|x_{il} - x_{jl}| - \sum_j y_{ij}p_{ij}|x_{il} - x_{jl}|) - \lambda\right)w_l \tag{6}$$

This derivative of the objective function yields the update equation for the gradient ascent. Algorithm 1 is a pseudocode of the neighborhood component feature selection procedure.

Algorithm 1: Neighborhood Component Feature Selection

NCFS(T , α, σ, λ, η) ◁ T : training set, α:initial step length, σ: kernel width, λ: regularization parameter, η: small positive constant;
Initialization: $w^{(0)} = (1,1,...,1)$, $\epsilon^{(t)} = -\infty$, t = 0
repeat
*for i= 1,··· , N **do***
Compute p_{ij} and p_i using $w^{(t)}$ according to (2) and (3)
*for l= 1,··· , d **do***
$\Delta_l = 2\left(\frac{1}{\sigma}\sum_i(p_i\sum_{j \neq i}p_{ij}|x_{il} - x_{jl}| - \sum_j y_{ij}p_{ij}|x_{il} - x_{jl}|) - \lambda\right)w_l^{(t)}$
t = t + 1
$w^{(t)} = w^{(t-1)} + \alpha\Delta$
$\epsilon^{(t)} = \xi(w^{(t-1)})$
If** $\epsilon^{(t)} > \epsilon^{(t-1)}$ **then
α= 1.01α
else
α= 0.4α
***until** $|\epsilon^{(t)} - \epsilon^{(t-1)}| < \eta$*
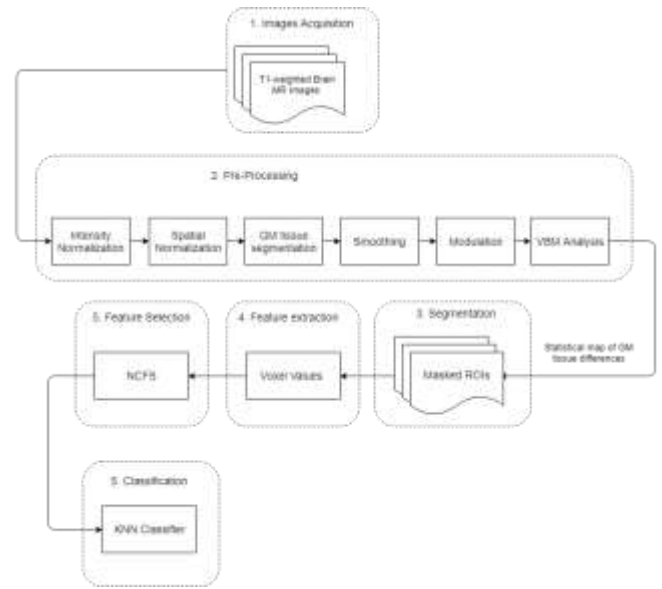*$w = w^{(t)}$*
***return** w*

Fig. 1. Detailed Schema of the Proposed CAD System.

Fig. 1 depicts the main components of the proposed system.

This research relies on the images acquired using structural MRI for their ability to produce high resolution images without any injected substance as well as their potential to capture the structure of brain. The atrophies that are a key symptom of AD pathology can be detected using this modality. The T1-weighted sequence is generally used for AD scans as stated in [14] and so will be the images in this framework. Before employing the images in creating a classification model, pre-processing steps need to be carried out. T1-weighted images might exhibit non-uniform intensities throughout the brain caused by a low-frequency smooth signal known as bias field. Bias field is introduced to the scans due to the heterogeneous magnetic field of MRI scanners. This affects the images by blurring them and consequently reducing their high frequency content such as edges besides changing the gray level distribution of tissues from the same class as proved in [27]. Therefore, MR images were corrected for this flaw prior to subsequent processing. Fig. 2 shows the bias field effect on the MR images. N4 intensity normalization method outlined in [33] as a variant of N3 algorithm known for bias correction in medical images, was also used for this purpose as suggested by ADNI. However, in order to produce precise statistics, images need to be aligned to reduce the variability between individuals since people have different sized and shaped brain structures. Spatial normalization (or registration) is the method that achieves that by transforming MR image of each individual to a reference frame called template.

