CLASSIFICATION OF DEMENTIA FROM FDG-PET PARAMETRIC IMAGES USING DATA MINING

Lingfeng Wen[1,2], *Member IEEE*, Michael Bewley[3], Stefan Eberl[1,2], *Member IEEE*, Michael Fulham[1,2], (David) Dagan Feng[1,4], *Fellow IEEE*

[1] Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia
[2] Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia
[3] School of Electrical and Information Engineering, University of Sydney, Australia
[4] Center for Multimedia Signal Processing (CMSP), Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong

ABSTRACT

It remains a challenge to identify the different types of dementia and separate these from various subtypes from the normal effects of ageing. In this paper the potential of parametric images from FDG-PET studies to aid the classification of dementia using data mining techniques was investigated. Scalar, joint, histogram and voxel-level features were used in the investigation with principal component analysis (PCA) for dimensionality reduction. The logistic regression model and the additive logistic regression model were applied in the classification. The results show that cerebral metabolic rate of glucose consumption *(*CMRGlc) was efficient in the classification of dementia and data mining using voxel-level features with PCA and the logistic regression model method achieving the best classification.

Index Terms— Dementia, data mining, classification, parametric image

1. INTRODUCTION

As the present population ages, dementia is an important global health and social problem. Functional imaging such as positron emission tomography (PET) plays an important role in early detection and differentiation of dementia [1]. Image interpretation remains problematic, however, because there is overlap between the changes found in normal ageing and those found in early dementia. Data mining has emerged as a promising technique in characterizing and classifying the dementias [2].

Artificial neural networks (ANN) have been used to classify the dementias where region-of-interest (ROI) uptake values are used to train the network [3]. The ROI usually requires prior knowledge of the pattern of disease involvement and manual delineation. The thresholding methods rely on simple values derived from data mining to differentiate disorders [4-6]. For generic reconstruction data, the value of fractal dimension was derived by fractal analysis according to the heterogeneity of functional changes [5, 6]. Another approach used sinogram data to derive a rotational sensitive feature [4]. Clustering, where a multidimensional co-occurrence matrix was classified by a K-mean clustering method has also been described [7]. Recently, the support vector machine (SVM) method has been applied in data mining of functional images [8]; spatial correlations of adjacent voxels are then taken into account to form a modified SVM [9].

Despite extensive literature in the field, to our knowledge, we could not identify work that takes advantage of quantitative parametric images. Quantitative parametric images are derived from functional imaging and can explicitly describe physiological processes. Thus, for this study the aim was to apply data mining techniques to parametric image data from FDG-PET to classify the dementia of Alzheimer's disease (AD), fronto-temporal dementia (FTD) and normals.

2. METHODS

2.1. Data acquisition

There were 196 neurological studies that were selected from the archives of Department of PET and Nuclear Medicine at Royal Prince Alfred Hospital. The majority (n=160) were patients who were referred with a diagnosis of a neurodegenerative disorder – dementia of the Alzheimer or fronto-temporal type; there were also 36 normal volunteers.

All studies were carried out on an ECAT 951R whole body PET scanner (Siemens/CTI, Knoxville, Tenn.). Approximately 400 MBq of FDG was infused at a constant rate over a 3-min period. Three arterialised-venous blood samples were taken at 0 min (before isotope injection) then at 10 min and 45 min post injection to calibrate a population-based input function [10]. PET scanning commenced at least 30 min after tracer injection with a scan

duration of twenty minutes. A post-injection transmission method was used for photon attenuation [11].

2.2. Spatial normalization of parametric images

Cerebral metabolic rate of glucose consumption (CMRGlc) is a functional parameter that describes *in-vivo* metabolism of glucose. Parameter estimation of CMRGlc requires a kinetic modeling technique with a measured blood curve. The autoradiographic method was used to derive voxel-wise parametric images of CMRGlc, assuming a three-compartment and four-parameter kinetic model for FDG [12].

To allow standardized brain structure definition to be used, the parametric images (128×128×56 voxels with a voxel size of 1.840×1.840×3.375 mm^3) were spatially normalized to a PET brain template using the SPM2 package (Wellcome Trust Centre for Neuorimaging, London, U.K.) [13], resulting in spatially normalized images of 91 by 109 by 91 voxels with a voxel size of 2×2×2 mm^3. No post filtering was applied to the spatially normalized images.

