Neural-Network Classification of Normal and Alzheimer's Disease Subjects Using High-Resolution and Low-Resolution PET Cameras

J. Shane Kippenhan, Warren W. Barker, Joachim Nagel, Cheryl Grady and Ranjan Duara

The Wien Center, Mt. Sinai Medical Center, Miami Beach, Florida; Departments of Biomedical Engineering Radiology and Neurology, University of Miami, Coral Gables, Florida; Laboratory of Neurosciences, National Institute on Aging, National Institutes of Health, Bethesda, Maryland

Neural-network classification methods were applied to studies of FDG-PET images of the brain acquired from a total of 77 "probable" Alzheimer's disease and 124 normal subjects at two different centers. **Methods:** Classification performances, as determined by relative-operating-characteristic (ROC) analyses of cross-validation experiments, were measured for FDG PET images obtained with either a 15-mm FWHM PETT V or a 6-mm FWHM Scanditronix PC-1024-7B camera for various methods of data representation. Neural networks were trained to distinguish between normal and abnormal subjects on the basis of regional metabolic patterns. For both databases, classification performance could be improved by increasing the "resolution" of the representation (decreasing the region size) and by normalizing the regional metabolic values to the value of a reference region (occipital region). **Results:** The optimal classification performance for Scanditronix data (ROC area = 0.95) was higher than that for PETT V data (ROC area = 0.87). Under Bayesian theory, the classification performance with Scanditronix data corresponded to an ability to change a pre-test probability of disease of 50% to a post-test probability of either 90% for a positive classification or 10% for a negative classification. **Conclusion:** This classification can be used to either strongly confirm or rule out the presence of abnormalities.

Key Words: neural network; Alzheimer's disease; fluorodeoxyglucose

J Nucl Med 1994; 35:7–15

Then diagnosis of a dementia such as Alzheimer's disease (AD) may involve neurological exams, neuropsychological testing, anatomical neuroimaging (x-ray computed tomography or magnetic resonance imaging) and other laboratory tests. Each of these modalities can be expected to contribute a certain amount of unique information with respect to the diagnosis of organic disease. Although in a large percentage of cases the diagnostic decisions based on these procedures are rather obvious, there are cases in which follow-up studies may be necessary to enable an accurate diagnosis, especially those involving distinctions between dementia and psychiatric disorders or normal aging. It is in these difficult cases that functional neuroimaging such as PET may be able to make an immediately valuable diagnostic contribution.

Quantitative techniques used to analyze PET scans are generally based on region of interest (ROI) analyses (1–5). The quantitative representations resulting from ROI analyses can be used as the basis of a system to objectively characterize PET scans as either normal or abnormal. Strother et al. (6) note that much of the research involving PET is directed toward improving the performance of particular links in the chain of highly complex data transformations that result in regional metabolic representations. The authors emphasize the importance of evaluating these advances in PET technology based on the degree to which they enhance PET's ability to detect disease. One important improvement over "first generation" PET cameras has been higher spatial resolution, which enhances the ability to accurately identify anatomical regions and reduces partial-volume effects.

The selection of a method or set of methods to be used for objectively measuring the diagnostic impact of these factors is important, since the choice of methods can greatly affect the outcome of a comparison. A particular type of classifier may be better suited for a normally distributed set of data than for a non-normally distributed set, for example, even though the "diagnostic information content" of the two sets may be the same.

In a pattern recognition setting, "signals" of interest must often be detected in the presence of "noise" and/or missing information. In the production of regional metabolic data using PET, any of the several "links" in the PET data chain represent potential sources of noise. Some sources, such as patient movement or problems with blood

	PETT V subjects	Scanditronix subjects
"Probable AD" subjects		
N	41	33
Age	70.9 ± 8.8 (range: 53–93)	65.8 ± 9.5 (range: 44–88)
Mini-Mental Status Exam Score	15.0 ± 7.3	15.1 ± 8.6
Gender (M,F)	21,20	20,13
Age-Equivalent normal controls		
N	50	74
Age	67.7 ± 8.9 (range: 50–84)	61.6 ± 10.8 (range: 45–90)
Gender (M,F)	25,25	34,40

collection or attenuation correction, are intrinsic to the PET acquisition process, while others, such as inaccuracies in anatomical ROI location, result from postacquisition analysis.

