
Comparison of Voxel- and Volume-of-Interest-Based Analyses in FDG PET Scans of HIV Positive and Healthy Individuals

Jeih-San Liow, Kelly Rehm, Stephen C. Strother, Jon R. Anderson, Niels Mørch, Lars Kai Hansen, Kirt A. Schaper, and David A. Rottenberg

Departments of Radiology and Neurology, University of Minnesota, Minneapolis; PET Imaging Center, Veterans Affairs Medical Center, Minneapolis, Minnesota; and Denmark Technical University, Copenhagen, Denmark

Abnormal glucose metabolic patterns in the brain have been reported for HIV-1 seropositive (HIV+) patients when compared with seronegative healthy individuals. The metabolic covariance pattern obtained from voxel- or volume-of-interest (VOI)-based multivariate data analysis techniques can be used to characterize disease and potentially to detect and monitor disease progression in the early stage of AIDS dementia complex. However, the arbitrary smoothing typically applied to PET data during reconstruction and processing to reduce noise has an unknown effect on the data, especially for the voxel-based analysis. **Methods:** To investigate the impact of smoothing on a discrimination task, we applied principal component analysis with scaled subprofile-model preprocessing (SSM/PCA) followed by Fisher discriminant analysis to FDG PET data that were reconstructed and processed with different degrees of smoothing. Receiver operating characteristic curves were used to compare the ability of derived metabolic covariance patterns to discriminate HIV+ patients from healthy volunteers. **Results:** For the voxel-based analysis, we found that the spatial distribution of voxel weights from the SSM/PCA analysis suggested edge effects along major tissue and cerebrospinal fluid boundaries, indicative of a disease-specific pattern of cerebral atrophy for the HIV+ patients. In terms of its discrimination performance, this metabolic covariance pattern is stable and insensitive to a wide range of smoothing kernels, except for ramp reconstruction and Hanning reconstruction with $7 \times 7 \times 7$ block smoothing. In these reconstructions, the discrimination performance decreased as a result of high image noise and excessive smoothing, respectively. Our results also indicated that if sufficient variance from the VOI measurements is included, the overall performance of a conventional VOI-based analysis can be similar to that of the voxel-based analysis for the same discrimination task. However, the VOI-based analysis performed poorly at low false-positive fraction and is less tolerant to noise in the metabolic covariance pattern than the voxel-based analysis. **Conclusion:** We have obtained a unique covariance pattern of brain glucose metabolism for HIV+ patients compared with healthy volunteers. Discrimination based on this covariance pattern was found to be insensitive to a wide range of image smoothness.

Key Words: HIV; scaled subprofile model; Fisher discriminant analysis; receiver operating characteristic

J Nucl Med 2000; 41:612-621

Several FDG PET studies of patients with the AIDS dementia complex (ADC) have shown regional hypermetabolism in the basal ganglia and thalamus compared with healthy individuals (1,2). These results, which suggested abnormal cortical-subcortical functional connectivity, can be used to characterize the disease process and to quantify disease severity and progression in the early stages of ADC. Multivariate statistical analysis followed by discriminant analysis can provide a more comprehensive view of the signature metabolic covariance pattern (2-9), which may then be used to classify individual HIV seropositive (HIV+) patients into ADC and non-ADC groups.

Traditionally, volume-of-interest (VOI)-based analysis techniques have been used in studying FDG PET data. In these techniques, a series of VOIs is defined on a standard template to which each individual's reconstructed volume is registered. Because most VOIs are relatively large and involve more than a few voxels, VOI-based techniques tend to be insensitive to smoothing at the voxel level. This kind of smoothing has typically been applied either during or after reconstruction to reduce noise or processing errors such as intersubject misalignment. One major drawback of the VOI-based analysis is that the differences in VOI size have implicitly imposed a spatially varying smoothing effect. The effect on the resolution and noise for local (small) structures within the VOI during the analysis is unknown. Another potential disadvantage for the VOI-based technique is that some important information may be overlooked, because large areas of brain are not sampled and analyzed. Voxel-based analysis techniques (10-14) have gained in popularity because of their robustness in processing and their sensitivity to small structures, compared with VOI-based techniques. In addition, the whole-brain volume rather than isolated regions is included in the analysis. However, by treating each voxel as a VOI not only is a voxel-based

Received Mar. 1, 1999; revision accepted Aug. 9, 1999.
For correspondence or reprints contact: Jeih-San Liow, PhD, PET Imaging Center (11P), VA Medical Center, One Veterans Dr., Minneapolis, MN 55417.

analysis more computationally demanding but the results also become increasingly dependent on the structure of small signals and background noise, which can be significantly affected by smoothing. Therefore, when using voxel-based data analysis techniques it is important to optimize the parameter for smoothing during the reconstruction and processing steps.

This article reports on a study involving a subset of individuals from a previous investigation (2). We compared the functionalized FDG PET scans reconstructed by filtered backprojection (FBP) with a range of different filter cutoffs, and postprocessed with different degrees of smoothing on the basis of group discrimination between HIV+ and healthy individuals. For discriminating among groups, we applied a scaled subprofile-model preprocessing followed by principal component analysis (SSM/PCA) and Fisher discriminant analysis (FDA). Receiver operator characteristic (ROC) curves were generated using the Fisher discriminant (FD) score to evaluate the impact of smoothing on the discrimination with a leave-1-out cross validation. Results were compared with those from the VOI-based SSM/PCA and FDA. Our objectives were first to investigate how smoothing parameters in reconstruction and processing impact the voxel-based analysis of group FDG PET data, and, second, to compare the voxel-based technique with the conventional VOI-based technique, for this specific discrimination task.

MATERIALS AND METHODS

SSM/PCA

The first part of our analysis is based on variance partitioning of the groups of cerebral metabolic rate of glucose (CMRglc) images. Data were preprocessed using the SSM followed by PCA (2,3,12,15). SSM/PCA is briefly summarized in the following expressions. The general form of the model for CMRglc images can be expressed as:

$$v_{iq} = g_q(r_i + z_{iq} + \epsilon_{iq}), \quad \text{Eq. 1}$$

where v_{iq} is the i th voxel from the q th subject ($q = 1..Q$), g_q is a global scaling factor, r_i is the mean image for all subjects after removal of g_q , z_{iq} is the voxel subject interaction, and ϵ_{iq} is random error. If we denote a data matrix i_{iq} that contains the interaction plus error after normalization (divided by g_q) and removal of mean image r_i , and apply PCA to this matrix:

$$i_{iq} = z_{iq} + \epsilon_{iq} = \sum_{k=1}^Q h_{ik} (\epsilon_k)^{1/2} s_{kq} = \sum_{k=1}^Q h_{ik} \text{SSF}_{kq}, \quad \text{Eq. 2}$$

where $\text{SSF}_{kq} = (\epsilon_k)^{1/2} s_{kq}$, the outcome of PCA yields orthonormal eigenimages h_{ik} , eigenvectors s_{kq} , and a scaling factor $(\epsilon_k)^{1/2}$. We shall define the scaled eigenvector $[(\epsilon_k)^{1/2} s_{kq}]$ as the subject scaling factor (SSF_{kq}). Compared with other similar multiplicative models, the most important difference for SSM/PCA is the assumption that the global scaling factor g_q is independent and uncorrelated with the SSFs and therefore requires an additional step to estimate (3).