2.3 Investigated features

Four types of features were investigated - scalar, joint, histogram, and voxel-level features. In order to apply the SVM method in data mining, the large number of input features contained in the images must be reduced to a relatively small number. Thus we applied two methods of dimensionality reduction - a ROI-based analysis, and principal component analysis (PCA). Automated labeling of sub-cortices of MNI brain space [14] was used to define 116 ROIs to cover the whole brain template used in the spatial normalization.

2.3.1. Scalar and joint features

The scalar features were then derived for each ROI. Two scalar features were obtained - the mean of CMRGlc ($Mean_{Glu}$), and the mean of gradients in three dimensions ($Mean_{Gra}$).

Intensity normalization is usually performed for non-quantitative uptake images to reduce effects of different scan time and dose of injected tracer [3, 9]. The cerebellum was used as the reference region to achieve normalized $Mean_{Glu}$ ($NMean_{Glu}$) to compare the effects of intensity normalization.

A joint feature was derived directly from the ROI-based $Mean_{Glu}$ and $Mean_{Gra}$, forming a two-element vector.

2.3.2. Histogram features

Voxels can also be potentially classified as normal or abnormal based on their quantitative values. Histogram analysis was applied with 2, 3, 4 and 12 bins. Since absolute values of CMRGlc were calculated in mg/min/100g, histogramming could be restricted to a particular range of CMRGlc , eg <10 mg/min/100g represents hypo metabolism for grey matter. The allowable maximum value (upper bound) of the analysis ranged from 2 to 15 mg/min/100g. The binned voxel numbers formed a vector as the histogram feature for the number of bins and the upper bounds. The histogram features were calculated for each ROI.

2.3.3. Voxel-level features

The aforementioned features require definitions of ROIs. To avoid requirement for prior knowledge of sub-cortices in the anatomy, voxel-level features were used instead.

The huge number of brain voxels (about 860,000) increased the computational complexity in data mining. Therefore, a Hanning low-pass filter was applied to down-sample the images in each dimension by a factor of 1.5.

2.4 Dimensionality reduction

Principal component analysis (PCA) is widely used as a tool for dimensionality reduction. PCA transforms original inputs to a new space by an orthogonal linear transformation to maximize overall variances of the estimated components. This results in the principal components with greater variances being retained while the "least important" components are discarded [15]. The singular value decomposition (SVD) technique was applied to find principal components in this investigation. The number of components retained in PCA ranged from 2 to 50.

2.5. Data mining

The regression model method is a well-known data mining technique to fit a model to a dataset with numerical inputs. Two methods were used for the logistic regression model (LR), and the additive logistic regression model (ALR).

The LR models the posterior class probabilities $\Pr(G=j|X=x)$ for the classes using the logit transform through formation of a multinomial logistic regression model with a ridge estimator (1).

$$j^* = \arg\max_j \Pr(G = j \mid X = x) \qquad (1)$$

where G is a class variable, X is the input vector [16].

The ALR method is a generalization of the LR method with forward stage-wise additive modeling method for boosting the regression. The Weka package (University of Waikato, Hamilton, New Zealand) was used to implement the two methods [17].

2.6 Evaluation

The studies were interpreted by an experienced neurologist and PET physician with 87 patients with AD and 73 with FTD. There are 36 normal volunteers. The clinical diagnosis was used as the known outputs in data mining.

The ten-fold stratified cross-validation method was applied to evaluate the performance of data mining for the methods and the features being investigated. The studies were randomly divided into ten groups with the same number of samples, and stratification was applied to reduce

variation of random sampling in grouping. For one instance, 9 groups of data were randomly selected to train the regression model and the remaining group of data was used in the test. The procedure was repeated ten times where each group was used once.

Quantitative parameters were derived from the ten-fold stratified cross-validation. The accuracy was derived by the number of correctly classified cases divided by the total number of cases. However, the accuracy doesn't take the number of correct classifications due to chance into account, which is a critical issue in the classification. Thus, the Kappa statistics, Kappa = $(N_O-N_R)/(N_T-N_R)$, was used to measure the agreement between predicted and actual classifications, where N_O is the number of correct predictions, N_T is the number of total cases, and N_R was the number of correct classifications by chance.

