30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Genetic Algorithm-Based PCA Eigenvector Selection and Weighting for automated identification of Dementia using FDG-PET Imaging

Yong Xia, *Member IEEE*, Lingfeng Wen, *Member IEEE*, Stefan Eberl, *Member IEEE*, Michael Fulham, and Dagan Feng, *Fellow, IEEE*

Abstract— **Parametric FDG-PET data offer the potential for an automated identification of the different dementia syndromes. Principal component analysis (PCA) can be used for feature extraction in FDG-PET. However, standard PCA is not always successful in delineating the features that have the best discrimination ability. We report a genetic algorithm-based method to identify an optimal combination of eigenvectors so that the resultant features are capable of successfully separating patients with suspected Alzheimer's disease and frontotemporal dementia from normal controls. We compared our approach with standard PCA on a set of 210 clinical cases and improved the performance in separating the dementia types with an accuracy of 90.0% and a Kappa statistic of 0.849. There was very good agreement between the automated technique and the diagnosis given by clinicians.**

I. INTRODUCTION

With the increasing ageing of the population, dementia has become an important world-wide public health problem. Positron emission tomography (PET), a functional imaging modality, can detect subtle changes in cerebral metabolism before there are changes on anatomical imaging and before a clinical diagnosis of probable dementia is made [1]. However, the interpretation of PET images remains a challenge because the changes can be subtle early in the course of the disease and there can be some overlap with other dementia types [2].

Alzheimer's disease (AD) and frontotemporal dementia (FTD) are two of the more common types of dementia [1, 3] and investigators have developed several methods to automatically separate these patients from healthy subjects. The methods in the literature have two main steps:- feature extraction and classification. The features that are considered include voxel values, gradients, simple statistics, histogram, fractal dimension [4, 5] and the T-map [6]. The features can be calculated from the whole brain, regions of interest (ROIs) and sinogram data [7]. The classification methods range from K-mean clustering [8], to artificial neural network (ANN) [9, 10] and a support vector machine (SVM) [11, 12].

PET scanning provides the additional advantage of deriving quantitative parametric images. In our previous work [13] we showed that voxel values of ^{18}FDG-PET parametric images are more robust than histograms, local statistics and gradients calculated from ROIs in identifying different conditions. Due to the huge number of voxels, dimensionality reduction is an essential prerequisite for classification when using voxel values as features. Principal component analysis (PCA) is the most common dimensionality reduction technique that is used [14]. PCA performs a linear transformation of data to a lower dimensional space spanned by a few eigenvectors of the covariance matrix of the data. These few eigenvectors are determined by sorting all eigenvectors in order of decreasing eigenvalue and choosing the first L eigenvectors which give an accumulative energy ν above a defined threshold, such as 90%. Although PCA maximizes the variance of the transformed data, it is not optimized for class separability [15], since larger eigenvalues do not necessarily guarantee the corresponding eigenvectors contain the "most important" aspects of the data in terms of classification.

Our aims were to remedy the disadvantage of PCA and provide a holistic solution for selecting an optimal combination of eigenvectors so that the resultant features have the best discrimination ability. Our eigenvector selection and weighting method is based on the genetic algorithm and is directed at maximizing the agreement between the result of automated dementia classification and the diagnosis given by experienced neurologists. Finally, we compared the proposed approach to standard PCA on a set of clinical PET images.

II. METHODS

A. Data Preparation

We selected 210 neurological studies from the PET archive, including 80 cases of AD, 60 cases of FTD, 35 cases of non-dementia, and 35 cases of healthy volunteers. Both the non-dementia and healthy cases were used as normal controls.

All studies were performed on an ECAT 951/R whole body PET scanner (Siemens/CTI, Knoxville, TN., U.S.A.). Approximately 400 MBq of ^{18}F-FDG was infused at a

Manuscript received April 7, 2008. This work was supported in part by ARC, PolyU/UGC grants.

Dr Yong Xia, Dr Lingfeng Wen, A/Prof Stefan Eberl, Prof Michael Fulham, and Prof Dagan Feng are with the Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia. (e-mail: y.xia@usyd.edu.au ; wenlf@ieee.org; feng@it.usyd.edu.au; stefan@cs.usyd.edu.au; mfulham@med.usyd.edu.au).

Dr Yong Xia, Dr Lingfeng Wen, and Prof Dagan Feng are also with the Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong.

Dr Lingfeng Wen, A/Prof Stefan Eberl and Prof Michael Fulham are also with the Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia.