FDA

For the discrimination task, we used FDA. The FD function, a linear model that maximizes the ratio of the between-group means

to the within-group variances using a set of input variables, can be written as:

$$y = \mathbf{X}\mathbf{a}, \quad \text{Eq. 3}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$. The input variables for each subject are $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iq}]$. The discriminant function \mathbf{a} is a linear vector obtained from the maximization, such that $y_i = \mathbf{x}_i \mathbf{a}$ for all subjects, satisfying the maximum separation condition described above. Details regarding FDA can be found in general statistics textbooks (16). In this article, we will use the SSFs (scaled eigenvectors) obtained from SSM/PCA as the input variables \mathbf{x} for FDA. We will also refer to y as the FD score and \mathbf{a} (the discriminant function) as the Fisher coefficients.

Patient Data

We analyzed 75 FDG PET scans from 20 HIV+ patients (ages, 43 ± 9 y; 10 patients with AIDS; some of the HIV+ patients with follow-up scans) and 41 healthy volunteers (ages, 45 ± 15 y; 3 with repeated scans). These data are a subset of the original 85 scans/subjects reported previously (2). Ten scans/subjects were removed from the study because of a coverage problem for the voxel-based analysis. Dynamic scans were acquired using an ECAT 953B/31 scanner (Siemens/CTI, Knoxville, TN) in 2-dimensional mode (17). Data were reconstructed by FBP with a Hanning filter cutoff at 80% of the Nyquist frequency (1.28 cycles/cm [Han04]). Attenuation was corrected with a transmission scan obtained using rotating rod sources. No scatter correction was applied. Gray-matter rate constants were estimated from the upper 25% of voxel values for each plane and a 3-compartment model (2). The average gray-matter rate constants from the 31 planes were then used to convert the last frame (30–40 min after injection) of the dynamic images into CMRglc images. To investigate the effect of different reconstruction filters, the same 75 PET FDG scans were also reconstructed by FBP with a ramp filter (Ramp) and 2 additional Hanning filters cut off at 100% (Han05) and 60% (Han03) of the Nyquist frequency, respectively. However, CMRglc images for the additional reconstructions were generated using the same rate constants obtained previously with the Han04 images.

Voxel-Based Processing and Analysis

A trained discrimination using all patients and volunteers was performed first. The sequence of processing and analyzing data is outlined in the following: (a) Intersubject registration with respect to a standardized PET volume in Talairach space (18) was accomplished using a 12-parameter version of the automatic image registration algorithm (AIR) (19). (b) To reduce the potential sensitivity of voxel-based SSM/PCA to registration errors, additional block smoothing was applied to all volumes before they were resampled into Talairach space. Three different block sizes ($3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ voxels) applied only to the Han04 reconstruction were investigated. (c) Before SSM/PCA, a global mask was obtained from a summed image of the 75 registered volumes reconstructed by Han04 with an upper 75% threshold. This mask was applied to all volumes for all reconstructions, in an attempt to exclude voxels of cerebrospinal fluid (CSF) from the analysis. Because of variations in axial coverage, voxels that did not exist in all 75 volumes were also rejected. (d) For each of the 4 reconstructions and 3 smoothings, SSM/PCA was applied to the 75 volumes. SSFs were obtained and used as input to FDA to calculate the FD score for each individual. Discrimination of HIV+ patients and healthy volunteers was initially compared using a single threshold. Accuracy, defined as the number of subjects that were

correctly classified in both groups (i.e., true-positive fraction [TPF] + true-negative fraction [TNF]), was compared for all reconstructions, smoothings, and number of SSFs used in calculating the FD scores. The effect of SSF number on the discrimination was studied in increments of 10 to provide a selection criterion for the number of SSFs used as input to FDA.

ROC Curves

To evaluate classification accuracy at different thresholds, i.e., discrimination that allows tradeoffs between types of classification error, ROC curves based on a double-Gaussian model of the population were fitted to the FD scores (20). For all reconstructions and smoothings, a paired 2-tailed *t* test that takes into account the correlation of data was applied to examine the significance of differences in ROC curves, both in overall TPF (area under the curve) and for TPF at different low false-positive fractions (FPFs).

Cross Validation

The discrimination performance estimated directly from all subjects is generally too optimistic. To obtain a less biased estimate of discrimination accuracy, we used a leave-1-out cross-validation technique similar to the jack-knife procedure (21,22). The basic idea of this technique is to omit 1 data point at a time, build the model with the remaining data, and see how well the model predicts the excluded point (8,23). We first dropped 1 patient and applied SSM/PCA to the remaining 74 patients, followed by FDA. The SSFs of the excluded patient were then estimated using the g_q and r_i (Eq. 1) obtained from the other 74. A "predicted" FD score was calculated by multiplying the estimated SSFs of the excluded patient with the Fisher coefficients also obtained from FDA of the 74 patients. For each reconstruction and smoothing, this process was repeated for all 75 patients and all volunteers. ROC curves were constructed using the "predicted" FD scores and evaluated. We will refer to this new set of ROC curves as "Tested Discrimination," in contrast with the "Trained Discrimination" (ROC curves) obtained using all 75 patients.

VOI-Based Analysis

A conventional VOI-based SSM/PCA was applied to a set of 40 VOIs (Table 1) for Han04 reconstruction. In this VOI-based analysis, the 40 VOI values from each individual replaced the voxel values as input to SSM/PCA. No masking or postreconstruction smoothing was applied. The SSFs were used to calculate the FD scores followed by ROC evaluation. Leave-1-out cross validation similar to that for the voxel-based analysis was performed.