This process is guided by an atlas which locates the position of the different anatomical regions in the template space. Namely, we relied on the MNI templates provided by The Montreal Neurological Institute [28]. Specifically, we adopted their ICBM152 standard from The McConnel Brain Imaging Centre [27]. After the pre-processing phase suggested in [25-26], instead of pre-selecting ROIs from GM tissue to

extract their features, a whole brain analysis is performed to find all the areas that are different between AD patients and healthy elderlies. Following VBM analysis and highlighting the areas that differ between the two groups, these areas are selected as ROIs. The considered features are the voxel values (or intensities) belonging to these ROIs. The algorithm NCFS is then applied to the training set to rank the extracted features according to their amount of positive influence on a *KNN* classifier in a leave-one-out classification. A subset containing the highly ranked features will then be selected as the representing features to be used while training the classification model.
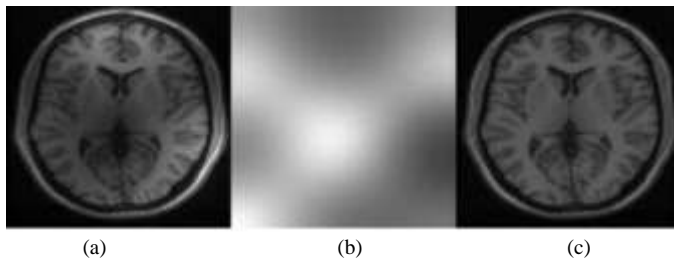

(a)   (b)   (c)

Fig. 2.   (a) Bias Field effect on a Slice of MR Image (b) Bias Field Frequency (c) Corrected Image.

IV.   EXPERIMENTS

We conducted our experiments on a collection MRI images which includes 100 AD cases versus 100 healthy instances. This dataset was provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI 2020). In particular, we compared the classification performance obtained using various classification methods. Specifically, some of the learned models were built using the original set of features, while others were built after some feature selection. Namely, the algorithms we used in these experiments are KNN classifier and SVM classifier (with two kernels: linear and RBF). The rationales behind this choice are: *(i)* NCFS is based on ranking the feature relevance according to their contribution in maximizing KNN classification performance. In other words, we intended to investigate whether selecting the top ranked features improves the performance of Alzheimer' disease detection. *(ii)* SVM is widely used in related CAD systems and testing it on the designed features would provide a better perception on the validity of the proposed system compared to state of the art solutions. Furthermore, two other existing methods were tested on the same dataset and environment in order to provide a good

ground for comparison. Specifically, the first method is based on PCA features reduction while the other relies on t-tests to sort the features and select the optimal ones.

A. VBM Analysis

After preprocessing the data, the gray matter tissue maps were analyzed with VBM to detect the atrophy regions that differentiated AD from HC group. For this particular analysis of the selected samples from AD and HC groups, three significant clusters which include different number of voxels were obtained. Fig. 3 illustrates the locations of these clusters while Table I provides their details. The values of the voxels belonging to the resulting clusters were then extracted from each sample then combined into one feature vector. This means that the feature vector is a concatenation of the voxels values from the three clusters per sample. In the following, we refer to this vector as raw features.

B. Optimization of SVM Hyper-Parameters

As depicted in [16], the SVM models rely on finding the optimal hyperplane which maximizes the margin between two classes. The optimization process is formulated using an objective function and resolved using typical mathematical optimization methods. This function is controlled by the parameters known as hyperparameters. Their settings affect the performance of the learning process. Therefore, these parameters need to be tuned to find the set of values which is optimal for solving a specific learning problem. Two kernels were investigated in this project. Namely, we used the linear and RBF kernels. For RBF kernel, a couple of parameters need to be optimized: the cost or regularization term and gamma or kernel width. The cost controls the trade-off between the misclassifications of training samples and width of the margin while the gamma controls the tradeoff between under-fitting loss and over-fitting loss. Both parameters were optimized in this work using Bayesian optimizer with 10-fold cross validation for every different set of features.