Prof Michael Fulham is also with the Faculty of Medicine, University of Sydney, Australia.

constant rate over a 3-min period to study in-vivo metabolism of glucose. Two arterialised-venous blood samples were taken at 10 min and 45 min post injection to calibrate the population-based input function [16]. PET scanning commenced at least 30 min after tracer injection with a scan duration of twenty minutes.

The autoradiographic method [17] was used to calculate parametric images of cerebral metabolic rate of glucose consumption (CMRGlc). The constructed CMRGlc images were spatially normalized to a PET brain template provided by the SPM2 package (Wellcome Trust Centre for Nueorimaging, London, U.K.) [18] to reduce anatomical and intensity variations and to ensure a meaningful voxel-wise comparison between the images.

B. PCA Eigenvector Computation

After spatial normalization, the CMRGlc image volume has a dimension of 91×109×91 and a voxel size of 2×2×2 mm^3. To reduce computational and spatial complexity, each dimension was down-sampled by a factor of 3, preceded by smoothing with a Gaussian filter (FWHM =8mm). Background and the cerebrospinal fluid (CSF) region were removed based on the regions defined in the MNI brain phantom [19]. As a result, a total of 8065 voxels are sampled from each CMRGlc image volume. The voxel value feature set of all images is denoted by a matrix $A_{D\times N}$ ($D = 8065$ and $N = 210$). The mean for each voxel across the studies was subtracted to yield the mean-subtracted matrix B

$$B = A - u \cdot h \tag{1}$$

where u is the mean of each row of A, and h is a $1 \times N$ row vector of all 1's. Finally, the singular value decomposition (SVD) [14] is applied to the covariance matrix of B to calculate its eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_D\}$ and the corresponding eigenvectors $\{X_1, X_2, \cdots, X_D\}$.

C. Eigenvector Selection and Weighting

From the calculated eigenvectors, we select a few to construct the transformation matrix

$$T_{D^*\times N} = \{X_{i_1}, X_{i_2}, \cdots, X_{i_{D^*}}\}. \tag{2}$$

The voxel value feature set is then projected into the D^*-dimensional space in the following way

$$F = T \cdot \left(\frac{B}{\sigma \cdot h} \right), \tag{3}$$

where the column vector σ is the main diagonal of the co-variance matrix of B . The optimum of eigenvector selection is assessed using the performance of classification obtained from the projected features. The classification performance is characterized by the accuracy and the Kappa statistic [21], which are calculated with the neurologists' diagnosis as the correct diagnosis. To efficiently utilize all the available studies, a 10 fold cross validation was performed. Each time, 90% of the studies in each category are used as training set, while the remaining 10% are used for assessment. This is repeated 10 times to ensure all studies contribute to the evaluation once. Since the training set is relatively small, the support vector machine (Lib SVM [20]) is adopted as the classifier in this research.

Selecting a set of eigenvectors to ensure the best classification performance is an optimization problem that can be solved by using the genetic algorithm (GA). Since most of the 8065 eigenvectors resulting from SVD are in fact trivial to classification, we initially sort them in order of decreasing eigenvalue and choose the first 100 as potential candidates. We then construct a binary coding GA (bGA) to further select eigenvectors from these candidates. In the bGA, each solution is represented as a 100-bit binary number, where 1 means the corresponding eigenvector is selected and 0 means it is discarded. A solution's fitness is defined as the Kappa statistic of the classification resulting from it. The new generation of population is created by several genetic operators, including best solution inheritance, roulette wheel selection, one-point crossover, random mutation, and gene modification. Since the classifier prefers lower dimensionality of the feature space, gene modification is designed to produce new solutions by modifying the current best one, discarding 1 to 3 selected eigenvectors which make the least contribution to the fitness. To avoid the optimization process being trapped in a local maximum, the proposed bGA adopts a variable mutation probability, given as follows

$$p_m^{(n+1)} = \begin{cases} p_m^{(n)}, & f_m^{(n)} > f_m^{(n-1)} \\ \alpha_{pm} \cdot p_m^{(n)}, & otherwise \end{cases} \tag{4}$$

where α_{pm} is the increasing rate of mutation probability , $p_m^{(n)}$ and $f_m^{(n)}$ are the mutation probability and highest fitness of the n th generation. When the mutation probability reaches its threshold T_{pm}, it will be reset to its initial value $p_m^{(0)}$ to prevent the bGA from degenerating to random searching. Another operator will be triggered when the evolution has been halted for more than 4 generations. In this case, all individuals whose fitness equals to the highest fitness will be replaced by their offspring produced by mutating. This operator aims to diversify genes in the population and thus speed up the evolution. Finally, the evolution will terminate when a predetermined number of generations is reached.