3. RESULTS

3.1. Scalar features

Fig.1 shows the effect of the number of PCA components for $Mean_{Glu}$ using the LR method. Initially, the performance of the classification improved with increasing number of PCA components. However, the classification tended to degrade as the number of components exceeded 23. This is not unexpected due to the high order PCA components provide little differentiating information and most likely reflect noise and gave rise to failed regression in data mining. In contrast, too few components could not disclose sufficient information for data mining.

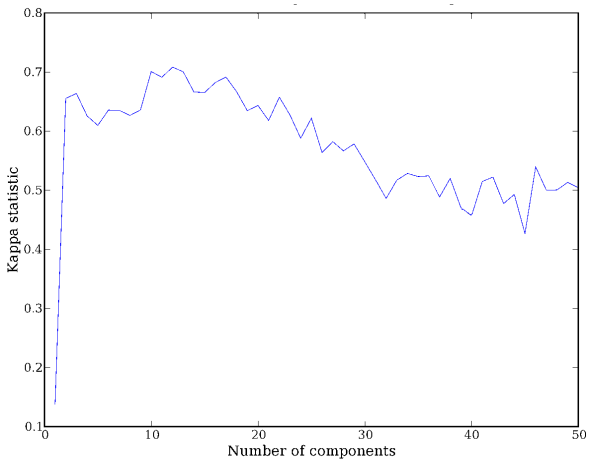

Fig.1 Plots of Kappa statistics for $Mean_{Glu}$ as a function of number of PCA components

The optimal value for $Mean_{Glu}$ was 0.71 with 12 components of PCA as shown in Fig.1. Similar trends were observed in data mining for other features with PCA, where the optimal components ranged from 11 to 19 to achieve highest value of the Kappa statistics.

Table.1 lists the accuracy and Kappa statistics for all the scalar features. The LR method gave the worst results and the Kappa statistics were lower than 0.59 because too many features were used in the regression; these findings are similar to those in Fig.1. The ALR method improved classification with the Kappa statistics above 0.65. This was mainly due to the involved boosting technique markedly improved the performance of classification. The LR method with PCA showed substantially improved overall performance in contrast to the results without PCA with the exception of the joint feature. This finding dose not contradict the conclusion in [7], because the joint feature in this study was a simple approach.

Table.1 Classification results: Accuracy % (Kappa statistics)

	$Mean_{Glu}$	$Mean_{Gra}$	$NMean_{Glu}$	$Mean_{Joint}$
LR	66% (0.47)	64% (0.46)	64% (0.43)	74% (0.59)
ALR	78% (0.65)	78% (0.65)	82% (0.71)	79% (0.66)
PCA-LR*	82% (0.71)	79% (0.66)	80% (0.68)	75% (0.60)

* The optimal number of components with the highest Kappa statistics

Interestingly, the ALR method with intensity normalization achieved better classification for $NMean_{Glu}$ while the results of the LR method with PCA were better for $Mean_{Glu}$. This implies that the classification method is dependent upon a selected feature in data mining. Furthermore, these findings indicate that intensity normalization with appropriate classification may achieve a similar performance to quantitative parametric images without intensity normalization.

3.2. Histogram features

Fig.2 shows the effect of the number of bins in histogram analysis using LR with PCA. Histogram analysis with 12 bins showed the largest variation as a function of varying upper bound. The analysis of 2 bins achieved a similar poor predication when the upper bound was less than 4 or greater than 12. This implies that the patterns of CMRGlc for data mining of dementia may be within the range from 4 to 12.