C. Optimization of Lambda in NCFS

A tuning via cross validation was conducted on a range of value starting from zero, where no regularization was enforced, to optimize the regularization parameter lambda. In each fold, the lambda values were used and the loss was assessed with Mean Squared Error (MSE). Fig. 4 reports the effect of choosing the appropriate regularization value showing the worst performance when the value was set to zero and an improved performance by more than 10% with 0.018.
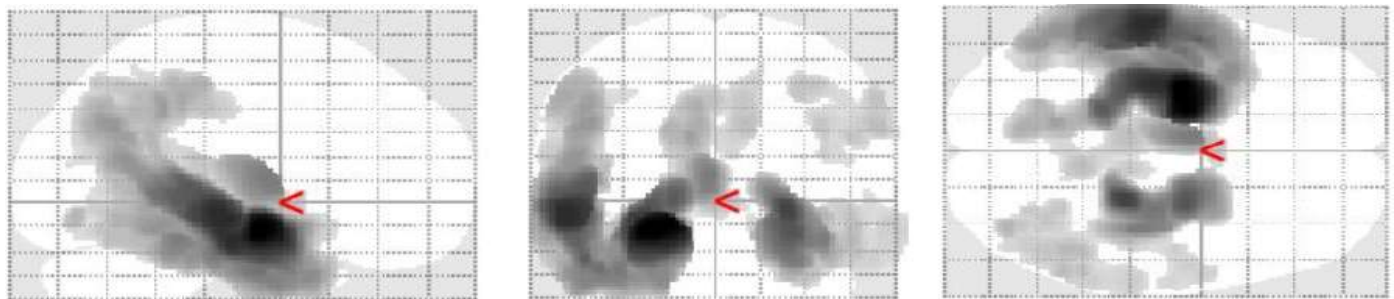

Fig. 3.   Glass view of Atrophy Clusters Highlighted by VBM Analysis where k = 1400.

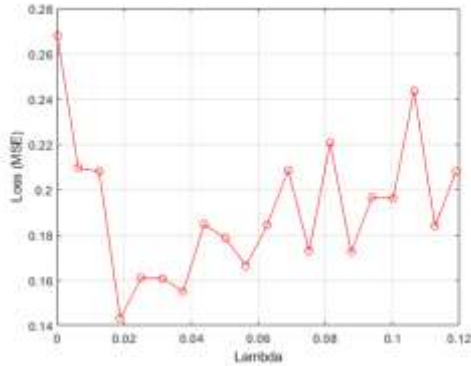| Cluster ID	MNI coordinates	K_E (voxels quantity)	T-value (peak-level)
1	-26, -9, -14	36607	14.51
2	24, -32, -6	8748	11.46
3	59, -56, 30	1824	6.35

Fig. 4.   Loss Results of Tuning the NCFS Regularization Parameter Lambda Measured by MSE.

D. Feature Weighting using NCFS

After optimizing NCFS parameters, the best value was used in the algorithm to rank the raw features. Then, to reduce the dimensionality of the ranked features, a threshold (or cutoff) was determined by iteratively adjusting the number of selected features and feed them to an SVM classifier where its performance is evaluated with 10-fold cross validation. The adjustment started by taking a wider range of values and large step size in order to have a rough estimation where the optimal number is located. After that, in the span that had the highest accuracy, a second adjustment was repeated within its bound with smaller step size as shown in Fig. 5.

E. T-Test based Feature Ranking

As previously mentioned, another type of feature selection was tested in this work. The raw features were ranked using T-test method according to their T-values. For selecting the optimal number of features, the same method used with NCFS features was applied here. The corresponding results are reported in Fig. 6. Though the authors in [7] proposed using fisher criterion to automatically determine the threshold, following our method in determining it won't have major effect as their classification model is also SVM which we use in the evaluation of the optimal number.

F. Features Reduction using PCA

The high dimensionality of raw features was reduced using PCA method. PCA was chosen in our comparative analysis for its effectiveness in significantly reducing the number of features and also being compared to t-test features. Similar to the previous work, using 10-fold cross validation, the orthogonal principal components of the raw features and training samples were extracted in every fold to train the models. The number of components was equal to the number of training samples and they were used directly as the features;

that resulted in a vector with 200 dimensions in our case. However, to improve the PCA performance we only selected the components that retained 95% of data information which further reduced the dimensions to only 75. The improvement in classification is described in Table II.

