32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Three-Fold Cross-Validation of Parkinsonian Brain Patterns

Phoebe G. Spetsieris, *Member, IEEE*, Vijay Dhawan, and David Eidelberg

Abstract—**Abnormal physiological networks of brain areas in disease can be identified by applying specialized multivariate computational algorithms based on principal component analysis to functional image data. Here we demonstrate the reproducibility of network patterns derived using positron emission tomography (PET) data in independent populations of parkinsonian patients for a large, clinically validated data set comprised of subjects with idiopathic Parkinson's disease (iPD), multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). Correlation of voxel values of network patterns derived for the same condition in different data sets was high. To further illustrate the validity of these networks, we performed single subject differential diagnosis of prospective test subjects to determine the most probable case based on a subject's network scores expressed for each of these distinct parkinsonian syndromes. Three-fold cross-validation was performed to determine accuracy and positive predictive rates based on networks derived in separate folds of the composite data set. A logistic regression based classification algorithm was used to train in each fold and test in the remaining two folds. Combined accuracy for each of the three folds ranged from 82% to 93% in the training set and was approximately 81% for prospective test subjects.**

I. Introduction

DESPITE enormous progress in medical imaging, the underlying functional pathology of many brain diseases remains to a large extent obscure, masked by primarily normal structural features. Functional modalities such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) generate maps of physiological parameters that can reveal abnormalities, but pathological differences between normal and disease states and between distinct pathological states are often subtle. The primary methods used to analyze functional neuroimaging data are based on mass univariate statistical models as opposed to multivariate methods of analysis [1]. However, in parkinsonism there are numerous unrelated conditions that are indistinguishable in early stages [2] and differences from expected norm values are minimal. In these cases there is an advantage in using multivariate methods of analysis that examine the covariance of voxels once mean effects are removed rather than mass univariate methods of analysis that rely on statistical differences in global mean normalized voxel values to determine areas of disease involvement [1],[3]-[6]. The scaled subprofile model of principal component analysis (SSM/PCA) is a multivariate approach that has revealed invaluable information regarding disease pathways in parkinsonism and other disorders and has provided powerful computational tools for single subject diagnosis [7]-[12].

In previous studies [10]-[12] we used scalar comparison and discriminant analysis of SSM/PCA subject scores derived from patient and control PET image data for multiple class differential classification using specialized techniques. Here, we focus on examining the reproducibility of disease specific patterns in independent populations. SSM-PCA networks are derived and tested prospectively in a new clinically validated data set. Representative resting state network patterns are determined for three distinct parkinsonian patient classes, idiopathic Parkinson's disease (iPD), multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). The voxel by voxel correlation of corresponding networks derived in different folds of the data set is demonstrated. Further, we employ customized MATLAB based routines to evaluate diagnostic performance based on comparisons of group accuracy and positive predictive values (PPV) in different subsets of the data. Logistic regression based classification [13] was used primarily because of its wider acceptance although performance evaluation compared to discriminant analysis based methods has generally shown small differences in results for accuracy [14]. Three-fold cross-validation was applied to SSM subject scores determined for each condition. A rigorous approach was used by requiring networks to be regenerated in each training fold. The determined classification parameters were then applied to test the subjects in the remaining two folds. Accuracy and predictive values for score dependent classification are compared for network sets in each fold.

II. Methods

A. Data Set

Positron emission tomography (PET) data of patients and normal controls in the resting state was acquired on a GE Advance tomograph (Milwaukee, WI, USA) at North Shore University Hospital [15] using *(F-18)2-fluoro-2-deoxy-D-glucose* (FDG) radiotracer. The PET data was initially spatially normalized into a common stereotactic space (MNI) using SPM software [16] so that there is a one to one correspondence of voxels between subjects.

The subjects tested in this analysis belong to a clinically, carefully validated data set that was classified by a gold standard of long term follow-up. This group is a composite

P. G. Spetsieris, V. Dhawan and D. Eidelberg are with the Center for Neurosciences, Feinstein Institute for Medical Research, North Shore – LIJ Health System, Manhasset, NY, 11030 USA and with the Department of Neurology and Medicine, New York University School of Medicine, New York, USA (516-562-1166; fax: 516-562-1008; email: pspetsie@nshs.edu, dhawan@nshs.edu, david1@nshs.edu).

set of 167 patients (96 iPD, 41 MSA and 30 PSP). Demographic details are presented in Table I and in a prior study of positive predictive methods [9]. For the three-fold cross-validation study performed here, the group was subdivided into three sets of approximately 56 subjects {32 PD, 14 MSA, 10 PSP} subjects, designated as F1, F2, and F3. Each subset of patients was paired with an equal number of normal controls.