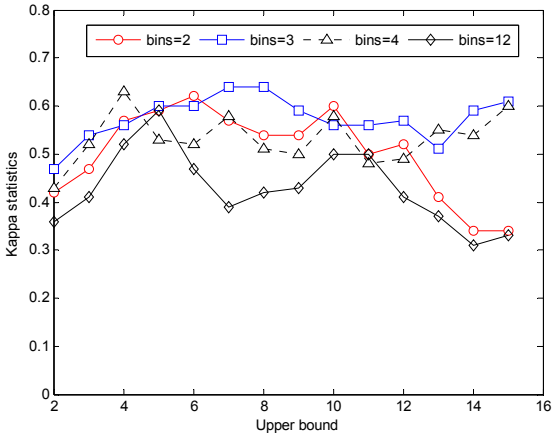

Fig.2 Plots of Kappa statistics for histogram features as a function of the upper bound in histogram analysis*

* The optimal number of components with the highest Kappa statistics

The results with 3-bin histogram analysis were observed to be more reliable with the kappa statistics around 0.6, while the upper bounds varied from 5 to 9. The best result was 0.64 with the upper bound of 7 and 8. This was identical to the clinical findings of hypometabolism having values lower than 10 mg/min/100g. This indicates that the histogram analysis has the potential to extract information from quantitative CMRGlc for better classification.

3.3. Voxel-level feature

The effect of classification for voxel-level features of CMRGlc is shown in Fig.3, where the LR method was used with PCA. The Kappa statistics was decreased when the number of PCA components was higher than 25. The optimal number of components was 15 with the accuracy of 83% and the Kappa statistics of 0.73 (see Fig.3).

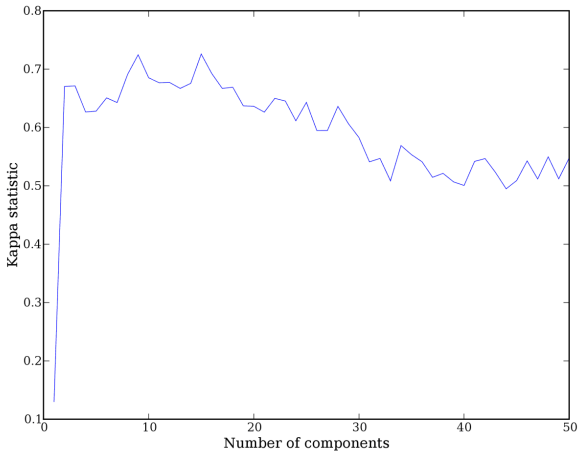

Fig.3 Plots of Kappa statistics of classification of voxel-level feature as a function of number of PCA components

Overall, the classification using voxel-level features with PCA and LR method achieved the best prediction in this study. Although this approach requires the optimal number of PCA components, it does not require definition of ROIs and knowledge about the pattern of disease involvement in dementia. Thus we suggest this approach is superior in any automated classifications of the dementias.

4. CONCLUSION

We investigated the effect of data mining of parametric images of CMRGlc in AD, FTD and normal data with scalar, joint, histogram and voxel-level features as well as the LR and ALR methods. Our results show that CMRGlc was efficient in the automated classification of dementia and data mining using voxel-level features with PCA and the LR method achieved the best classification. Further investigations will focus on the derivation of a probabilistic classification of dementia as well as comparisons with different data mining techniques and combined features.

5. ACKNOWLEDGEMENT

This work was supported in part by ARC, PolyU/UGC grants.

6. REFERENCES

[1] M. D. Devous, "Functional brain imaging in the dementias: role in early detection, differential diagnosis, and longitudinal studies," European Journal of Nuclear Medicine and Molecular Imaging, vol. 29, pp. 1685-1696, 2002.

[2] V. Megalooikonomou, J. Ford, L. Shen, et al., "Data mining in brain imaging," Statistical Methods in Medical Research, vol. 9, pp. 359-394, 2000.

[3] M. P. A. Page, R. J. Howard, J. T. Obrien, et al., "Use of neural networks in brain SPECT to diagnose Alzheimer's disease," Journal of Nuclear Medicine, vol. 37, pp. 195-200, 1996.

[4] 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.

[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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] M. E. Daubewitherspoon, R. E. Carson, and M. V. Green, "Post-Injection Transmission Attenuation Measurements for PET," IEEE Transactions on Nuclear Science, vol. 35, pp. 757-761, 1988.

[12] 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.

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

[14] 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.

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

[16] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," Machine Learning, vol. 59, pp. 161-205, 2005.

[17] I. H. Written and E. Frank, Data Mining: Practical machine learning tools and techniques, 2 ed. Amsterdam; Boston: Morgan Kaufman, 2005.