We have previously demonstrated that neural networks can be trained, on the basis of quantitative regional representations, to classify fluorodeoxyglucose (FDG) PET scans with an accuracy comparable to that of an expert PET reader (7,8). We describe here the application of neural networks to FDG-PET data acquired with two different PET cameras from patients with AD and from age-equivalent normal subjects, with particular emphasis on the diagnostic value associated with recent advances in PET technology. Diagnostic significance was determined directly by evaluating cross-validation classification performances for given conditions. The objectives of the investigation were:

1. To enable the formation of recommendations for optimizing the representation and analysis of PET data for the diagnosis of AD.
2. To demonstrate and compare the optimal ability of PET to discriminate between normal and AD subjects with the two cameras.
3. To determine the most generally applicable metabolic profiles that discriminate normal from AD subjects.

MATERIALS AND METHODS

Normal and AD subjects who underwent FDG-PET studies with a PETT V scanner (9) (inplane and axial image resolution of 15 mm) were recruited for brain imaging studies at the Wien Center, Mount Sinai Medical Center, Miami Beach. Recruiting procedures are described in detail elsewhere (10). A second group of normal and AD subjects were studied with a Scanditronix PC 1024-7B scanner (11) (inplane resolution of 6 mm, axial resolution of 10 mm) at the Laboratory of Neurosciences at the National Institute on Aging, National Institutes of Health, Bethesda, Maryland (12). At both sites, the patients used in these studies had been diagnosed with "probable AD" according to NINCDS-ADRDA criteria (13). Table 1 summarizes the composition of the two experimental groups.

PET scans for both PETT V and Scanditronix subjects were obtained in the "resting state" (in a quiet, darkened room with

eyes and ears occluded). Both procedures involved injection of approximately 5 mCi of [^{18}F]FDG, and both involved the use of head-positioning devices.

In the PETT V procedures, "arterialized" venous blood was collected in order to measure plasma radioactivity and glucose (14). Regional cerebral metabolic rate of glucose (rCMRglc) values were calculated using standard rate constants, a lumped constant of 0.42 and an operational equation (15). A contour-based attenuation correction procedure was performed. Data were collected for 67 ROIs in the brain using previously-published methods (16). For each region, the average metabolism in mg/100 g/min was determined. Values for certain regions were appropriately averaged to provide values for the 25 "lobules" shown in Table 2 and for 4 bilateral lobar regions (frontal, parietal, temporal and occipital).

In the Scanditronix procedures, arterial blood was sampled in order to measure plasma radioactivity and glucose (12). The operational equation of Brooks et al. (17) was used to estimate rCMRglc using a lumped constant of 0.42 (18). Transmission scans, performed prior to the injection of FDG, were used to perform attenuation correction. ROI procedures are described by Kumar et al. (12). For classification studies, small structures from the Scanditronix database were combined to obtain regional representations equivalent to those of the PETT V data at both lobular and lobar levels.

