Dimension and Complexity Study for Alzheimer's disease Feature Extraction

Mohamed M. Dessouky[1], Mohamed A. Elrashidy[1], Taha E. Taha[2], and Hatem M. Abdelkader[3]

[1]Menoufiya University, Faculty of Electronic Engineering,
Department of Computer Science and Engineering, Egypt
Mohamedmoawed2003@hotmail.com

[1]Menoufiya University, Faculty of Electronic Engineering,
Department of Electronics and Electrical Communications, Egypt.
taha117@hotmail.com

[1]Menoufiya University, Faculty of Computers and Information, Egypt
Hatem6803@yahoo..com

***Abstract:** This paper discusses the problem of dimensionality and computational complexity analysis for feature extraction proposed algorithms of Alzheimer's disease. The effective features are very useful for some of the discrimations and to assist the physicians in the detection of abnormalities. This paper concern two main issues that must be confronted which are: The first one concern the study of how the classification accuracy depends on the dimensionality (i.e. the number of features). The second issue is the computational complexity of designing the classifier. As the number of features increases, the classification error decreases which consequently improve the accuracy of the classifier.*

Keywords: Diagnosis, Alzheimer's disease, Computer Aided Diagnosis, Feature Extraction, Dimensionality, Complexity Analysis, Feature reduction, and Support Vector Machine.

1. Introduction

Dementia is a loss of brain function that occurs with certain diseases. Alzheimer's Disease (AD) is one of the most common term of dementia that gradually get worse over time which affects memory, thinking and behavior. It is a neurodegenerative disorder which is happened to old people. Till now, there is no treatment for AD. The problem that AD is discovered at late stage, which the cure slows the impairment. So, the early diagnosis of the AD helps in improving the treatment [1, 2].

Brain images can be used as a sign of the AD. There are different types of brain images that could be used in AD discovery like the cerebrospinal fluid (CSF), positron emission tomography (PET) and structural imaging which based on magnetic resonance imaging (MRI) [1,2].

Morphometry analysis is one of the most useful tools which used for brain anatomy studies. It is used for measuring the differences of the brain anatomy structure among groups. Voxel-based morphometry (VBM) [3] is an approach which determines the differences in the brain tissue by comparing among multiple brain images. The steps of VBM includes the spatial normalization, segmentation, smoothing to reduce noise, and finally statistical tests. Statistical Parametric Map (SPM) [4] tool is used to compute the difference in contrast which is thresholded according to the Random Field theory. Computer Aided Diagnosis (CAD) aims to provide a computer output as

an opinion to assist physicians in the detection of abnormalities quantification of disease progress and differential diagnosis of lesion. Computer Aided Diagnosis (CAD) tools are widely used in classification of structural and functional brain images to distinguish which of them are normal or there is neurodegenerative disorder. There are different classifiers that could be used in the classification step. The linear or non-linear Support Vector Machine (SVM) is one of the most important classifiers that gives an excellent results with AD discovery. [5-9]

2. Accuracy and Dimension

The classification accuracy depends on the dimensionality .Dimension means the number of features. If the features are independent, there are some theoretical results that suggest the possibility excellent performance. For linear, two class multivariate normal cases, the class conditional densities are Gaussian with equal covariance matrices. A vector-valued random variable $X = [X_1 \ X_2 \ \dots \ X_n]^T$ is said to have a multivariate normal (or Gaussian) distribution with mean μ and covariance matrix $\sum$. The probability density function of normal distribution (N) is given by [10, 11, 12]

$$N(x/\mu, \textstyle\sum) = \frac{1}{(2\pi)^{\frac{d}{2}}|\sum|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(x-\mu)^T \textstyle\sum^{-1}(x-\mu)\right). \quad (1)$$