RESULTS

Image Resolution

Examples of CMRglc images of a healthy volunteer obtained using the reconstruction and smoothing filters discussed in this study are displayed in Figure 1. All images shown are after reformatting into Talairach space before the analysis. The resolution (measured from a 2-dimensional line source) ranged from 0.6 cm full width at half maximum (FWHM) for ramp filter to 0.82, 0.95, and 1.18 cm for Han05, Han04, and Han03, respectively. With additional block smoothing applied to the Han04 images, the FWHMs became 1.15, 1.57, and 2.07 cm for Han04_3 \times , Han04_5 \times , and Han04_7 \times , respectively. The shape of the line source

TABLE 1
VOIs Defined in VOI Analysis

VOI area
Midbrain*
Caudate (L and R)*
Putamen (L and R)*
Angular gyrus (L and R)*
Cuneus (L and R)*
Supramarginal gyrus (L and R)*
Anterior cingulate gyrus (L and R)*
Cerebellum (L and R)
Hippocampus (L and R)
Inferior frontal gyrus (L and R)
Inferior temporal gyrus (L and R)
Medial frontal gyrus (L and R)
Operculum (L and R)
Orbitofrontal gyrus (L and R)
Postcentral gyrus (L and R)
Posterior cingulate gyrus (L and R)
Precentral gyrus (L and R)
Superior temporal gyrus (L and R)
Thalamus (L and R)
Transverse temporal gyrus (L and R)
Pons

*VOIs that received the heaviest weighting by FDA analysis.

remained essentially Gaussian after block smoothing. As a result of the resampling process, a slight degradation of the resolution was observed in the Talairach image compared with the raw reconstructed image.

Variance Distribution

Figure 2A depicts the variance contributed by each of the 75 SSFs (i.e., the eigenvectors) after SSM/PCA for all reconstructions. The variance decreases rapidly after the first 5 SSFs, approaching a constant but different background asymptotically for different reconstructions. The Ramp reconstruction has the highest variance among the reconstructions, followed by Han05, Han04, Han03, Han04_3 \times , Han04_5 \times , and Han04_7 \times . This trend is consistent with the decreasing resolution (increasing smoothing) in the reconstructed images and their corresponding visual appearance (Fig. 1). Also plotted is the variance of the VOI-based SSM/PCA, which lies between Han04_5 \times and Han04_7 \times . Figure 2B illustrates the cumulative variance as a fraction of the total variance (sum of the variances for all SSFs in Fig. 2A). Whereas the total variance decreases as the smoothing increases, the relative contribution of variance from the early SSFs increases. The smoother the image becomes, the larger the fraction of variance appears in the early SSFs. The VOI-based analysis has the highest cumulative variance within its 40 SSFs. (Note that the number of SSFs from SSM/PCA is determined by the smaller of either the number of individuals/scans or the number of voxels/VOIs. Therefore, there are only 40 instead of 75 SSFs for the VOI-based analysis.)

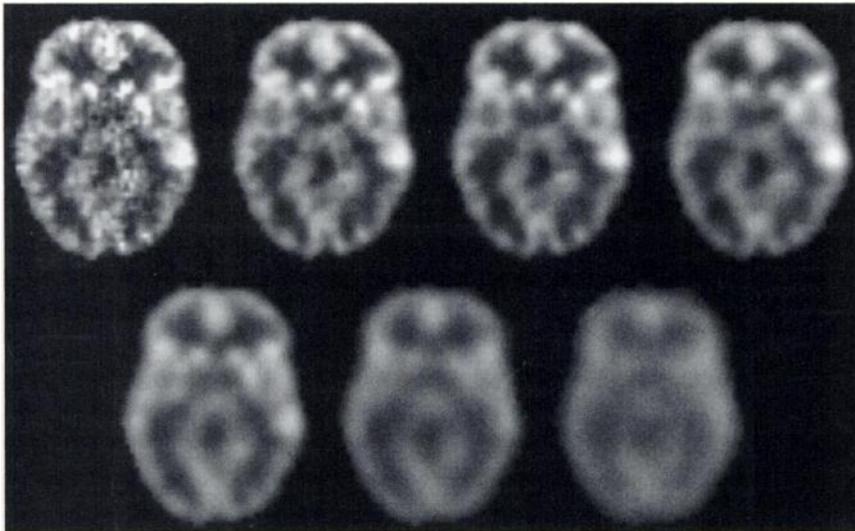


FIGURE 1. CMRglc images of healthy volunteer reconstructed with Ramp; Han05, Han04, and Han03 (top row, left to right); and Han04 with additional $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ block smoothing (bottom row, left to right). All images were created after resampling in Talairach space and scaled by global scaling factor.

Accuracy

A trained discrimination was first performed using all 75 scans. For all reconstructions, FD scores were calculated using the first 10 and 20 SSFs. Figure 3A illustrates an example of FD scores from Han04 reconstruction with 20 SSFs (accounts for $\approx 55\%$ of the total variance) plotted as a

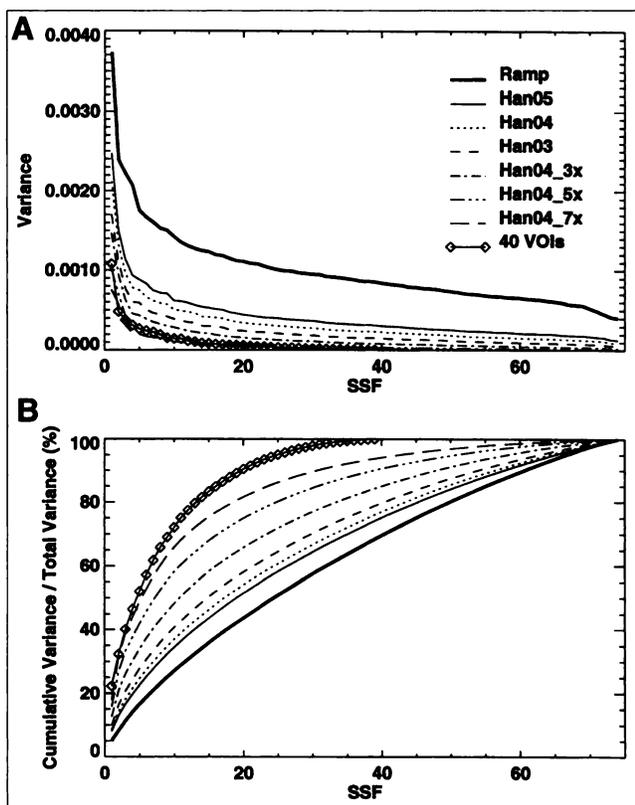


FIGURE 2. (A) Variance contribution from each of SSFs after voxel-based SSM/PCA. (B) Cumulative variance as percentage of total variance for all reconstructions and smoothings and for VOI-based SSM/PCA with Han04.

function of the individuals' age. There was no apparent age dependency for FD scores in either healthy or HIV+ individuals. When a discrimination threshold was selected at the mean of the 2 group means, the healthy volunteers and HIV+ patients were separated with an accuracy (defined as the percentage of cases correctly labeled) of 92%. Accuracies based on the group-mean threshold are given in Table 2 for all reconstructions using the first 10 and 20 SSFs. The overall accuracy decreased as the number of SSFs used decreased from 20 to 10. Presumably, this occurred because less information was available in the model for discrimination. For 20 SSFs, the accuracy did not vary significantly across reconstructions (91%–93%). For 10 SSFs, the accuracy was the worst for Ramp (77%) and improved but remained similar as the smoothing increased (83%–89%) up to Han04_7 \times . The decreased accuracy for Han04_7 \times in both 10 (81%) and 20 (91%) SSFs relative to other reconstructions indicated that the data were probably over smoothed and that structures essential to the discrimination may have been suppressed. The VOI-based analysis using 40 VOIs resulted in an accuracy of 87% for 10 SSFs and 93% for 20 SSFs (Table 2). These accuracies are comparable with those obtained from the voxel-based technique.