For quantitative analyses, subjects were thus represented by

TABLE 2
Identification of Regions Used in Lobule-Level Experiments

Region numbers	Region(s)
1,2	Right, left prefrontal
3,4	Right, left premotor
5,6	Right, left orbitofrontal
7,8	Right, left motor-sensory
9,10	Right, left superior parietal
11,12	Right, left inferior and medial parietal
13,14	Right, left occipital
15,16	Right, left superior temporal
17,18	Right, left mid-temporal
19,20	Right, left medial temporal
21,22	Right, left middle gray (Basal ganglia, thalamus)
23,34	Right, left cerebellum
25	Paracentral

their regional metabolic patterns. These n-dimensional vectors (n = 8 for lobar representations, n = 25 for lobular representations) served as inputs when training and testing the classifiers. Classification performances were evaluated on the basis of relative operating characteristics (ROC) analyses (19,20), in which the area under the ROC curve was used as the figure of merit. The ROC area measures a diagnostic system's performance at several different settings of the decision criteria, and is a more complete representation of a diagnostic system's performance than, for example, the report of a single pair of sensitivity and specificity values. Classification performances within each of the two databases were evaluated for both lobular and lobar representations for various methods of classification and data preprocessing.

Classifiers were evaluated by cross-validation studies similar to those described in our earlier work (8). In order to make the cross-validation results as general as possible (i.e., to reduce the results' dependence on any special properties of a particular combination of training and testing sets), different cross-validation configurations were used for each experiment. To create training and testing sets, normal subjects were randomly apportioned into two groups, as were the abnormal subjects. Either of the normal groups could be paired with either of the abnormal groups to form a training set, with the remaining subjects forming an independent set used for testing. Since training and testing sets could be interchanged, this resulted in four unique ways to perform cross-validation for each experiment. For neural-network experiments, training was repeated 20 times for each configuration, each time with random network initialization in order to eliminate any potential bias attributable to particular initial conditions. Each ROC value reported for neural network performance was thus the mean value of ROC results for 80 experiments (20 for each of four cross-validation configurations). "Within-database" cross-validation was performed for most of the experiments, i.e., training and testing sets were created from the subjects within one database. For some of the experiments, networks were either trained with patterns from both databases or were trained with patterns from one database and tested on patterns from the other. In all such cases, however, testing sets were kept strictly independent of training sets.

Neural network training was performed using back-propagation techniques described elsewhere (8,21,22). The regional metabolic patterns of all members of the training sets (resulting from ROI analyses of normal and AD PET scans) were repeatedly presented at the input layer. For each presentation, networks "learned" by comparing the calculated value of the output unit with the predetermined target values for the given pattern and then adjusting the internal weights so that the calculated outputs would be closer to the target values. Normal subjects were assigned a target value of 1.0. Abnormal subjects were assigned targets in a graded manner, in order to include dementia-severity information in the training. These targets were based on scaled Mini-Mental Status Exam scores (23), all of which were obtained within 30 days of the PET procedure, a period of time during which changes in mental status would have been minimal.

Once a network was trained in this way, "unknown" testing patterns were classified by comparing the network's output response (after presentation of the input pattern) to a numerical decision threshold. ROC curves were constructed by computing true-positive and false-positive ratios for a range of output decision thresholds. Detailed descriptions of the methods for training and optimization of the networks used in this study are given elsewhere (7,8,24).

Classification results for neural networks were compared to results using discriminant analysis, which were implemented with SAS statistical software (25), regional metabolic values as independent variables and diagnosis as the classification variable. To cross-validate, the discriminant function obtained for a "training set" was applied to patterns within a "testing set." The SAS procedure employed an optimization strategy which used either linear or quadratic discriminant analysis, depending on the results of tests of the intra-class and pooled covariance matrices (26). ROC curves were constructed by choosing a range of prior probabilities (from 0 to 1) for the discriminant procedure.