Where N is Normal densities or Gaussian distribution, $\mu = E[x]$ is the mean vector, $\sum = cov[x]$ is d×d symmetric positive

definite matrix, known as the covariance matrix, and $\sum^{-1}$ is known as the precision matrix. [10, 11, 12]
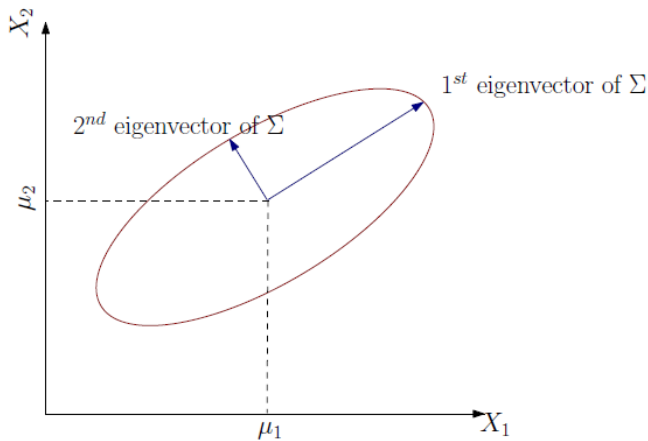

Figure 1: 2 dimensional Gaussian density [10, 11, 12].

Figure 1 shows a 2-dimensional Gaussian density. The random vectors span two dimensions and are denoted in the plot by X_1 (x-axis) and X_2 (y-axis). The means of X_1 and X_2 are μ_1 and μ_2 respectively. The density at μ is highest, and as the random vector moves away from μ, the density goes down. The first and second eigenvectors of the covariance matrix are orthogonal to each other as shown in the figure 1. The first eigenvalue is the direction of maximum variance in the multivariate normal distribution; the second eigenvector is orthogonal to the first [10, 11, 12].

The Bayes risk (i.e. the error produced by the normal classifier) is given by [10, 12]:

$$p(error) = \frac{1}{\sqrt{2\pi}} \int_{r/2}^{\infty} e^{-u^2/2} \, du \qquad (2)$$

Where $\quad r^2 = (\mu_1 - \mu_2)^T \sum^{-1}(\mu_1 - \mu_2) \qquad (3)$

Where r is the distance between μ_1 and μ_2. P (error) is related to cumulative distribution function of normal distribution.

If the features are independent, the covariance matrix is diagonal, then [10, 12]

$$r^2 = \sum_{r=1}^{d} \frac{(\mu_{1i} - \mu_{2i})^2}{\sigma_i^2} \qquad (4)$$

This shows that the independent features reduces the classification error. Thus using many enough independent features, the Bayes risk can be made arbitrarily small. For $\mu_1 = 0$ and $\mu_2 = 1$, the covariance matrix is an identity matrix. Figure 2 gives the variation of the Bayes risk as a function of the number of features in a classifier.
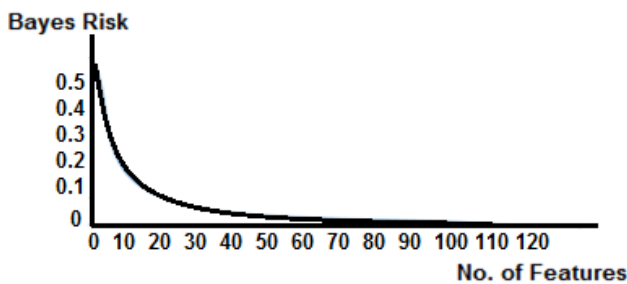

Figure 2: Bayes risk with number of features

From this figure, it is shown that the Bayes risk decreases as the number of features increases. Adding additional new features will increase the accuracy of the classifier, but if these features increases to beyond a certain point additional features lead to worse performance. Classifier design is more

computationally complex problem than the classifier evaluation. Learning the model for the class is more complex than deciding which model (or class) generated the measured features.

3. Materials and Database

One hundred and twenty subjects of men and women (aged 18-96 years) were selected from the Open Access Series of Imaging Studies (OASIS) database [13]. OASIS database has a cross-sectional collection of 120 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. More than one T1-weighted magnetic resonance imaging scan (three or four) captured in single imaging sessions. For the present study there are 49 subjects who have been diagnosed with very mild to mild AD and 71 nondemented. A summary of subject demographics and dementia status is shown in Table 1 [1, 2].

Table 1: Summary of subject demographics and dementia status [1, 2, 13].