G. Results obtained using SVM Models

In our experiments, every feature set was used to train and test three classification models. More specifically, we used KNN along with a linear-kernel SVM and a Gaussian RBF SVM. First, we present the results obtained using SVM classifiers.

Table III reports the accuracy, sensitivity, specificity and AUC measures attained using the two SVM classifiers associated with the considered features. As it can be seen, linear SVM yielded noticeably higher accuracy. In particular, it resulted in a 4.5% increase for the raw features while the others did not exhibit large differences in their accuracies as they did not go above 1%. On the other hand, from the sensitivity perspective, which is considered an important measure in this study and the medical field in general, SVM with RBF kernel attained 4.5% improvement in the classification of AD patients with the t-test features compared to the linear kernel. NCFS and PCA features remained rather consistent across the two models but giving better overall performance with the linear kernel. Raw and PCA features scored the highest accuracy among the other features in the linear model and PCA in the RBF model.
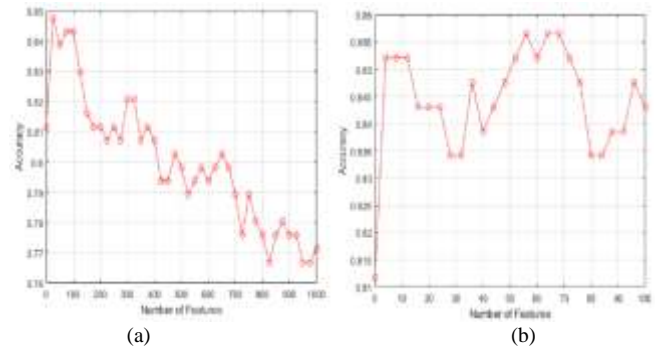

(a)   (b)

Fig. 5.   The Accuracies from Training an SVM Model with NCFS Features of different Dimensions. (a) The First Adjustment Results with Higher Range and Step Size (b) The Second Adjustment within the Bound of the Highest Accuracy Span and Small Step Size.
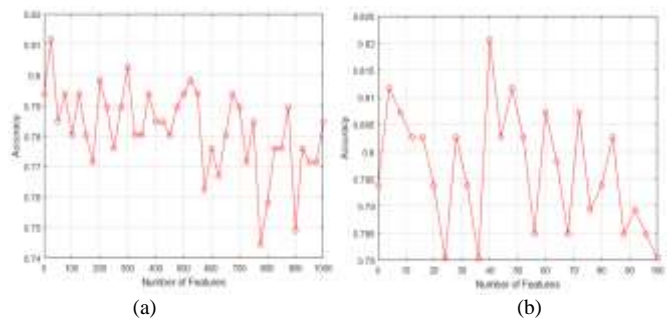

(a)   (b)

Fig. 6.   The Accuracies from Training an SVM Model with T-Test Features of different Dimensions. (a) The First Adjustment Results with Higher Range and Step Size (b) The Second Adjustment within the Bound of the Highest Accuracy Span and Small Step Size; Threshold = 40.

TABLE II.   DIFFERENCES IN CLASSIFICATION RESULTS OF PCA WITH ALL THE COMPONENTS AND THE COMPONENTS WITH 95% PRESERVE

	Linear SVM			
Features	**Acc %**	**Sen %**	**Spe %**	**AUC %**
PCA-All	77.58	73.15	81.74	85.83
PCA	87.00	83.33	90.43	92.38

TABLE III.   CLASSIFICATION RESULTS OF SVM MODELS ON EACH TYPE OF FEATURE

	Linear SVM				RBF SVM			
Features	Acc %	Sen %	Spe %	AUC %	Acc %	Sen %	Spe %	AUC %
Raw	87.0	87.0	86.9	93.3	82.5	78.7	86.0	91.2
NCFS	84.3	82.4	86.0	91.5	83.8	79.6	87.8	90.2
T-test	81.6	75.0	87.8	88.7	82.5	79.6	85.2	85.0
PCA	87.0	83.3	90.4	92.3	86.1	82.4	89.5	91.7

Fig. 7 and Fig. 8 depict the ROC curves produced by plotting the true and false positive rates from each classifier. The higher the curve is, the larger AUC value which measures the area under this curve and the better the classifier. T-test features showed the least AUC values with RBF kernel which can be perceived in how low its curve is in Fig. 8 compared to the other curves. The unexpected performance of the raw features can be attributed to two facts. The first one is that these features are a result of a statistical analysis that discriminates between the two groups. This yields a fair separability between the data instances from the two classes. The second fact is that kernel SVM is a powerful classifier when the hyperparameters are optimized and can handle efficiently highly dimensional data.