Several different methods were used to preprocess the data, each of which preserved certain aspects of the data's information content while eliminating others. Posthoc analyses of the classification results for different preprocessing methods thus allowed assessments of the relative importances of the different types of information present. The first method was simple scaling by a constant. This was necessary to avoid "saturation" of the transfer function used in the neural networks' processing units. Each value in the numerical representations was divided by 15.0, which scaled the maximum values to approximately unity. It is important to note that this operation (equivalent to representing the same data with different units) did *not* affect the covariance structure of the datasets—absolute metabolism information was retained. A second preprocessing method involved removal of the mean value from each pattern (the mean value of each n-dimensional pattern was subtracted from each of the n dimensions). Two additional preprocessing methods were used: global normalization, i.e., division of each value by the subject's global metabolism; and occipital normalization, i.e., the division of each value by that of the (assumedly unaffected) occipital region (3). The different methods of normalization have different implicit assumptions: removal of mean values would be expected to decrease the influence of spurious effects (noise) which were additive in nature, whereas normalization of data by division would be expected to counteract *multiplicative* noise. All but the first preprocessing method, however, affected the covariance structure of the data, removing any references to absolute metabolism.

Posterior "probabilities of disease" were calculated by employing concepts from Bayesian theory and information theory. Classification at the point of maximum "information content" on a given ROC curve (27) yielded an effectively dichotomous test with a specified sensitivity and specificity, and the conditional probability of disease, given the results of this classification and a specified prior probability, was calculated from Bayesian theory (28).

Methods to extract discriminating profiles from trained neural networks are described elsewhere (7). Briefly, weight vectors from many different trained networks were subjected first to analyses of sensitivity and orientation and then to clustering analyses to describe, on a probabilistic basis, the regions in the weight-vector space which represented the most important discriminating profiles.

RESULTS

Figures 1 and 2 illustrate the mean pattern vectors for the two subject groups within each database. Mean values of CMRglc are plotted by region, with a two-standard-deviation range shown (high = mean + 1 s.d., low = mean − 1 s.d.). Figure 1 shows the mean pattern vectors for lobar representations, while Figure 2 shows the mean vectors for lobular representations. At both levels of represen-
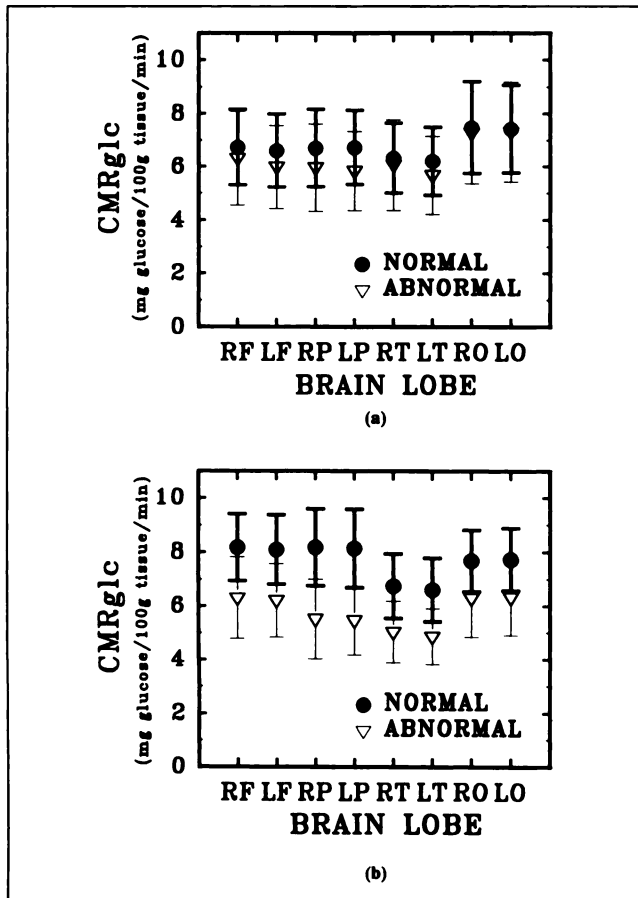
FIGURE 1. Mean regional metabolism in the four bilateral lobes of the brain (right and left frontal, parietal, temporal and occipital) within normal and abnormal groups for (a) PETT V subjects and (b) Scanditronix subjects. A 2 s.d. range is shown (high = mean + s.d., low = mean − s.d.). There was a larger degree of overlap between normal and abnormal groups in the PETT V group than was present between the normal and abnormal groups in the Scanditronix group. These visible differences were confirmed by the classification experiments.
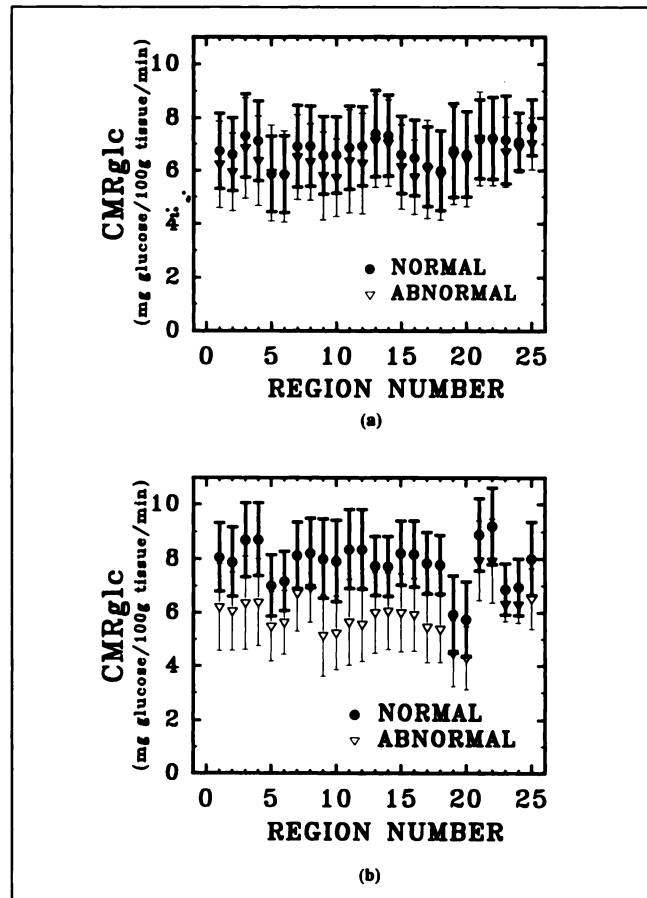
FIGURE 2. Mean regional metabolism, according to "lobule" (see Table 2 for lobule identification) within normal and abnormal groups for (a) PETT V subjects and (b) Scanditronix subjects. As in Figure 1, one can observe the larger degree of overlap between normal and abnormal groups in the PETT V group than that observed between the normal and abnormal groups in the Scanditronix group.