Group	Very mild to mild AD	Normal
No of subjects	49	71
Age	63-96	33-94
CDR	0.5 – 1 – 2	0
MMSE	16-30	25-30

Where: the Clinical Dementia Rating (CDR) : range 0.5, 1, 2 for patient and 0 for normal. And the Mini-Mental State Examination (MMSE) score ranges from 0 (worst) to 30 (best) [1, 2, 13].

Several T1-weighted Magnetic Resonance Imaging (MRI) scan (three or four) captured in single imaging sessions for the same person to test if this person has the disease or not. Image parameters: TR= 9.7 msec, TE= 4.0 msec, Flip angle= 10, TI= 20 msec, TD= 200 msec, 128 sagittal 1.25 mm slices without gaps. All photos are 3-D and its dimensions are 176 X 208 X 176 voxels size, as shown in figure 3 [1, 2].
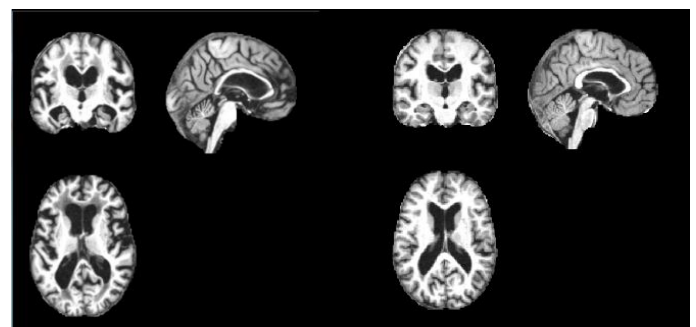
(a) Demented and mild   (b) Nondemented
with AD subject   with AD subject
Figure 3: Demented and nondemented subjects with AD

4. Proposed Algorithms

In this paper, the two proposed algorithms given in [1] and [2] will be discussed briefly to examine the accuracy and the complexity as a function of the number of extracted features.

4.1 First proposed Algorithm

The first proposed feature selection and extraction algorithm described in details in [1] and compared with the two important feature extraction algorithms (Principal Component Analysis (PCA) [1, 14] and Linear Discriminate Analysis (LDA) [1, 15]). Voxel Intensity (VI) of features extraction will be used for this comparison study. Parameter optimization was performed within a nested Cross Validation (CV) procedure [1, 16]. Classification was carried out using the Support Vector Machine (SVM) technique with linear kernel [1, 17]. Figure 4, illustrates the proposed algorithm steps.
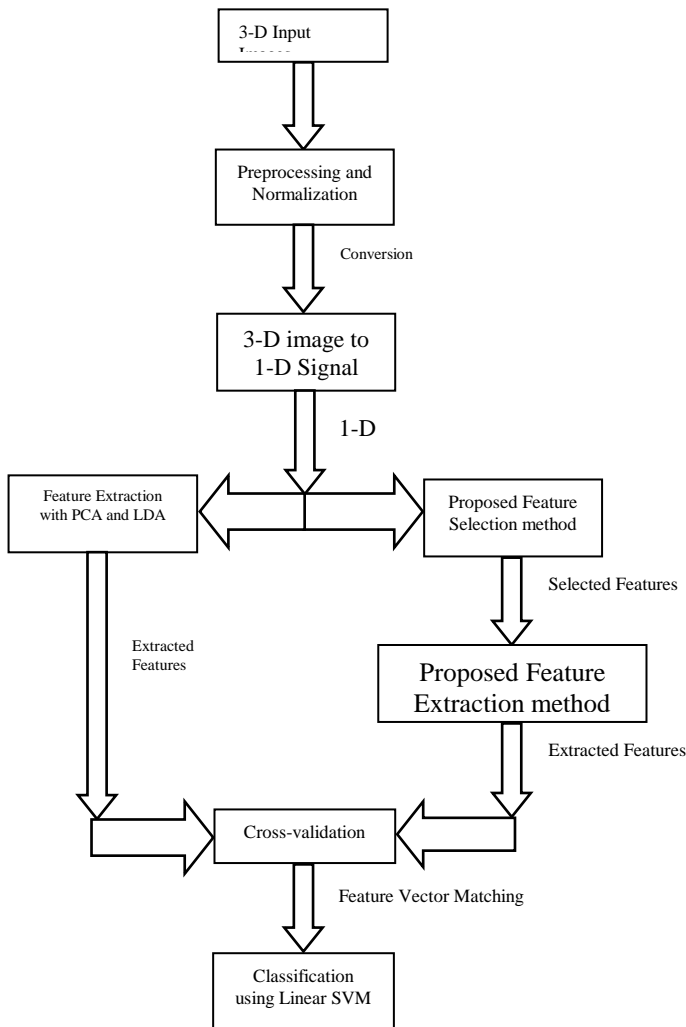