Up to this point, the same set of individuals had been used both in training the model and in prediction. With the leave-1-out cross validation, Figure 3B illustrates the predicted FD scores as a function of the individuals' age for Han04 calculated using 20 SSFs. Compared with trained discrimination of the same reconstruction (Fig. 3A), individuals are more tightly clustered by the predicted FD scores in this tested discrimination, and less separation is achieved between the 2 groups. The accuracy of tested discrimination dropped 5%–13% for all reconstructions using 20 SSFs and 7%–15% using 10 SSFs (Table 2). However, 3 trends remained consistent: (a) the accuracy of overall discrimination for the VOI-based analysis was comparable with that of

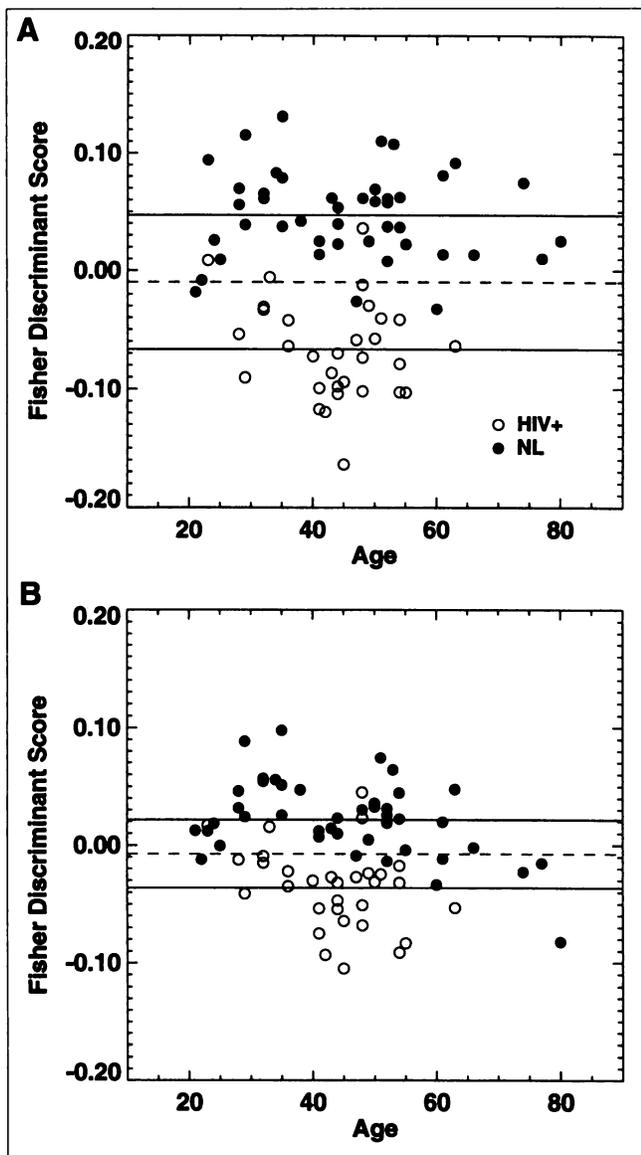


FIGURE 3. Trained FD scores (A) and predicted FD scores (B) calculated using 20 SSFs for Han04 reconstruction plotted versus individuals' age. Two solid lines are group means for healthy volunteers and HIV+ patients. Dashed line represents average of 2 group means.

the voxel-based analysis; (b) Ramp and Han04_7× reconstructions were worse than the other reconstructions, as a result of no smoothing and excessive smoothing, respectively; and (c) there was no apparent correlation between the predicted FD scores and ages of individuals.

ROC Curves

The choice of a single threshold at the midpoint between the 2 group means is optimal only if both population distributions are Gaussian with the same FWHM. In practice, this assumption is rarely satisfied, and it is therefore more useful to discuss discrimination with tradeoffs, such as those represented by the ROC curves. Table 3 lists the area under the ROC curves obtained with trained FD scores for

TABLE 2

Accuracy of Trained and Tested Discrimination for Different Reconstructions and Smoothings Using 10 and 20 SSFs

Reconstructions and smoothings	Accuracy (%)			
	Trained		Tested	
	SSF 1–10	SSF 1–20	SSF 1–10	SSF 1–20
Ramp	77	92	69	80
Han05	89	92	79	87
Han04	88	92	79	84
Han03	83	92	77	84
Han04_3×	89	93	79	87
Han04_5×	87	92	81	85
Han04_7×	81	91	69	85
VOI	87	93	81	83

all reconstructions using the first 10 and 20 SSFs. As in measures of accuracy, the area under the ROC curves, an indicator of overall discrimination power, did not vary significantly across different reconstructions. For 10 SSFs, the area was the smallest for Ramp (0.86), increased as the smoothing increased, and decreased again for Han04_7× (0.91), possibly as a result of excessive smoothing. For 20 SSFs, all areas increased compared with 10 SSFs, and the differences among reconstructions were insignificant. For the VOI-based analysis using 40 VOIs, the area under the ROC curves increased from 0.94 with 10 SSFs to 0.97 with 20 SSFs (Table 3). Once again, the performance of the VOI-based method was comparable with the voxel-based methods in discriminating HIV+ patients from healthy volunteers. For the tested discrimination, the area dropped 7%–11% for 20 SSFs and 8%–15% for 10 SSFs compared with the trained discrimination for all reconstructions (Table 3).

In the trained discrimination, as the number of SSFs used continued to increase, the area under the ROC curve approached unity (perfect discrimination) after 35 SSFs (results not shown). It was not surprising that with all 75 scans, perfect discrimination could be achieved if enough SSFs (variance) were included. However, this is usually not

TABLE 3

Area Under ROC Curve for Trained and Tested Discrimination for Different Reconstructions and Smoothings Using 10 and 20 SSFs