tation, it is evident that there was a much larger degree of overlap between the normal and abnormal groups in the PETT V database than was present between the normal and abnormal groups in the Scanditronix database. These visible differences were confirmed by classification experiments. When subjects were classified strictly according to global metabolism, the ROC area for Scanditronix data was 0.90 in comparison to 0.60 for PETT V data.

Optimization experiments indicated the use of four hidden units at both levels of representation. Thus, the networks used for classification were either 8-4-1 networks (for lobar representations) or 25-4-1 networks (for lobular representations). Figure 3 shows ROC curves for neural networks applied to the PETT V data at lobar and lobular levels for two different methods of preprocessing: simple scaling (which preserved absolute metabolic information) and occipital normalization. As this figure shows, the highest ROC area was attained for a lobular representation

using occipital-normalized data, while simply-scaled lobar data resulted in the lowest performance.

Figure 4 shows equivalent ROC curves for neural networks applied to the Scanditronix data. For this database, maximum ROC area was achieved with lobular data (with approximately equal performances for both types of preprocessing), while occipital-normalized lobar data gave the lowest performance. Neural-network classification results for both databases and for all preprocessing methods are summarized in Table 3. Small variations in these ROC values were highly significant. T-tests indicated that differences in ROC area of 0.02 were significant (p = 0.005).