Figure 4: The First Proposed algorithm steps

Pseudo-code for the first proposed Algorithm:
1- Read the MRI images.
2- Perform the Preprocessing and Normalization for the input images.
3- Convert 3-D images to 1-D signal.
4- Extract special features using PCA, LDA,
5- Select features from brain shape features using proposed feature selection method. Then extract special features from selected feature using the proposed feature extraction method.
6- Use cross validation technique for dividing the data into 5 or 10 groups (folds). Four used for training the classifier and the last for testing.
7- Perform classification using Linear Support Vector Machine (SVM) classifier.
8- Results.

This algorithm uses the AD images from database given in [13]. These images are 3-D, so it first converted from 3-D to 1-D signal by using MATLAB reshape function. The strategy of this function is to find the linear representation and recut based on the number of rows in the reshaped array [18]. The number of features of each image equal 6443008 (176 X 208 X 176) features. Preprocessing and Normalization is applied on the images to enhance the images and reduce the dimensions of images. The number of features after Preprocessing and Normalization step reduced to 2122945 (121 X 145 X 121) features. But it still a very large dimension for each image which need very large memory size. Then applying the proposed Feature Selection method [1] reduces the number of features to 690432 features. Then extracting the special features using the proposed feature extraction algorithm to reduce number of features to 2000 features for each image. Use k-folds cross validation to randomly partitioning the 120 subjects (49 demented and 71 nondemented) into 5 folds. Four of these folds will be used for training the Linear Support Vector Machine classifier and the last fold for testing SVM then take another four folds for training and fifth for testing and so on. The detailed discussion for the first proposed algorithm given in [1].

4.2 Second Proposed Algorithm

The second proposed algorithm given in [2], reduces the number of extracted features more than the first proposed algorithm by making use of Mel-Scale Frequency Cepstral Coefficients (MFCC) technique [19, 20, 21]. The reduction of the number of extracted features is necessary and also increasing the accuracy. The flow chart of the second algorithm is shown in Fig.5.

Pseudo-code for the second proposed Algorithm:
1- Read the MRI images.
2- Convert 3-D images to 1-D signal
3- Select and Extract features using proposed feature extraction algorithm.
4- Apply MFCC to reduce number of extracted features.
5- Use cross validation for training and testing the results.
6- Perform classification using Linear Support Vector Machine (SVM) classifier.
7- Results.