Reconstructions and smoothings	Area under ROC curve			
	Trained		Tested	
	SSF 1–10	SSF 1–20	SSF 1–10	SSF 1–20
Ramp	0.86	0.97	0.73	0.86
Han05	0.94	0.98	0.85	0.90
Han04	0.95	0.98	0.86	0.91
Han03	0.94	0.97	0.84	0.90
Han04_3×	0.95	0.98	0.87	0.91
Han04_5×	0.94	0.97	0.87	0.90
Han04_7×	0.91	0.96	0.83	0.90
VOI	0.94	0.97	0.91	0.90

the case for the tested discrimination because of the problem of over fitting the noise. In addition, the shape of the ROC curves, which could have very different implications in terms of the discrimination, is not revealed by the overall area and therefore must be examined also. Figure 4 illustrates ROC curves generated with predicted FD scores calculated using 10, 20, 30, and 40 SSFs. ROC curves were obtained for up to 70 SSFs but are not shown because the changes were insignificant beyond 40 SSFs. In general, the areas increased as more SSFs were added in calculating the FD scores. This area increase was also associated with a

change in the shape of the ROC curve, consisting of a slight decrease of TPF at high FPF but a significant increase of TPF at low FPF, which occurred between 20 and 30 SSFs (Fig. 4C). Although some reconstructions appeared to be slightly better than others, depending on the number of SSFs used, paired 2-tailed *t* tests indicated that only 2 differences were significant ($P < 0.05$). If only 10 SSFs were used, the area under the ROC curve for Ramp reconstruction was significantly smaller than that of all other reconstructions ($P < 0.003$). For Han04_7x reconstruction, the area under the ROC curve was initially comparable with other recon-

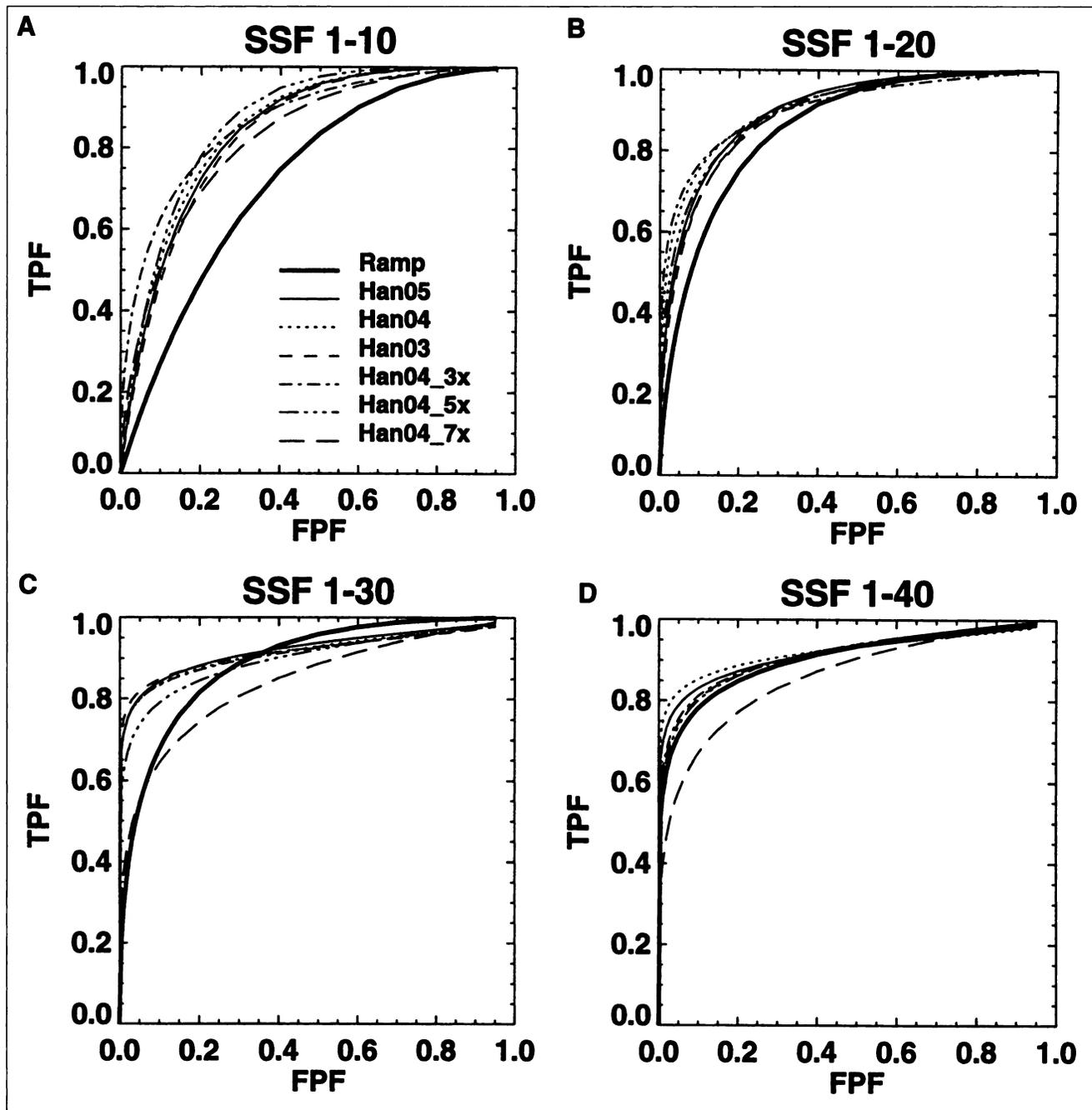


FIGURE 4. ROC curves of predicted FD scores for 4 reconstructions and 3 smoothings calculated using 10 (A), 20 (B), 30 (C), and 40 SSFs (D).

structions but became significantly worse after 30 SSFs. For Ramp reconstruction, the changes in shape at the high and low FPF observed for other reconstructions also did not occur until 30 SSFs. Figure 5 illustrates the ROC curves calculated using the predicted FD score for VOI-based analyses. The area under the ROC curve was similar for 2 and 5 SSFs, increased for 10 and 20 SSFs, and then decreased rapidly for 30 and 40 SSFs. The sharp decrease in area for 30 and 40 SSFs occurred simply because too many SSFs were included, and noise as well as signals were fitted in the model, degrading the ability to discriminate. For this VOI-based analysis, 2-tailed *t* test on the areas under the ROC curves indicated that 40 SSFs were significantly worse than all other SSFs ($P < 0.0001$), and 10 SSFs were significantly better than all other SSFs ($P < 0.08$). Unlike the voxel-based analysis, no obvious shape changes in ROC curves were observed for VOI-based analysis as the number of SSFs increased.