TABLE I
GROUP DEMOGRAPHICS

Group [No. Subjects]	No. of Male/ Female	Age Yrs (stdev)	Disease Duration Yrs (stdev)
Group F1			
iPD [32]	22/10	55.0 (8.7)	6.2 (4.1)
MSA [14]	4/10	62.7 (6.7)	4.1 (1.5)
PSP [10]	5/5	70.1 (6.7)	2.8 (1.4)
Group F2			
iPD [32]	18/14	57.6 (11.4)	4.0 (3.9)
MSA [13]	7/6	56.2 (8.8)	4.7 (3.0)
PSP [10]	4/6	70.0 (7.1)	2.6 (1.3)
Group F3			
iPD [32]	26/6	59.6 (8.2)	7.7 (4.7)
MSA [14]	7/7	64.0 (9.0)	2.9 (1.3)
PSP [10]	5/5	68.0 (8.5)	3.7 (1.5)
Normal			
NRM [32]	16/16	53.8 (13.5)	
NRM [14]	6/8	49.0 (10.1)	
NRM [13]	6/7	49.6 (10.3)	
NRM [10]	5/5	48.1 (8.7)	

B. SSM/PCA Basic Formulation

SSM analysis was performed on a personal computer, MATLAB (ver. 2006b) based platform (Mathworks, Sherborn, MA) using customized software (scanvp ver. 6.1) [17]. This neuroimaging toolbox [10], [12] incorporates MATLAB and C-language based image processing and statistical routines.

The SSM-PCA method is an established procedure based on a regional approach detailed in [7] and [18] and extended to voxel based approaches summarized in [8], [10] and [19]. The basic steps are reviewed here:

Spatially normalized image data corresponding to paired groups of patients and normal controls is thresholded or masked to remove low value voxels and artifacts. Each subject's data $\mathbf{P_j}$ is mapped into a single row vector *j* of the group data matrix. The group matrix is log transformed and doubly centered by subject voxel (row) centering followed by group voxel (column) centering. These operations insure the removal of irrelevant global, subject mean (GMR_j) and group mean profile (**GMP**) effects [1], [5], [7]. As a result data for each subject *j* is represented by the doubly centered residual row *j* values of the group matrix described as the subject residual profile $\mathbf{SRP_j}$ (1).

$$\mathbf{SRP_j} = \text{Log } \mathbf{P_j} - GMR_j - \mathbf{GMP} \quad (1)$$

A singular value decomposition (SVD) is performed on the subject by subject (*row by row*) covariation matrix, derived from the group data matrix. Each of the resulting eigenvectors accounts for a variable percent (*vaf*) of the total variance, and is associated with subject dependent scores as previously described [7], [8], [10], [19]. Consequently, for each subject *j* the SRP_j data can be represented as a linear combination of orthogonal principal components (PCs) described as group invariant subprofile image vectors or patterns ($\mathbf{GIS_k}$) and associated subject scores ($Score_{kj}$).

$$\mathbf{SRP_j} = \sum_k Score_{kj} \times \mathbf{GIS_k} \quad (2)$$

SSM GIS image values do not represent absolute metabolic values but are patterns of the covariance structure of combined normal and patient data. Z-scored GIS voxels represent standard deviations above the vector normalized GIS voxel mean. Each ($\mathbf{GIS_k}$) image pattern is tested for disease related significance by analyzing the *t-test p* value of associated patient/normal scores. In cases where disease effects are distributed over more than one PC a linear combination is obtained using multiple regression of associated scores. Once a GIS or linear combination of GIS vectors is established as a disease related pattern, individual subject scores for prospective patients can be determined as the inner product of the transpose of the subject SRP vector and the GIS vector. (This relationship is evident from (2) considering the orthogonality of GIS vectors):

$$Score_{kj} = \mathbf{SRP^T_j} \cdot \mathbf{GIS_k} \quad (3)$$

It has been shown that subject scores correlate with clinical measures of individual disease status and can be used to evaluate the efficacy of interventional treatment [8]. Moeller, et al. [20] demonstrated the reproducibility of regional SSM patterns in independent populations of iPD patients. Here, we verify these results in voxel space for a large group of new subjects including two additional parkinsonian syndromes. Further, we demonstrate the cross validation of differential diagnostic test scores based on sets of independent disease networks derived in separate folds of the combined patient and normal data set.