H. Results Obtained using KNN

The classification results obtained using KNN yielded different but predictable results. The Manhattan distance (or city block) was used as the distance metric to decide the nearest neighbors of any given sample. As it can be seen in Table IV, the number of neighbors K was set to be 3, 5 and 7 respectively to assess its effect on the performance. NCFS features resulted in the highest performance measures and ROC curve, followed by t-test features as confirmed by Fig. 9. As NCFS ranked the features based on their contribution in maximizing KNN classification, this had an advantage of giving it a classification power similar to SVM classifier. Unfortunately, the sensitivity dropped by about 5% compared to SVM results. This means that more AD instances were misclassified in all number of neighbors variations and could indicate that the initial features weren't discriminative enough. Likewise, the performance with raw features declined in this round in all the three settings. Unlike SVM, KNN doesn't have the ability to deal with the high dimensionality in data which in turn affected the distance by the irrelevant features.
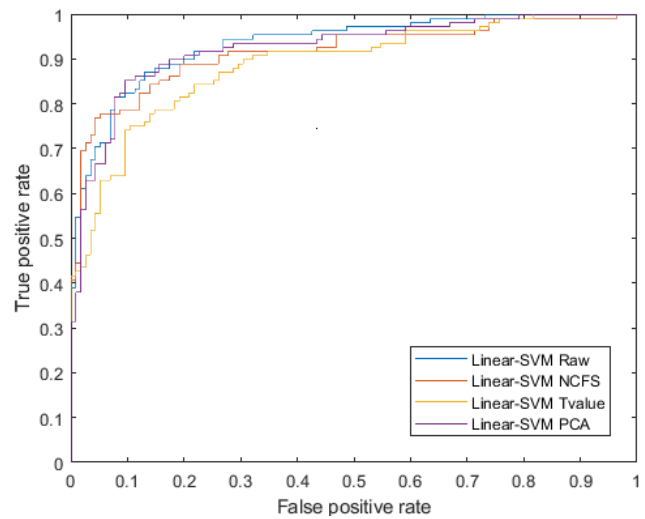

Fig. 7.   ROC Curves for Linear SVM Classifications with the different Features Types.
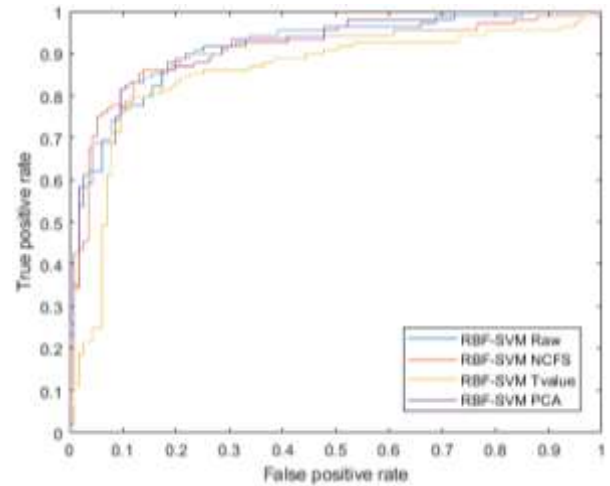

Fig. 8.   ROC Curves for RBF-SVM Classifications with the different Features Types.
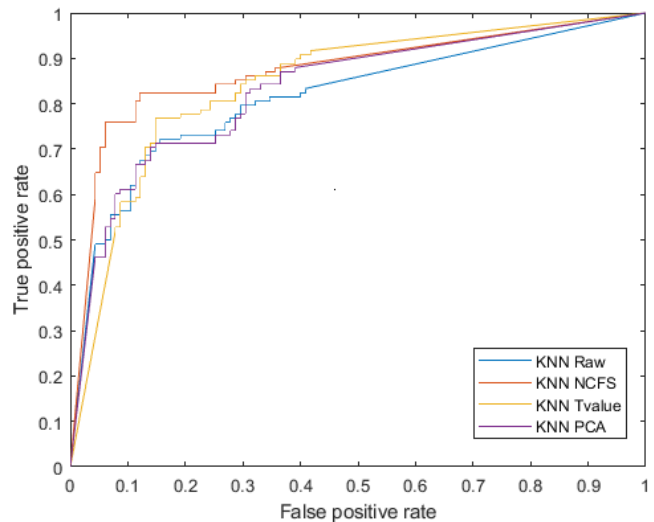

Fig. 9.   ROC Curves for KNN Classifications with the different Features Types.