To further assess the effect of shape changes in the ROC curves on the discrimination performance, comparison of TPF at 2 specific low FPFs (low number of misclassified healthy volunteers) was made for all reconstructions. Figure 6 illustrates TPF at FPFs of 0.15 and 0.05 taken from the ROC curves in Figures 4 and 5. For FPF = 0.15, all voxel-based analyses showed major increases in TPF between 10 and 30 SSFs. For all but Ramp and Han04_7 \times , this improvement leveled out after 30 SSFs, reaching a TPF of 0.85–0.90 at 70 SSFs (Fig. 6A). Both Ramp and Han04_7 \times eventually were able to achieve a TPF of about 0.80 at 70 SSFs but were worse than all other reconstructions. For FPF = 0.05, the same trend was observed (Fig. 6B). For VOI-based analysis at FPF = 0.15, the TPF was 0.70 initially, rose to 0.80 at 10 SSFs, but dropped rapidly after 20 SSFs. At FPF = 0.05, the TPF started at 0.50,

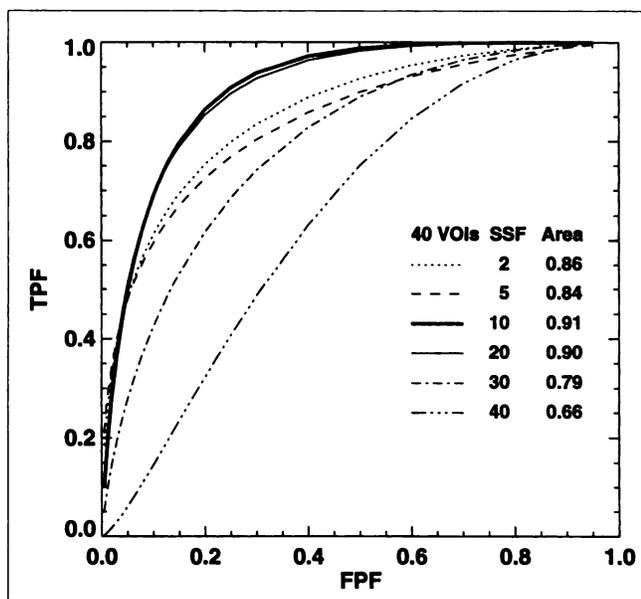


FIGURE 5. ROC curves of predicted FD scores for VOI-based analysis using number of SSFs ranging from 2 to 40.

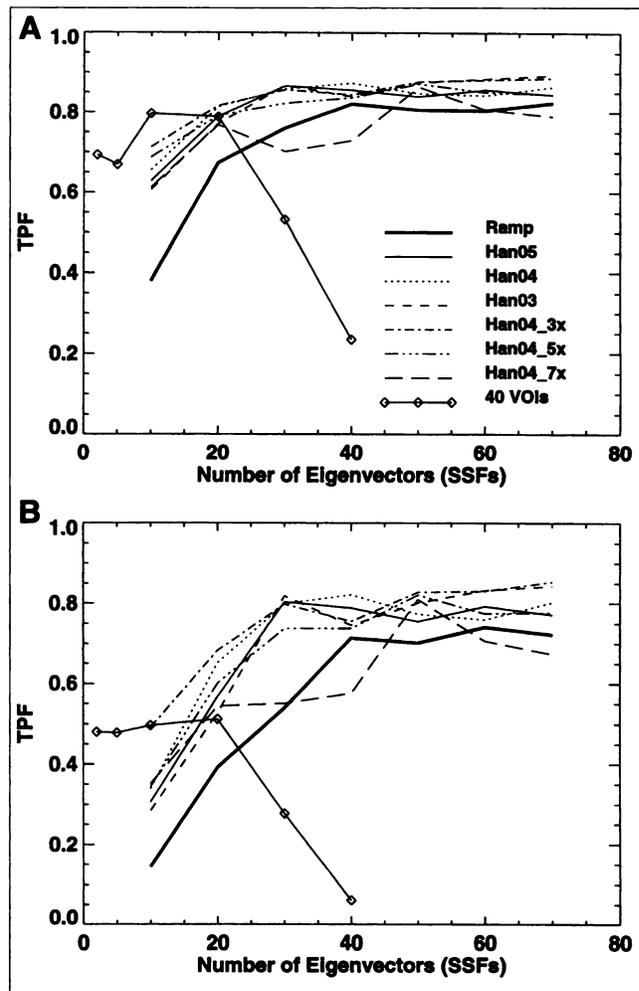


FIGURE 6. TPFs as function of number of SSFs used at FPFs of 0.15 (A) and 0.05 (B) for tested discrimination of all 4 reconstructions, 3 smoothings, and for VOI-based analysis.

followed by a sharp decline after 20 SSFs. Two-tailed *t* tests applied to TPF at both FPFs showed that for the VOI-based analysis, 40 SSFs were significantly worse than all other SSFs ($P < 0.0001$ for FPF = 0.15, $P < 0.0008$ for FPF = 0.05).

Eigenimage

Figure 7 shows selected slices of the corresponding FDA eigenimage calculated using the first 20 SSFs for all reconstructions and smoothings. Only the positive portion of the eigenimage is shown, depicting decreased rates of glucose metabolism (lower voxel weights) for HIV+ patients compared with healthy volunteers. For HIV+ patients, these areas of reduced glucose metabolism occur in anterior frontal, occipital, and perisylvian cortex; in cortex abutting the anterior and posterior interhemispheric fissures; and in relation to the anterior lateral and third ventricles. This pattern of relative hypometabolism suggests a disease-specific pattern of cerebral atrophy. The signature pattern of hypermetabolism in the putamen (higher voxel weights) for HIV+ patients was in the negative portion of the eigenim-