Each eigenvector characterizes a specific pattern in the brain and hence can be considered as an "eigenbrain" as shown in Fig. 3. Since various regions have different importance to dementia classification, the significance of eigenbrains should be different ie higher weights should be given to eigenvectors or eigenbrains which highlight regions which provide strong differentiation between the study types. Therefore, we assign each selected eigenbrain a weight and hence modify the transformation matrix given in Eq. (2) as follows

$$T' = \{\omega_{i_1} X_{i_1}, \omega_{i_2} X_{i_2}, \cdots, \omega_{i_{D^*}} X_{i_{D^*}}\}. \tag{5}$$

We adopt a real-valued GA (rGA) [21] to optimize the weight vector $\{\omega_{i_1}, \omega_{i_2}, \cdots, \omega_{i_{D^*}}\}$. In the rGA, a solution consists of 100 real numbers, each of which takes a value from $\left[-\beta\sqrt{2}, \sqrt{2}\right]$. A positive number represents a weight

while a negative number means the corresponding eigenbrain is discarded. The bias coefficient β interprets the prior that solutions with good performance usually select a small number of eigenbrains. The fitness function and the termination condition are the same as those of bGA.

III. Results

The bGA-based eigenbrain selection is first carried out and the resultant optimal solution is included in the initial population of rGA. In our experiments, the initial mutation probability $p_m^{(0)}$, mutation probability threshold T_{pm}, mutation probability increment, α_{pm}, and bias coefficient β are empirically set to be 0.01, 0.05, 1.09, and 5, respectively. Although inappropriate parameters may slow down the convergence, the GA algorithm is generally not sensitive to these parameters. The indexes and weights of the obtained optimal combination of eigenbrains are listed in Table 1. In Fig. 1, the classification performance of the features derived from these eigenbrains is compared with that of standard PCA, which selects the eigenvectors (or eigenbrains) with the highest eigenvalues. For PCA, the classification performance initially improves with increasing number of eigenbrains. However, as the number of selected eigenbrains exceeds 12, although the accumulated energy continues to increase, the performance tends to worsen. This clearly illustrates that the accumulated energy is not the optimal criterion for eigenbrain selection. Fig. 1 reveals that the proposed eigenbrain selection and weighting method outperforms PCA in dementia classification. By weighting the contribution of each selected eigenbrain, we achieved a classification accuracy of 90.0% and a Kappa statistic of 0.849, a very good agreement between the automated classification result and the diagnosis given by neurologists.

Since GA is likely to be trapped in local maximum, we run the proposed feature selection and weighting algorithm 15 times. The mean weight of each eigenbrain and the average classification performance are depicted in Fig. 2. It shows that there are 17 significant eigenbrains with weights above 0.8.

When displaying the absolute voxel values of some eigenbrains in Fig. 3, the "highlighted" area represents the region to which the feature resulted from this eigenbrain mainly responds. For example, the features derived from the 6th eigenbrain characterize the information of right temporal lobe, and the features produced by the 3rd eigenbrain mainly reflect the situation in frontal lobe and occipital lobes. The frontal, temporal, parietal, occipital lobes and the

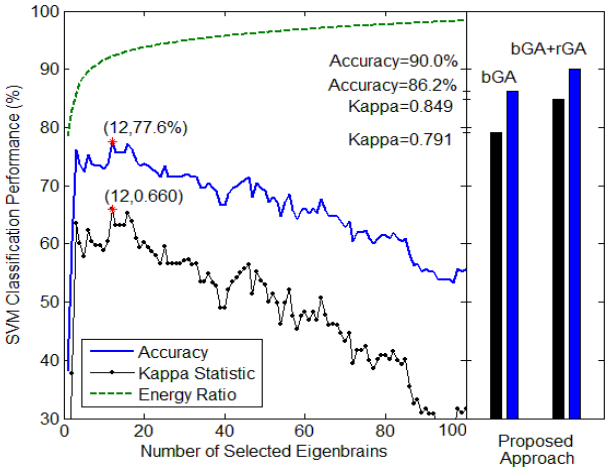

Fig. 1 Performance comparison between PCA and proposed feature selection method.