3-D Input Images

↓ Conversion

Preprocessing and
Normalization

↓

3-D image to 1-D
Signal

↓ 1-D signal

The Proposed Feature
Selection method

↓ Selected Features

The Proposed Feature
Extraction method

↓ Extracted Features

MFCC Feature
Extraction

↓

Cross-Validation

↓ Feature Vector Matching
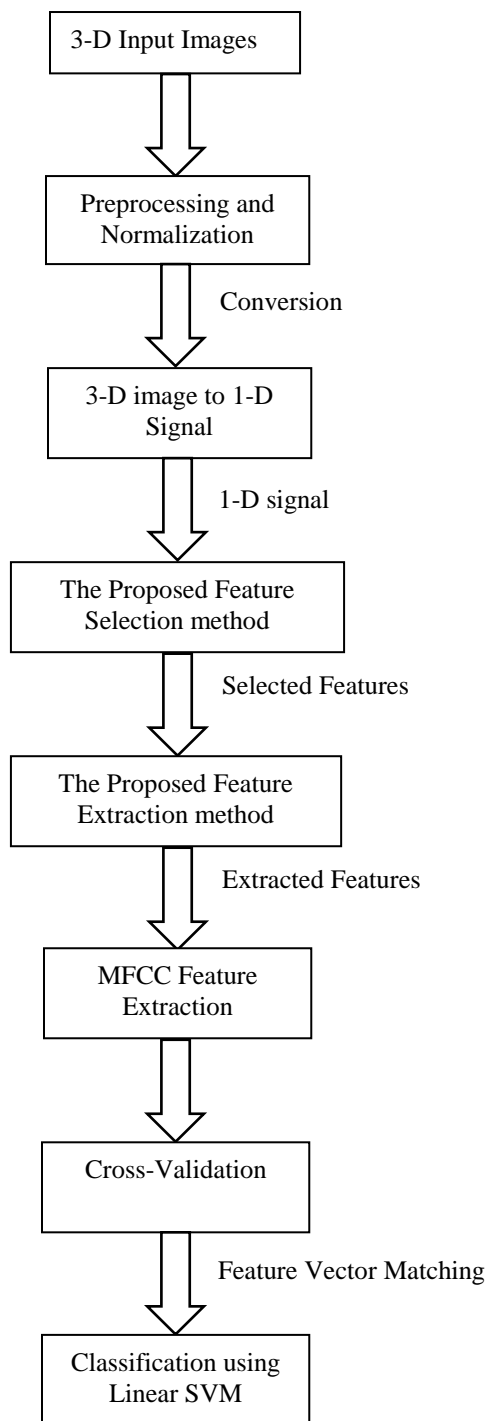
Classification using
Linear SVM

Figure 7: The Second proposed approach flow chart

For the second proposed algorithm, the 120 of AD images are converted from 3-D to 1-D signal. Then applying the proposed feature selection method. The image matrix will be (120 X 618228) where each image has 618228 features. Next, the proposed feature extraction approach given in [1] applied to image matrix. The image matrix will be (120 X 6610) as each image will have the 6610 features that obtained using the proposal given in [1]. Then, applying the steps of MFCC. It is found that each image will have only 50 features and the image matrix will be (120 X 50). After that, applying the cross validation on the 120 images will be randomly in five folds. Four of these folds will be used in training the Linear SVM which will be used as a classifier and the fifth used for testing, then take another four for training and fifth for testing and so on. The detailed analysis and discussion for the second proposed algorithm given in [2].

5. Metric Parameters

The performance of the classification system is evaluated by using the following metric parameters:
1. Accuracy.
2. Stability.
3. Complexity (speed or processing time).

The metric parameters is computed using a laptop DELL INSPIRON N5110. The main specifications of this Laptop are summarized as following: processor Intel Core i5 2.5 GHz, 8 GB RAM, 64-bit Windows 8 Enterprise Operating System, and 500 GB Hard Disk.

5.1 Accuracy

The accuracy is an important parameter used for measuring the performance of the classifier. It is usually represented by the ratio of correct classifications. The accuracy of each class is determined by the number of points that are correctly assigned to a given class. The general steps for estimating the accuracy is as follows. First, a part of the database (called the training set) to train the classifier. The trained classifier is then tested on the rest of data (the test set) and the results are compared to the actual classification that is assumed to be available. The percentage of correct decisions in the test set is an estimate of the accuracy of the trained classifier, provided that the training set is randomly sampled from the given data. There are many methods which can be used to enhance the accuracy of a classifier for artificially generated data sets or real ones, such as bagging, boosting, stacking, and their variants. [22]

To test the results the true positive, true negative, false positive and false negative, positive means that this person is a patient and negative means that the person is normal, which they are defined as:

- **True Positive (TP):** positive samples correctly classified as positive.
- **False Positive (FP):** positive samples incorrectly classified as negative.
- **True Negative (TN):** negative samples correctly classified as negative.
- **False Negative (FN):** negative samples incorrectly classified as positive.