TABLE IV.   CLASSIFICATION RESULTS OF KNN MODELS ON EACH TYPE OF FEATURE WITH DIFFERENT NUMBER OF NEIGHBORS K

	K = 3					K = 5					K = 7			
Features	Acc %	Sen %	Spe %	AUC %		Acc %	Sen %	Spe %	AUC %		Acc %	Sen %	Spe %	AUC %
Raw	78.5	72.2	84.4	81.4		78.0	66.7	88.7	83.9		78.9	68.5	88.7	85.0
NCFS	83.0	76.9	88.7	87.2		81.6	77.8	85.2	88.6		83.0	78.7	87.0	89.2
T-test	80.3	76.9	83.5	84.4		80.7	75.9	85.2	86.8		82.5	76.9	87.8	87.8
PCA	77.6	66.7	87.8	83.1		79.8	69.4	89.6	85.0		80.3	70.4	89.6	85.3

V. CONCLUSIONS

In this paper, we proposed a fully-automated CAD system able to detect AD cases. A detailed description of the different component of the proposed system was provided. In particular, the Neighborhood Component Analysis and Feature Selection (NCFS) approach was combined with K-Nearest Neighbor (KNN) as the supervised learning algorithm intended to perform feature selection and AD detection. The obtained results confirmed the effectiveness of the proposed unsupervised learning approach as well as the discriminative power of the features used to encode the image visual properties.

REFERENCES

[1] Abásolo, D., Hornero, R. and Espino, P. (2009) Approximate Entropy of EEG Background Activity in Alzheimer's Disease Patients, Intelligent Automation & Soft Computing, 15(4), 591-603.

[2] Ahmed, O. B., Mizotin, M., Benois-Pineau, J., Allard, M., Catheline, G. and Amar, C. B. (2015) "Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex," Computerized Medical Imaging and Graphics, 44, 13-25.

[3] Alzheimer's Association 2020. Available: https://www.alz.org/ alzheimers-dementia/what-is-Alzheimers.

[4] Alzheimer's Association, 2019. Available: https://www.alz.org/ alzheimers-dementia/facts-figures.

[5] Beheshti, I. and Demirel, H. (2015). "Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease," Computers in Biology and Medicine, 64, 208-216.

[6] Beheshti, I. and Demirel, H. (2016). "Feature-ranking-based Alzheimer's disease classification from structural MRI," Magnetic resonance imaging, 34(3), 252-263.

[7] Casanova, R., Hsu, F., Sink, K., Rapp, S., Williamson, J. and Resnick, S. (2013). "Alzheimer's disease risk assessment using large-scale machine learning methods," PLOS ONE, 8(11).

[8] Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti R. (2011). "Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease," Neuroimage, 58(2), 469-480.

[9] Cho, Y., Seong, J.-K., Jeong, Y. and Shin, S. Y. (2012). "Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data," Neuroimage, 59(3), 2217-2230.

[10] Craig-Schapiro, R., Fagan, A. M. and Holtzman, D. M. (2009). "Biomarkers of Alzheimer's disease," Neurobiology of Disease, 35, 128 - 140.

[11] Ding, Y., Zhang, C., Lan, T., Qin, Z. Zhang, X. and Wang, W. (2015). "Classification of Alzheimer's disease based on the combination of morphometric feature and texture feature," in IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC.

[12] Dubey, R., Zhou, J., Wang, Y., Thompson, P. M. and Ye, J. (2014). "Analysis of sampling techniques for imbalanced data: An n= 648 ADNI study," NeuroImage, 87, 220-241.