Figure 5 illustrates ROC curves which compare the performances of neural networks and discriminant analysis for particular datasets. For Scanditronix lobular data, the performances of neural networks and discriminant analysis were approximately equal; for PETT V lobular data, the performance of neural networks was somewhat higher than
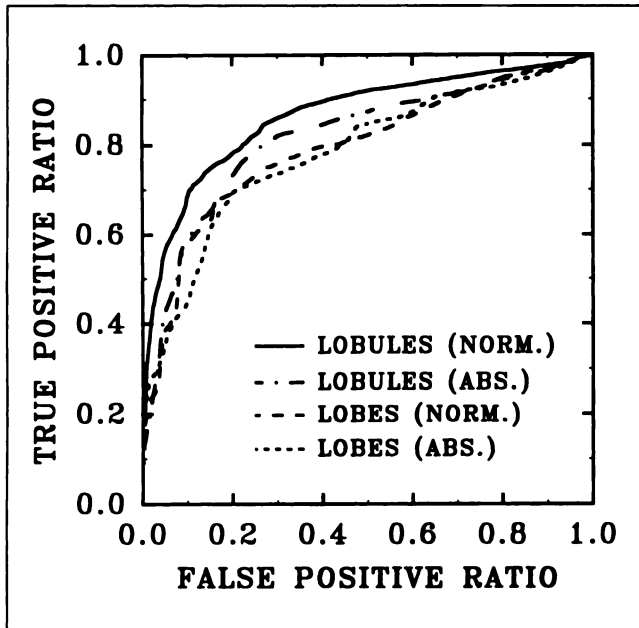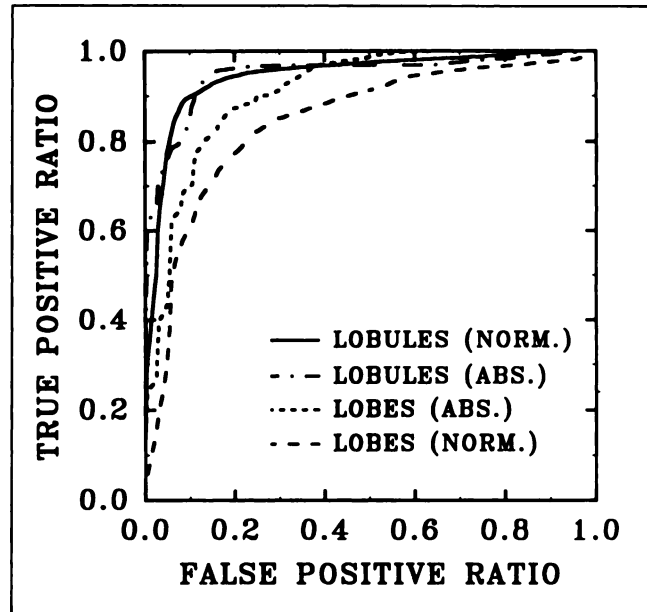
FIGURE 3. ROC curves illustrating classification performance for neural networks applied to the PETT V data at lobar and lobular levels for two different preprocessing methods: simple scaling (which preserved absolute metabolic information) and occipital normalization. The highest ROC area was attained for a lobular representation using occipital-normalized data, whereas simply scaled lobar data resulted in the lowest performance. The curves in Figures 3 and 4 are averages of ROC curves from 80 trials (20 trials for each of four cross-validation configurations).

FIGURE 4. ROC curves illustrating classification performance for neural networks applied to the Scanditronix data at lobar and lobular levels for two different preprocessing methods: simple scaling and occipital normalization. For this database, maximum ROC area was achieved with lobular data (with approximately equal performances for both types of preprocessing), whereas occipital-normalized lobar data gave the lowest performance.

that of discriminant analysis. Classification performances for discriminant analysis are summarized in Table 4.