III. RESULTS

A. Three-Fold Derivation of GIS Network Patterns

Network patterns were generated in each fold (F1, F2, F3) for each of the three parkinsonian syndromes using corresponding patients and a matching number of normal subjects. In each case, composite disease specific GISs were constructed as a linear combination of one or more component PCs that satisfied the criterion $p < 0.15$, *vaf* > 9%. The resultant GISs had values: $p < 1.5 \times 10^{-5}$ and *vaf*> 23% (Fig. 1). The squared values of Pearson's linear correlation coefficients (R^2) for network patterns derived in different folds were calculated on a *voxel by voxel* basis for common non-zero voxels (Table II).

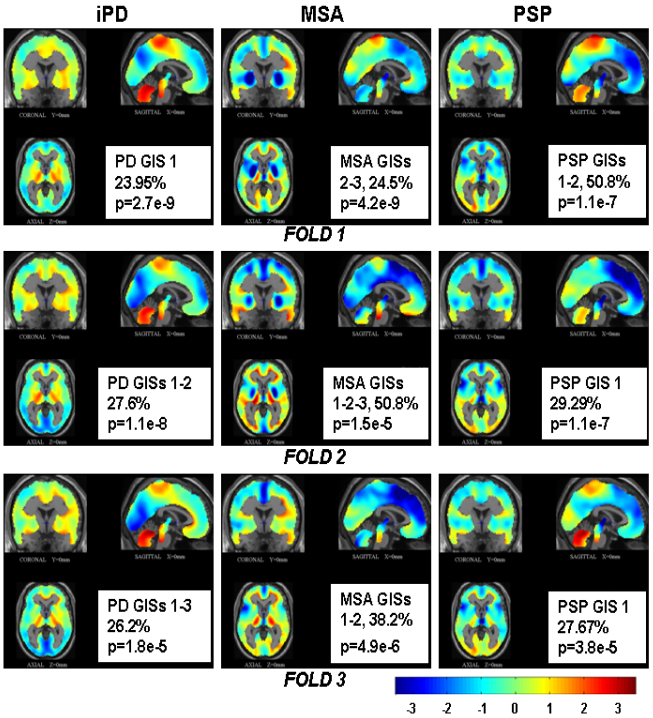

Fig. 1. Characteristic parkinsonian metabolic brain network patterns for iPD, MSA and PSP derived using SSM/PCA in corresponding groups of patient and normal subject PET metabolic image data repeated for three separate patient cohorts (Folds 1, 2, 3). The *vaf* and derivation group *p* values for each GIS or combination of GISs are noted (where, e=10). The displays represent cross-sectional planes (axial, coronal and sagittal) through the origin of normalized MNI space on a structural MRI background.

TABLE II
R^2 VOX CROSS-CORRELATION OF PATTERNS DERIVED IN DIFFERENT FOLDS

FOLD	*I* (iPD)			*M* (MSA)			*P* (PSP)		
	1	2	3	1	2	3	1	2	3
I									
1	1.00	0.75	0.71	0.03	0.02	0.03	0.24	0.07	0.40
2	0.75	1.00	0.85	0.04	0.05	0.04	0.13	0.04	0.22
3	0.71	0.85	1.00	0.02	0.04	0.01	0.10	0.01	0.21
M									
1	0.03	0.04	0.02	1.00	0.40	0.36	0.38	0.38	0.18
2	0.02	0.05	0.04	0.40	1.00	0.54	0.20	0.37	0.17
3	0.03	0.04	0.01	0.36	0.54	1.00	0.26	0.55	0.21
P									
1	0.24	0.13	0.10	0.38	0.20	0.26	1.00	0.65	0.66
2	0.07	0.03	0.01	0.38	0.37	0.55	0.65	1.00	0.49
3	0.40	0.22	0.21	0.18	0.17	0.21	0.66	0.49	1.00

It can be seen that patterns derived for iPD in different folds were highly correlated with each other ($.71 \le R^2 \le .85$). The correlation for MSA patterns was significant but not as strong ($.36 \le R^2 \le .54$). Good correlation was evident for PSP patterns ($.49 \le R^2 \le .66$). Patterns derived for iPD were not correlated with MSA patterns ($R^2 \le .05$). Some iPD correlation was evident with two of the PSP patterns ($.01 \le R^2 \le .4$). The PSP patterns were also correlated to some extent with the MSA patterns ($.17 \le R^2 \le .55$). The somewhat high values in these cases illustrate the extent of common areas affected in these diverse conditions also reflected in the difficulty in discriminating patients.