[13] Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P. and Thompson, P. M. (2010). "The clinical use of structural MRI in Alzheimer disease," Nature Reviews Neurology, 6(2), 67.

[14] Goldberger, J., Hinton, G. E., Roweis, S. T. and Salakhutdinov, R. R. (2005). "Neighbourhood components analysis," Advances in neural information processing systems, 513--520.

[15] Han, F., Lei, M., Zhao, W., and Yang, J. (2012). New Support Vector Machine for Imbalance Data Classification. Intelligent Automation & Soft Computing, 18(6), 679–686.

[16] Herrera, L. J., Rojas, I., Pomares, H., Guillén, A., Valenzuela, O. and Baños, O. (2013). "Classification of MRI Images for Alzheimer's Disease Detection," in International Conference on Social Computing, Alexandria, VA, USA.

[17] Kenny. E. R., Blamire. A. M., Firbank. M. J. and John. T. (2011). "Functional connectivity in cortical regions in dementia with Lewy bodies and Alzheimer's disease," a journal of neurology, 135, 569-81.

[18] Khazaee, A., Ebrahimzadeh, A. and Babajani-Feremi, A. (2015). "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory," Clinical Neurophysiology, 126, 2132-2141.

[19] Kazemifar, S., Manning, K. Y., Rajakumar, N., Gómez, F. A., Soddu, A, Borrie, M. J., Menon, R. S. and Bartha, R. (2017). "Spontaneous low frequency BOLD signal variations from resting-state fMRI are decreased in Alzheimer disease," PLoS One, 12(6).

[20] Khedher, L., Ramirez, J., Gorriz, J. M., Brahim, A. and 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, 139-150.

[21] Liu, M., Cheng, D., Wang, K. and Wang, Y. (2018). "Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis," Neuroinformatics, 16(3), 295-308.

[22] Liu, M., Zhang, D. and Shen, D. (2013). "Identifying informative imaging biomarkers via tree structured sparse learning for AD diagnosis," Neuroinformatics, 12(3), 381-394.

[23] Meinshausen, N. and Buhlmann, P. (2010). "Stability selection," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(4), 417-473.

[24] Mikl, M., Marecek, R., Hlustik, P., Pavlicova, M., Drastich, A., Chlebus, P., Brazdil, M. and Krupa, P. (2008). "Effects of spatial smoothing on fMRI group inferences," Magnetic resonance imaging, 26(4), 490-503.

[25] Poldrack, R. A., Mumford, J. A. and Nichols, T. E. (2011). "Spatial normalization," in Handbook of Functional MRI Data Analysis, Cambridge University Press, 53-69.

[26] Poldrack, RA. 2007. Region of interest analysis for fMRI. Soc Cogn Affect Neurosci, 2(1), 67–70.

[27] The McConnel Brain Imaging Centre (2020), Available: http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009.

[28] The Montreal Neurological Institute (2020). "Montreal Neurological Institute and Hospital,". Available: https://www.mcgil l.ca/neuro/.

[29] The National Institute on Aging, 2020. Available: https://www.nia .nih.gov/health/alzheimers/treatment.

[30] The Saudi Alzheimer's Disease Association, 2020. http://alz.org.sa/ ?page_id=2187.

[31] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. and Gee, J. C. (2010). "N4ITK: improved N3 bias correction," IEEE transactions on medical imaging, 29(6), 1310-1320.

[32] Urka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C. (2004). "visual categorization with bags of keypoints," In: ECCV workshop on statistical learning in computer vision, 1, 1-22.

[33] Vos, F. d., Koini, M., Schouten, T. M., Seiler, S., Grond, J. v. d., Lechner, A., Schmidt, R., Rooij, M. d. and Rombouts, S. A. (2018). "A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease," NeuroImage, 167, 62-72.

[34] Wang, Z., Zheng, Y., Zhu, D. C., Bozoki, A. C. and Li, T. (2018). "Classification of Alzheimer's Disease, Mild Cognitive Impairment and Normal Control Subjects Using Resting-State fMRI Based Network Connectivity Analysis," IEEE Journal of Translational Engineering in Health and Medicine, 6, 1-9.

[35] Web MD, 2020. Available: https://www.webmd.com/alzheimers/ guide/making-diagnosis-tests#1.