In this paper, the Accuracy is defined as the following [1, 2]:

$$ACC = ((TN + TP))/((TP + TN + FP + FN)) \quad (5)$$

The Matthews correlation coefficient (MCC) is calculated which gives an accurate description to the Accuracy. As shown in next equation [1, 2]:

$$MCC = (((TP * TN) - (FP * FN)))/\sqrt{(((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)))}$$

$$(6)$$

5.2 Stability

The stability of the system for the updated data depends on the flexibility of the classifier and the ability of the feature to remain constant over time. A classifier is considered as being stable if bagging does not improve its performance. If small changes of the training set lead to a varying classifier performance after bagging, the classifier is considered to be an unstable one. The unstable classifiers are characterized by a high variance although they can have a low bias. On the

contrary, stable classifiers have a low variance, but they can have a high bias. [22]

5.3 Complexity (Processing Speed)

The speed of the system includes all the processing or execution time required to perform all the operation stages for the proposed algorithm. The processing time depends on the number of operation performed using the proposed approaches for all operation steps.

The execution time (T_1) for the first proposed algorithm can be calculated using the following equation:

$$T_1 = t_1 + t_2 + t_3 + t_4 + t_5 \qquad (7)$$

Where t_1 is the processing time needed to perform all operation steps for the proposed feature selection method,

t_2 is the processing time to perform all operation steps for proposed feature extraction method,

t_3 is the processing time for descending sorting the extracted features,

t_4 is the processing time for making the Cross Validation and portioning the images into 5-folds,

t_5 is the processing time for training and testing the classifier using SVM algorithm.

For the second proposed algorithm, the execution time (T_2) can be calculated using the following equation:

$$T_2 = t_1 + t_2 + t_3 + t_4 + t_5 + t_6 \qquad (8)$$

Where t_1, t_2, t_3, t_4, and t_5 are the same parameters defined previously for the first proposed algorithm. The additional term t_6 in Eq. (8) is the processing time required to extract special features using MFCC algorithm.

6. Experimental Results

Table 2 and figure 8 present the Accuracy and MCC of the classifier using the first proposed algorithm as a function of the number of extracted features. Table 3 gives the processing or execution time used as a metric parameter for measuring the complexity for the first algorithm.

Table 2: Accuracy of the classifier and MCC for the First Algorithm

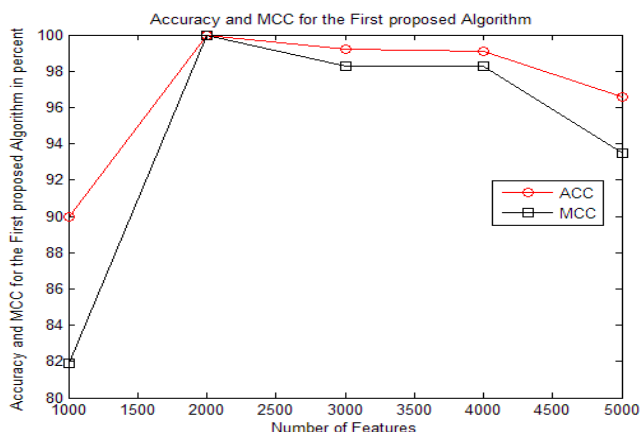| First Proposed Algorithm	No. of Features				
	5000	4000	3000	2000	1000
Accuracy	96.6	99.1	99.2	100	90
MCC	93.5	98.3	98.3	100	81.9

Figure 8: Accuracy and MCC of the classifier for the First Algorithm

Table 3: Execution time for the First proposed Algorithm

Proposed Algorithm	Execution Time in seconds					
	t_1 (Feature Selection)	t_2 (Feature Extraction)	t_3 (Descending Sorting)	t_4 (Cross Validation)	t_5 (Training & Testing SVM)	Total execution Time (T)
First proposed Algorithm	2.1885	0.6401	0.0508	0.0035	0.2077	3.0906

The variation of the accuracy and MCC of the classifier versus the number of extracted feature using the second proposed algorithm is given in Table 4 and depicted in figure 9. Table 5 presents the metric parameter used for measuring the complexity which is the processing (execution) time calculated for the second proposed algorithm.