B. *Cross Validation Based on Brain Pattern Scores*

We applied K-fold cross-validation (K=3) [13] to subject score sets for disease networks derived in each of the folds (F1, F2, F3). The classification algorithm (based on MATLAB's ver. R2006b logistic *mnrfit* and *mnrval* routines and customized software) was trained using the SSM scores of subjects used to derive the networks in each fold and the resulting training parameters were used for re-evaluating the training subjects as well as testing the subjects in the remaining two folds. Individual class sensitivity and predictive values and composite patient values are presented in Table III for the train and corresponding test set subjects.

TABLE III
THREE-FOLD TRAIN/TEST CLASS SENSITIVITY AND PPV

	iPD Sens. (PPV) %	MSA Sens. (PPV) %	PSP Sens. (PPV) %	Composite Patient ACCURACY % PPV**
TRAIN FOLD	[32]	[14]*	[10]	[56]*
1	97 (97)	86 (86)	90 (90)	93
2	97 (91)	77 (83)	80 (89)	89
3	94 (88)	79 (79)	50 (63)	82
Mean	96 (92)	81 (83)	73 (81)	88
TEST FOLDS	[64]	[28]*	[20]	[112]*
(2, 3)	94 (82)	67 (82)	60 (75)	81
(1, 3)	91 (84)	64 (82)	75 (71)	81
(1, 2)	97 (83)	61 (77)	55 (73)	80
Mean	94 (83)	64 (80)	63 (73)	81

The number of subjects in each group is shown in brackets.
* The number of MSA subjects in the train set for fold 2 is 13 and total patients are 55. For test sets (2, 3) and (1,2) MSA subjects are 27 and total subjects are 111.
** Composite accuracy and PPV values are the same.

Sensitivity was evaluated as the ratio of correct predictions to the total number of class subjects. The positive predictive value (PPV) for each class is the ratio of correct positive predictions to total (true and false) positive predictions. Classification was assigned for each subject based on maximum predicted probability.

IV. DISCUSSION

Fig. 1 visually illustrates the commonality of features specific to different disease patterns derived in independent data sets using SSM/PCA methodology. These observations are numerically supported in Table II by the *voxel by voxel,* R^2 correlation values of patterns. High and low areas help us pinpoint distinguishing markers of each disease and direct us to areas for therapeutic intervention [8]. Evidence of shared common features for different conditions is also apparent in some cases which may explain the difficulty in distinguishing early cases clinically. Results shown in Table

III for the most likely differential diagnosis based on distinct disease networks derived and tested in independent data sets showed notably similar outcomes. As expected accuracy and PPV values were greater in the derivation sets used for training. Composite values ranged from 82% to 93%. The overall accuracy in the test sets was approximately the same in each case (mean 81%). However, there were some differences in sensitivity and PPV values for the individual disease groups with the best values achieved for the much larger iPD group (train: 96 (92)%, test: 94 (83)%).

These results demonstrate the basic invariance of disease networks derived in independent patient populations using SSM/PCA methodology and help validate the status of such network patterns as disease biomarkers. Error rates and differences between groups are a reflection of the diversity of subjects and disease characteristics within different groups and are more pronounced for the smaller groups. For differential diagnosis, where false positive results can have detrimental consequences in treatment, improvement in both accuracy and positive predictive values is essential. In addition to using larger and better matched cohorts of patients and controls, and perhaps including additional PCA components to derive patterns, higher sensitivity can be obtained by using masking constraints and other strategies to derive modified metabolic differential networks dedicated to diagnosis [10], [11]. On the other hand, higher PPV values can be achieved at the expense of lower sensitivity by the exclusion of underdetermined cases with predictive classification probability below a predetermined cutoff as illustrated by Tang [9]. In future studies we hope to improve rates by incorporating these and other concepts.

The overall robustness of networks derived using SSM/PCA methodology is significant. By contrast, voxel-wise effects using mass univariate analysis are not direct representations of functional connectivity and are not as reproducible because they can include more global or cohort dependent non-disease related statistical variance which can overshadow disease related features [1], [3], [4]. Various newer MRI based approaches are still in early stages of development [21]. In addition to providing important information regarding the spatial topography of disease pathways, SSM/PCA provides individual scores for quantifying single-subject disease status that is valuable in diagnostic support, especially in early cases, and in the evaluation of drug and treatment efficacy.

ACKNOWLEDGMENT

Many thanks to Dr. Thomas Chaly, Claude Margouleff, Spiro Kavathas and Ralph Mattachieri for PET and Cyclotron support and Rosie Persaud for administrative assistance.

REFERENCES

[1] K. M. Peterson, T. E. Nichols, J. B. Poline, A. P.Holmes, "Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models," *Philos. Trans. R. Soc. Lond., B Biol. Sci.* vol. 354, 1999, pp. 1239-1260.