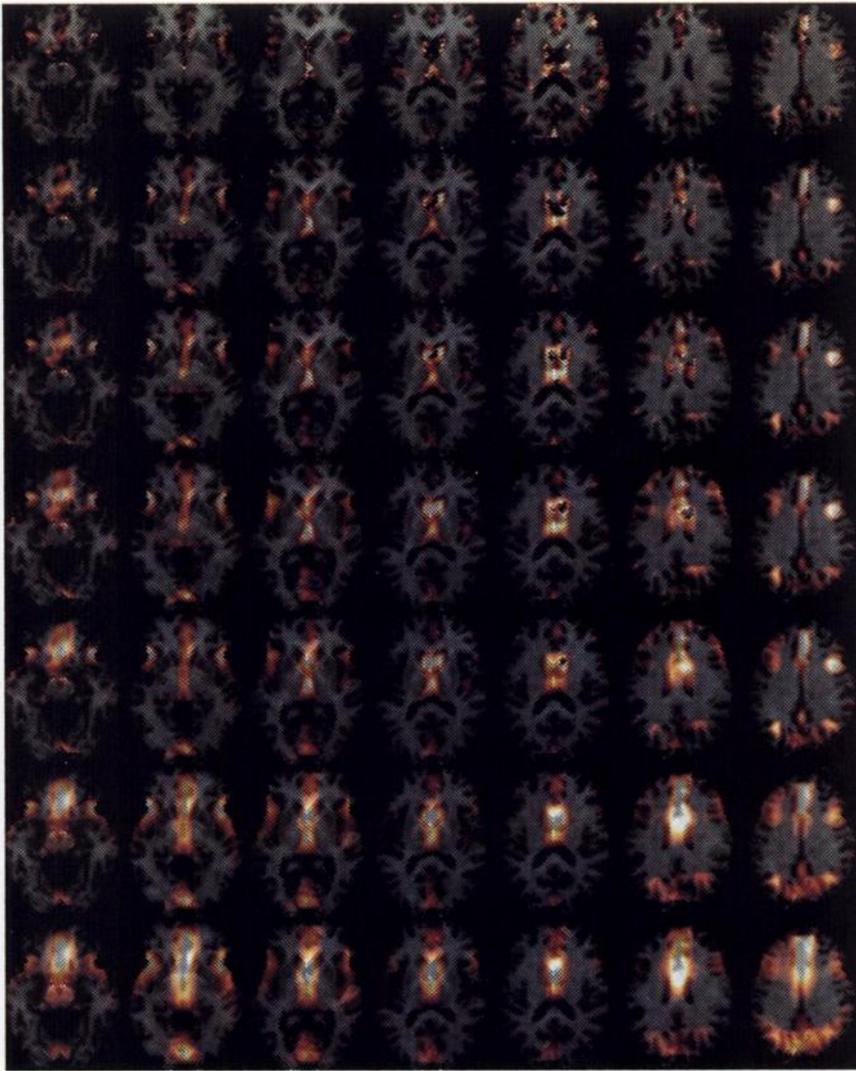


FIGURE 7. Selected slices of FDA eigenimage corresponding to trained FD scores calculated using 20 SSFs merged with T1-weighted MRI template for all reconstructions and smoothings. Eigenimages are (from top to bottom rows) Ramp; Han05, Han04, and Han03; and Han04 with additional $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ block smoothing during processing (Han04_3 \times , Han04_5 \times , and Han04_7 \times). Only positive portion of eigenimage is displayed. Intensity corresponds to reduced rates of glucose metabolism for HIV+ patients compared with healthy individuals.

age (not shown in Fig. 7) (2). This hypermetabolic pattern, believed to be specific for HIV+ patients, constitutes much less variance than the “atrophy effect.” The effect of smoothing on these eigenimages is generally similar to that on the original images (Fig. 1). For Ramp reconstruction, spatial locations of voxels with higher weights show a discontinuous pattern as the result of noise, making their interpretation with respect to structures difficult compared with other reconstructions (e.g., Han04). For Han04_7 \times reconstruction, the loss of spatial resolution on the pattern as a result of excessive smoothing was obvious. Although the ability to discriminate between healthy volunteers and HIV+ patients is relatively insensitive to most reconstructions and smoothings, as indicated previously by accuracy and ROC curves and by similarity in the covariance patterns, higher resolution images may still be more advantageous for the purpose of visual interpretation.

To determine how many of the significant voxels/VOIs that were involved in the discrimination between the voxel- and VOI-based analysis overlapped, we compared the voxels retained by thresholding the eigenimage of Han04 at

the 98th percentile with the mask of all 40 VOIs, as well as 13 VOIs that are heavily weighted in the discrimination according to VOI-based FDA (Table 1). Among the 436 “important” voxels retained by thresholding the eigenimage, 126 (28.9%) were not covered by the mask of all VOIs and 310 (71.1%) were not covered by the mask of the 13 significant VOIs. Most of the voxels left out by the VOI-based analysis again are located at the major tissue–CSF boundaries.

DISCUSSION

The advantage of a multivariate approach for the discrimination of different subject groups is that the effects of functional differences, errors, and confounds, as well as the interaction among them, are assessed at the voxel/VOI level. This means that l does not have to model regional interactions, for example, by performing a step-wise multiple regression between regions before characterization (8). The spatiotemporal correlations among regions are all explicitly included in the analysis. However, a potential disadvantage

is that too many variables are being fitted for the available data when including all spatiotemporal correlations. FDA is the 1-dimensional case CVA, a more general multidimensional approach that maximizes the separation between different groups based on a set of input variables. The outcome of this optimization is a set of scores known as canonical variates. These variates characterize the distributed signals. Because the selection of a threshold for discrimination in a multidimensional space is nontrivial, often only the 1-dimensional analysis is considered. In this study, SSFs from the output of SSM/PCA were used as the input variables for the 1-dimensional linear model, and classification was based on the resulting individuals' FD scores. Because FDA is a 1-dimensional process, it is possible to select a single threshold or multiple thresholds (depending on the problem posed), to classify each individual as belonging to a particular group.

For discrimination between healthy volunteers and HIV+ patients using SSM/PCA and FDA, our results indicate that some form of smoothing is necessary to reduce Poisson noise or intersubject variance potentially arising from registration errors. The impact of Poisson noise was clearly demonstrated by the fact that (in both the trained and the tested discrimination) Ramp reconstruction was worse than the Hanning reconstruction with some degree of smoothing. However, it seemed to make little difference whether the smoothing was performed through filters during reconstruction or through a simple block smoothing after reconstruction. Our results also show that, as long as the data are not over smoothed (as in Han04_7 \times), all Hanning reconstructions (over a wide range of spatial resolution) appear to perform similarly when used as the basis for group discrimination. This could imply that the metabolic covariance pattern that differentiates HIV+ patients from healthy volunteers is sufficiently robust that image resolution has little or no impact on it. The discrimination performance decreased for Han04_7 \times compared with other Hanning reconstructions, suggesting that the intersubject variance that allows for group discrimination may have been reduced by over smoothing. This is certainly very likely if the intersubject variance is produced by edge effects as a result of cerebral atrophy, as indicated by the spatial patterns in the eigenimage (Fig. 7). Although our studies showed that, for most reconstructions, resolution does not seem to play a critical factor in their discrimination performance, for the purpose of visualization, higher resolution images may still be desirable so that the functional information can be better mapped to anatomic structures for interpretation. Similar discrimination performance obtained from data both with and without postreconstruction smoothing (e.g., Han04 versus Han04_3 \times and Han04_5 \times) also suggests that for this dataset, registration errors may not be as significant as initially expected.

For all voxel-based analyses in the tested discrimination, even though the TPF at low FPF shows little or no decrease as the number of SSFs increases, the number of SSFs used to

calculate the FD score should be limited to avoid over fitting. The SSFs obtained from SSM/PCA are arranged in order of decreasing variance. However, not all SSFs are necessarily useful for discriminating between HIV+ and healthy individuals. The fact that some SSFs represent noise or signal unrelated to group discrimination may help explain why the TPFs oscillate after 30 SSFs (Fig. 6). Instead of adding the SSFs in order of variance and by the increment of 10, as we did during our evaluation, an ongoing research project is focusing on adapting more sophisticated strategies, such as applying a *t* test to each SSF to determine its significance for the separation of the 2 groups, followed by selecting a subset of SSFs based on some *P* value threshold to calculate the FD scores. The appropriate smoothing, regardless of whether performed during or after reconstruction, can effectively help achieve a similar purpose by reducing the noise contribution to the variance as well as simultaneously moving signal variances into the early SSFs (Fig. 2). This supposition is supported by the fact that the TPFs for most Hanning reconstructions rose quickly to 85% after 30 SSFs, whereas it required 40 SSFs for Ramp reconstruction to reach the same level (Fig. 6).