Table 4: Accuracy of the classifier and MCC for the second Algorithm

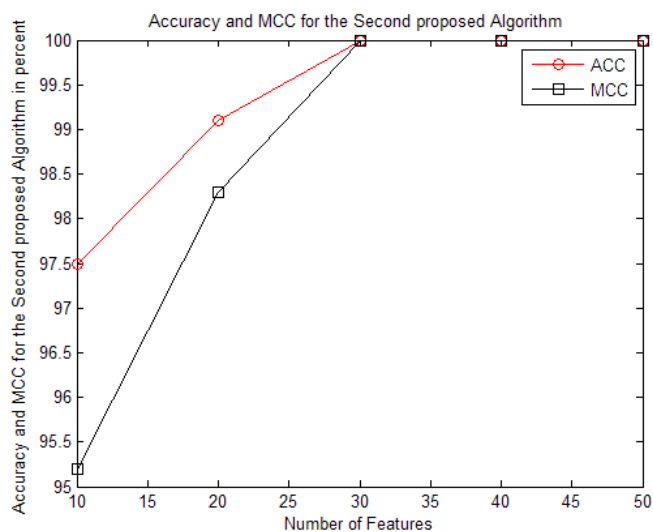| Second Proposed Algorithm	No. of Features				
	50	40	30	20	10
Accuracy	100	100	100	99.1	97.5
MCC	100	100	100	98.3	95.2

Figure 9: Accuracy of the classifier and MCC for the Second Algorithm

Table 5: Execution time for the Second proposed Algorithm

Proposed Algorithm	Execution Time in seconds						
	t_1 (Feature Selection)	t_2 (Feature Extraction)	t_3 (Descending Sorting)	t_4 (Cross Validation)	t_5 (Training & Testing SVM)	t_6 (MFCC feature extraction)	Total execution Time (T)
Second proposed Algorithm	2.0875	0.6415	0.0495	0.0036	0.0918	0.6806	3.5545

7. Result Discussion

From the obtained results presented in Table 2 and illustrated in figure 8, it is noticed that the accuracy and the

MCC of the classifier values reached to 100% using the first proposed algorithm for feature extraction of Alzheimer's disease with number of features equal to 2000 features. If number of features increase more than 2000 features, the system performance decreases than as at 2000 features. Table 3 presents the execution (processing) time required to perform the all operation steps for the first proposed algorithm equal to 3.09 seconds.

Table 4 and figure 9 give the accuracy and MCC of the classifier using the second proposed algorithm which equal to 100% with the number of features equal to 30 features and the classifier becomes stable which the accuracy becomes constant over time as the number of features increases to 30, 40, and 50 features as indicated in figure 9.

Table 5 shows the processing time required to realize the second proposed algorithm steps equal to 3.55 seconds. The difference between both proposed algorithms processing time is small and negligible.

From the obtained results it is clear that using the MFCC technique in the second proposed algorithm, the accuracy and MCC coefficients of the classifier equal 100% with small number of extracted features (30 features). Thus, comparing the two proposed algorithms it is clear that the number of extracted features needed to realize the same value of accuracy and MCC are reduced by about 70% (from 2000 to 30 features (2000/30 = 70%)) which is very significant for the memory size reduction with a negligable increase in the processing or execution time.

As the number of features reduced, the error will be reduced and consequently increases the accuracy of the classifier. This will make the classifier design be simpler.

8. Conclusion

This paper discussed the complexity and the dimensionality problem for computer aided diagnosis of Alzheimer's disease. Two proposed feature extraction algorithms are presented. The obtained accuracy equal to 100% for both proposed algorithms. The first proposed algorithm gives the accuracy equal to 100% using number of features equal to 2000 features. For the second proposed algorithms the accuracy of the classifier are kept at the same level (100%) but with small number of features equal to 30 features. In addition, the system stability using the second proposed algorithm is better. Thus, the memory size reduced by about 70% which significantly reduces the hardware implementation for designing the classifier. This leads to reduction the economic cost.

9. References

[1] M.M.Dessouky, M.A.Elrashidy, T.E.Taha, and H.M Abdelkader, "Selecting and Extracting Effective Features for Automated Diagnosis of Alzheimer's Disease", International Journal of Computer Applications, Vol. 81 – No.4, 2013, pp 17-28.