[2] E. Tolosa, G. Wenning, W. Poewe, "The diagnosis of Parkinson's disease," *Lancet Neurol*., vol. 5, pp. 75-86, 2006.

[3] C. Habeck, and Y. Stern, "Neural network approaches and their reproducibility in the study of verbal working memory and Alzheimer's disease," *Clin. Neurosci. Res.*, vol. 6, no. 6, pp. 381-390, 2007.

[4] C. Habeck, N. L. Foster, R. Perneczky, A. Kurz,, P. Alexopoulos, R. A. Koeppe, A. Drzezga, Y. Stern, "Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease," *Neuroimage*, vol. 40, pp. 1503-1515, 2008.

[5] I. T. Joliffe, *Principal Component Analysis*, 2nd edition, New York: Springer Series in Statistics, Springer-Verlag, 2002, pp. 388-391.

[6] A. C. Rencher, *Methods of Multivariate Analysis.* New York: Wiley-Interscience, 1995

[7] G. E. Alexander, J. R. Moeller, "Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: A principal component approach to modeling brain function in disease," *Hum. Brain Mapp.* vol. 2, pp. 1-16, 1994.

[8] D. Eidelberg, "Metabolic brain networks in neurodegenerative disorders: a functional imaging approach," *Trends in Neurosciences,* vol. 32, no. 10, pp. 548-57, 2009.

[9] C. C. Tang, K. L. Poston, T. Eckert, A. Feigin, S. Frucht, M. Gudesblatt, V. Dhawan, M. Lesser, J. P. Vosattel, S. Fahn, D. Eidelberg, "Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis," *Lancet Neurol.*, vol. 9, no. 2, pp. 149-58, 2010.

[10] P. G. Spetsieris. Y. Ma, V. Dhawan, D. Eidelberg, "Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features," *Neuroimage,* vol. 45, no. 4, pp.1241-1252, 2009.

[11] P. G. Spetsieris. Y. Ma, V. Dhawan, T. Eckert, D. Eidelberg, "New strategies for automated differential diagnosis of degenerative brain disorders," in *Conf. Proc. 2007 IEEE Eng. Med. Biol. Soc.,* pp. 3421-3425.

[12] P. G. Spetsieris. Y. Ma, V. Dhawan, J. R. Moeller, D. Eidelberg, "Highly automated computer-aided diagnosis of neurological disorders using functional brain imaging," in Proceedings of SPIE, Medical Imaging: Image Processing, Reinhardt JM, Pluim JPW, eds., San Diego, 2006, vol. 6144, pp. 61445M1-12.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2001, ch. 4 ch.7, pp. 214-216.

[14] R. Higdon, N. L. Foster, R. A. Koeppe, C. S. DeCarli, W. J. Jagust, C. M. Clark, et al., "A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's," *Statistics in medicine*, vol. 23, pp. 315-326, 2004.

[15] T. Eckert, A. Barnes, V. Dhawan, S. Frucht, M. Gordon, A. S. Feigin, D. Eidelberg, "FDG PET in the differential diagnosis of parkinsonian disorders ", *Neuroimage*, vol 26, no. 3, pp. 912-921, 2005.

[16] spm99. Available: http://www.fil.ion.ucl.ac.uk/spm

[17] scanvp, ver. 6.1. Available: http://www.feinsteinneuroscience.org/

[18] D. Eidelberg, J.R. Moeller, V. Dhawan, P. Spetsieris, S. Takikawa, T. Ishikawa, et al., "The metabolic topography of parkinsonism," *J Cerebral Blood Flow and Metabolism,* vol. 14, pp. 783-801, 1994.

[19] Y. Ma, C. Tang, P. Spetsieris, V. Dhawan, D. Eidelberg., "Abnormal metabolic network activity in Parkinson's disease: test-retest reproducibility," *J Cerebral Blood Flow and Metabolism*, vol. 27, pp. 597-605, 2007.

[20] J. R. Moeller, T. Nakamura, M. J. Mentis, V. Dhawan, P. Spetsieris, A. Antonini, J. Missimer, K. L. Leenders, D. Eidelberg, "Reproducibility of regional metabolic covariance patterns: comparison of four populations," *J. Nucl. Med.,* vol. 40, pp. 1264-1269, 1999.

[21] G. Rizzo, P. Martinelli, D. Manners, C. Scaglione, C. Tonon, et al., "Diffusion-weighted brain imaging study of patients with clinical diagnosis of corticobasal degeneration, progressive supranuclear palsy and Parkinson's disease," *Brain*, vol. 131, pp. 2690-2700, Oct 2008.