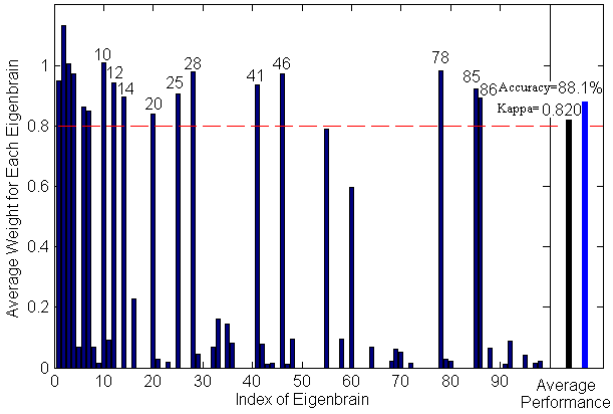

Fig. 2 Average weight of each Eigenbrain and the average classification performance

TABLE I
INDEX AND WEIGHTS OF SELECTED EIGENBRAINS

Method	Index or Weights
bGA	1 2 3 4 6 7 10 12 14 20 25 28 41 46 55 60 78 85 86
bGA + rGA	W_1=0.8867, w_2=1.2420, w_3=1.0051, w_4=0.7875, w_6=0.5441, w_7=0.4773, W_{10}=0.9996, w_{12}=1.0100, w_{14}=1.0040, w_{20}=0.7378, w_{25}=0.7995, w_{28}=1.0159, W_{41}=0.9958, w_{44}=0.2296, w_{46}=0.8186, w_{47}=0.1513, w_{55}=1.0049, w_{60}=0.5466, W_{78}=1.0002, w_{85}=1.0057,w_{86}=0.5825

hippocampal formation are important structures in AD and FTD [23, 24]. We calculate the average absolute voxel values of these tissues in each eigenbrain and interestingly find that the largest averages are mostly found in the eigenbrains selected by our approach. Fig. 4 compares these largest averages with the average absolute voxel values of the corresponding eigenbrain. It reveals that the majority of these tissues are "highlighted" by the selected eigenbrains. This result demonstrates that the output of the proposed eigenbrain selection approach is largely in accord with the pathological knowledge.

IV. Conclusion

We investigated the automated differentation of AD and FTD from normal subjects from the perspective of selecting an optimal combination of eigenbrains to derive the features that had the best differentiation ability. Compared with the standard PCA algorithm, the proposed eigenbrain selection and weighting method improves the performance of dementia classification. However, the spatial resolution of current

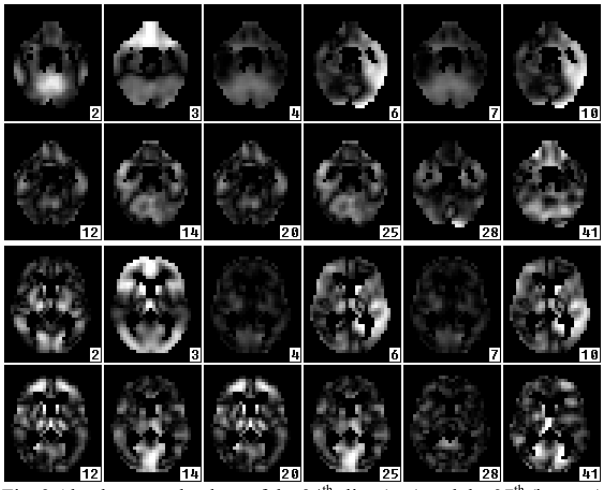

Fig. 3 Absolute voxel values of the 24th slice (top) and the 37th (bottom) of 12 selected eigenbrains. The number at bottom right is the index of each eigenbrain.

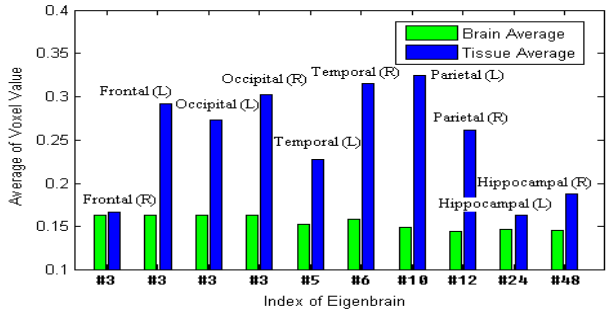

Fig. 4 Comparison of the average absolute voxel values.

eigenbrains is relatively low to limit the complexity. Our future work is aimed at improving the resolution of eigenbrains and mining the information embedded in the eigenbrains.

ACKNOWLEDGMENT

This work was supported by ARC and PolyU grants.

REFERENCES

[1] D. H. S. Silverman, "Brain ^{18}F-FDG PET in the diagnosis of neurodegenerative dementias: Comparison with perfusion SPECT and with clinical evaluations lacking nuclear imaging," *Journal of Nuclear Medicine*, vol. 45, pp. 594-607, 2004.