The spatial pattern of the eigenimage (Fig. 7) showed that a major portion of the variance occurs along the edges of several large structures. Because it is not possible for the 12-parameter AIR registration to create local misalignments and the FD scores did a reasonably good job in discriminating HIV+ from healthy individuals even for images with little smoothing, we believe that the eigenimage pattern obtained along major tissue boundaries is real and disease related. The pattern possibly reflects tissue loss in those areas resulting from cerebral atrophy (2). However, given that the FD scores for the HIV+ patients do not correlate well with their ages or with atrophy scores based on their structural MR images (i.e., some HIV+ subjects have large negative FD scores but have no physical atrophy) (2), another possibility would be a loss of tissue functionality (reduced glucose metabolism).

Compared with the voxel-based analysis, the VOI-based analysis is simpler and less computationally demanding. However, unlike the voxel-based analysis, in which the discrimination remains relatively stable over all SSFs, the discrimination for VOI-based analyses seems to be much more sensitive to noise and less tolerant to over fitting. Therefore, 1 must be careful when selecting the SSFs for VOI-based analysis. Although the overall discrimination (area under the ROC curve) for the VOI-based analysis with 40 VOIs appears to be as good as most of the voxel-based analyses in both the trained and tested cases, it underperforms significantly at low FPFs. For example, at FPF = 0.05, all voxel-based analyses eventually reached TPF > 0.7, whereas TPF for the VOI-based analysis was ≤ 0.5 . It is unclear whether this discrepancy is the result of the additional "important" voxels in the voxel-based analysis that are not covered in the VOI-based analysis. Finally, the voxel-based analysis provides a spatial pattern of the

difference between the 2 groups of individuals through its statistically informed eigenimage. Visualized in conjunction with anatomic images such as CT scans or MR images, the eigenimage can be very important in identifying and understanding the disease process. The use of higher resolution images does not necessarily improve the discrimination, although it might be preferable for visual interpretation.

CONCLUSION

To maximize discrimination between HIV+ patients and healthy volunteers using voxel-based SSM/PCA and FDA, our results indicate that moderate smoothing applied either through reconstruction filters or separately as a postprocessing operation is necessary. For this dataset, discrimination performance is insensitive to a wide range of image smoothness except for Ramp reconstruction, in which smoothing appears to be insufficient, and Hanning reconstruction with the additional $7 \times 7 \times 7$ block smoothing, in which smoothing becomes excessive. Our results also suggest that group discrimination based on a global measure, i.e., the entire image volume, is better than that achieved with VOI-based techniques in which only a few isolated regions are sampled.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grants R01-NS25701, R01-NS33718, R29-NS33721, and MH57180.

REFERENCES

1. Rottenberg DA, Moeller JR, Strother SC, et al. The metabolic pathology of the AIDS dementia complex. *Ann Neurol.* 1987;22:700-706.
2. Rottenberg DA, Sidtis JJ, Strother SC, et al. Abnormal cerebral glucose metabolism in HIV-1 seropositives with and without dementia. *J Nucl Med.* 1996;37:1133-1141.
3. Moeller JR, Strother SC. A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *J Cereb Blood Flow Metab.* 1991;11:A121-A135.
4. Clark CM, Ammann W, Martin WR, Ty P, Hayden MR. The FDG/PET methodology for early detection of disease onset: a statistical model. *J Cereb Blood Flow Metab.* 1991;11:A96-A102.
5. Kippenhan JS, Barker WW, Pascal S, Nagel J, Duara R. Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects. *J Nucl Med.* 1992;33:1459-1467.
6. Strother SC, Moeller JR, Sidtis JJ, Rottenberg DA. A disease-independent pattern of cerebral metabolic variation: the reticular activating system? In: Uemura K, Lassen NA, Jones T, Kanno I, eds. *Quantification of Brain Function: Tracer Kinetics and Image Analysis in Brain PET.* Amsterdam, The Netherlands: Elsevier Science; 1993:491-498.
7. Hill TC, Holman BL. Discriminant function analysis: toward a more rigorous approach to SPECT interpretation. *J Nucl Med.* 1994;35:1455-1456.
8. Azari NP, Pettigres KD, Pietrini P, Horwitz B, Schapiro MB. Detection of an Alzheimer disease pattern of cerebral metabolism in Down syndrome. *Dementia.* 1994;5:69-78.
9. Burn DJ, Sawle GV, Brooks DJ. Differential diagnosis of Parkinson's disease, multiple system atrophy, and Steele-Richardson-Olszewski syndrome: discriminant analysis of striatal ^{18}F -dopa PET data. *J Neurol Neurosurg Psychiatry.* 1994;57:278-284.
10. Friston KJ, Firth CD, Liddle PF, Frackowiak RSJ. Comparing functional (PET) images: the assessment of significant change. *J Cereb Blood Flow Metab.* 1991;13:5-14.
11. Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab.* 1992;12:900-918.
12. Strother SC, Anderson JR, Schaper KA, et al. Principal component analysis and the subprofile scaling model compared to intersubject averaging and statistical parametric mapping: I. Functional connectivity of the human motor system studied with ^{15}O water PET. *J Cereb Blood Flow Metab.* 1995;15:738-753.
13. Strother SC, Lange N, Savoy RL, et al. Multidimensional state-spaces for fMRI and PET activation studies. *Neuroimage.* 1996;3(part 2):S98.
14. Friston KJ, Poline JB, Holmes AP, Frith CD, Frackowiak RSJ. A multivariate analysis of PET activation studies. *Hum Brain Mapping.* 1996;4:140-151.
15. Eidelberg D, Moeller JR, Dhawan V, et al. The metabolic topography of Parkinsonism. *J Cereb Blood Flow Metab.* 1994;14:783-801.
16. Mardia KV, Kent JT, Bibby JM. *Multivariate Analysis.* San Diego, CA: Academic Press; 1979:318-320.
17. Spinks TJ, Jones T, Bailey DL, et al. Physical performance of a positron tomograph for brain imaging with retractable septa. *Phys Med Biol.* 1992;37:1637-1655.
18. Talairach J, Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme; 1988.
19. Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. *J Comput Assist Tomogr.* 1998;22:139-152.
20. Metz CE. FORTRAN programs LABROC and CLABROC (1996 distribution). Available from CE Metz, Department of Radiology, University of Chicago, Chicago, IL.
21. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol.* 1992;27:723-731.
22. Efron B. The jackknife, the bootstrap and other resampling plans. In: *Regional Conference Series in Applied Mathematics.* Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982:5-6.
23. Gilman S, St. Laurent RT, Koeppe RA, Junck L, Kluijn KJ, Lohman M. A comparison of cerebral blood flow and glucose metabolism in olivopontocerebellar atrophy using PET. *Neurology.* 1995;45:1345-1352.