[2] M.M.Dessouky, M.A.Elrashidy, T.E.Taha, and H.M Abdelkader, "Effective Features Extracting Approach Using MFCC for Automated Diagnosis of Alzheimer's Disease", CiiT International Journals for Data Mining and Knowledge Engineering Journal, Vol 6, No 2, 2014, pp. 49-59.

[3] VBM8: http://dbm.neuro.uni-jena.de/vbm/

[4] SPM8: http://www.fil.ion.ucl.ac.uk/spm/

[5] Glenn Fung, Jonathan Stoeckel," SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease using Spatial Information", Knowledge and Information Systems, Volume 11, Issue 2, pp 243-258, 2007

[6] Dong Hye Ye, Kilian M. Pohl and Christos Davatzikos. "Semi-Supervised Pattern Classification: Application to Structural MRI of Alzheimer's Disease", IEEE International Workshop on Pattern Recognition in NeuroImaging, 2011.

[7] Simon F. Eskildsen, Pierrick Coupé, Daniel García-Lorenzo, Vladimir Fonov, Jens C. Pruessner, D. Louis Collins and the Alzheimer's Disease Neuroimaging Initiative, "Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning", NeuroImage, Volume 65, 15 January 2013, pp. 511–521.

[8] Katherine R. Gray, "Machine learning for image-based classification of Alzheimer's disease", thesis submitted for the degree of Doctor of Philosophy, Department of Computing, Imperial College London, 2012.

[9] Alexandre Savio, Maite García-Sebastián , Manuel Graña, Jorge Villanúa, "Results of an Adaboost approach on Alzheimer's Disease detection on MRI", Bioinspired Applications in Artificial and Natural Computation ,Lecture Notes in Computer Science Volume 5602, 2009, pp 114-123.

[10] Sergios Theodoridis, " Pattern Recognition", Elsevier, 2006, 3Ed.

[11] Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer Science + Business Media, LLC, 2006.

[12] Richard O. Duda, Peter E. Hart, David G.Stork, "Pattern Classification", Wiley Interscience, 2000, 2Ed.

[13] OASIS database: http://www.oasis-brains.org

[14] Lindsay I. Smith," A tutorial on Principal Components Analysis", University of Otago, New Zealand 2002.

[15] Shih, Frank Y,"Image processing and pattern recognition: fundamentals and techniques.", IEEE,2010.

[16] Chia-Yueh C. CHU, "Pattern recognition and machine learning for magnetic resonance images with kernel methods", Doctor of Philosophy thesis, University College London, 2009.

[17] Katherine R. Gray, "Machine learning for image-based classification of Alzheimer's disease", thesis submitted for the degree of Doctor of Philosophy, Department of Computing, Imperial College London, 2012.

[18] http://www.mathworks.com/help/matlab/ref/reshape.html

[19] N. Tawfik, M. Eldin, M. Dessouky, and F. AbdEl-samie, "processing of Corneal Images With a Cepstral Approach", ICCTA,2013.

[20] B. Martin, and V. Juliet, "Extraction of Feature from the Acoustic Activity of RPW using MFCC", Recent Advances in Space Technology Services and Climate Change (RSTSCC), 2010, pp. 194-197.

[21] M.R.Devi, and T.Ravichandran, "A novel approach for speech feature extraction by Cubic-Log compression in MFCC", International Conference on Pattern Recognition, Informatics and Mobile Engineering, 2013.

[22] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting and variants," Machine Learning, vol. 36, pp. 105–142, 1999.

Author Profile

Mohamed M. Dessouky was born in Egypt, 27 April 1984. Graduated from department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufiya University, Egypt at 2006. Demonstrator at 2007, Assistant Lecturer at 2011. Now he is a PhD student. The major field of study is image processing and artificial intelligence.

Mohamed has More than six years of teaching experience as an assistant lecturer, and Teaching Assistant for a variety of undergraduate courses in different Computer science and Engineering fields.

Dr. Mohamed is CISCO Certified Instructor and got award from CISCO as a best instructor for more than 5 years.