[2] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "Alzheimer's disease and models of computation: Imaging, classification, and neural models," Journal of Alzheimers Disease, vol. 7, pp. 187-199, 2005.

[3] M. Sjögren and C. Andersen, "Frontotemporal dementia - A brief review," *Mechanisms of Ageing and Development*, vol. 127, pp. 180-187, 2006.

[4] M. Nagao, Y. Sugawara, M. Ikeda, et al., "Heterogeneity of cerebral blood flow in frontotemporal lobar degeneration and Alzheimer's disease," European Journal of Nuclear Medicine and Molecular Imaging, vol. 31, pp. 162-168, 2004.

[5] T. Yoshikawa, K. Murase, N. Oku, et al., "Statistical image analysis of cerebral blood flow in vascular dementia with small-vessel disease," Journal of Nuclear Medicine, vol. 44, pp. 505-511, 2003.

[6] K. Herholz, E. Salmon, D. Perani, et al., "Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET," Neuroimage, vol. 17, pp. 302-316, 2002.

[7] A. Sayeed, M. Petrou, N. Spyrou, et al., "Diagnostic features of Alzheimer's disease extracted from PET sinograms," Physics in Medicine and Biology, vol. 47, pp. 137-148, 2002.

[8] M. Pagani, V. A. Kovalev, R. Lundqvist, et al., "A new approach for improving diagnostic accuracy in Alzheimer's disease and frontal lobe dementia utilising the intrinsic properties of the SPET dataset," European Journal of Nuclear Medicine and Molecular Imaging, vol. 30, pp. 1481-1488, 2003.

[9] D. Hamilton, D. Omahony, J. Coffey, et al., "Classification of mild Alzheimer's disease by artificial neural network analysis of SPET data," Nuclear Medicine Communications, vol. 18, pp. 805-810, 1997.

[10] M. Nagao, Y. Sugawara, M. Ikeda, et al., "Heterogeneity of cerebral blood flow in frontotemporal lobar degeneration and Alzheimer's disease," European Journal of Nuclear Medicine and Molecular Imaging, vol. 31, pp. 162-168, 2004.

[11] R. Higdon, N. L. Foster, R. A. Koeppe, et al., "A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging," Statistics in Medicine, vol. 23, pp. 315-326, 2004.

[12] G. Fung and J. Stoeckel, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," Knowledge and Information Systems, vol. 11, pp. 243-258, 2007.

[13] L. Wen, M. Bewley, S. Eberl, M. Fulham, and D. Feng, "Classification of dementia from fdg-pet parametric images using data mining," to be published in Proc. of ISBI'08.

[14] S. Haykin, "Principal Components Analysis," in Neural Networks: A Comprehensive Foundation, 2 ed. New Jersey: Prentice Hall, 1999.

[15] K. Fukunaga, Introduction to Statistical Pattern Recognition (2nd Ed.), Academic Press, New York, 1990.

[16] S. Eberl, A.R.Anayat, R.R.Fulton, et al., "Evaluation of two population-based input functions for quantitative neurological FDG PET studies," European Journal of Nuclear Medicine, vol. 24, pp. 299-304, 1997.

[17] G. D. Hutchins, J. E. Holden, R. A. Koeppe, et al., "Alternative Approach to Single-scan Estimation of Cerebral Glucose Metabolic Rate Using Glucose Analogs with Particular Application to Ischemia," Journal of Cerebral Blood Flow and Metabolism, vol. 4, pp. 35-40, 1984.

[18] R. S. J. Frackowiak, K. J. Friston, C. D. Frith, et al., Human Brain Function. Amsterdam; Boston: Elsevier Academic Press, 2004.

[19] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," Neuroimage, vol. 15, pp. 273-289, 2002.

[20] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training SVM," Journal of Machine Learning Research, vol. 6, pp. 1889-1918, 2005.

[21] J. L. Fleiss, "Measuring nominal scale agreement among many raters." Psychological Bulletin, vol. 76, no. 5 pp. 378—382, 1971.

[22] C.R. Houck, J. Joines, and M. Kay, "A genetic algorithm for function optimization: A Matlab implementation," NCSU-IE TR 95-09, 1995.

[23] M.D. Devous Sr, "Functional brain imaging in the dementias: role in early detection, differential diagnosis and longitudinal studies" Eur. J. Nucl. Med., vol. 29, pp. 1685-1696, 2002.

[24] J.R. Petrella, R.E. Coleman and P.M. Doraiswamy, "Neuroimaging and early diagnosis of Alzheimer's disease: a look to the future," Radiology, vol. 226, no. 2, pp. 315-336, 2003.