Table 5 summarizes the results of experiments which investigated the extent to which neural networks could identify groups in one database after being trained with sets that included subjects from the other database. These experiments were conducted with occipital-normalized data. At the lobar level, all between-database, cross-validation tests yielded ROC areas that were within 0.02 units of the corresponding results from the last column of Table 3. At the lobular level, this was true only when testing with Scanditronix subjects after training with a mixture of PETT V and Scanditronix subjects. The between-database performances when testing on lobular PETT V data, in particular, were much lower than the corresponding result (A = 0.87) from Table 3.

Figures 6 and 7 illustrate the most important and generalizable discriminating profiles learned by the neural networks. The profiles in both figures were learned during "between-database" experiments corresponding to results from the second row of Table 5. The two lobar profiles in Figure 6 are similar in that they both have relatively low values for parietal and temporal lobes, particularly for parietal lobes. In particular, the first profile emphasizes low values in the left-parietal region, whereas the second gives approximately equal weights to both parietal lobes. The major features of the two lobular profiles in Figure 7 are: generally low values

in the parietal (regions 9–12) and temporal (regions 15–18) regions combined with consistently higher values in the motor-sensory (regions 7 and 8) and occipital regions (regions 13 and 14). The first profile shows consistent asymmetries (right-side coefficients higher than left-side coefficients) in the frontal, temporal and parietal regions, whereas the second profile shows asymmetries in the opposite direction (left-side coefficients higher than right-side coefficients) in the parietal and temporal regions.

DISCUSSION

The results of this work illustrate that PET can be a powerful tool for discriminating between normal and demented subjects. The ROC area values presented here can be compared with values from the literature (19) which describe the diagnostic performances of various medical imaging techniques, such as the detection of brain lesions on CT (A = 0.97) and on radionuclide scanning (A = 0.87) and the detection of adrenal disease (0.93 for CT, 0.81 for ultrasound). The highest ROC area of the PET-neural-network combination was 0.95, which compares well with these values and supports similar performances shown by Friedland et al. (29) for a similar subject group.

Some general trends were immediately apparent. First, classification performance for Scanditronix data was generally higher than that for PETT V data. As seen from

TABLE 3
Summary of Neural-Network Classification Performances for Various Data Representation Methods Using Two Different PET Databases

| | | | Data preparation method | |
Data/Representation	Simple scaling (Absolute metab.)	Mean removal	Global normalization	Occipital normalization
PETT V/Lobes	0.78	0.78	0.80	0.80
PETT V/Lobules	0.81	0.84	0.87	0.87
Scanditronix/Lobes	0.91	0.86	0.86	0.86
Scanditronix/Lobules	0.95	0.94	0.95	0.95

Table 4, ROC areas for corresponding experiments were from 0.06 to 0.13 higher for Scanditronix data than for PETT V data. The difference between the respective levels of highest performance was 0.08 (A = 0.87 versus A = 0.95). By comparing Tables 3 and 4, it is also evident that the performance of neural-networks was generally higher than that of discriminant analysis, with the two exceptions of mean-removed and occipital-normalized lobule-level Scanditronix data. The ability of the neural-network classifiers to form nonlinear and nonparametric decision boundaries apparently allowed it to be generally more robust for these classification problems.

The strength of the cross-validation method is that it indicates the degree to which a classification system is able to *generalize*, i.e., to apply learned "classification rules" to previously unseen data. The backpropagation neural network, a highly adaptive classifier, places much of the burden for determining generalization ability *on the training data*. The ease by which disease "signals" were detected depended on the manner of representation, as illustrated in Table 3, where the interactive effects of variations in representation level (lobar versus lobular) and preprocessing method for each of the two databases can be seen.

Examination of the first row of Table 3 shows that the classification performances for PETT V lobar data with either global or occipital normalization were slightly higher than performances with either simple scaling or mean removal. This trend also held true for PETT V lobular data, with the difference that performance for mean removal, although lower than performances for normalization, was higher than that for simple scaling. The results with both lobar and lobular data affirm previously published observations (8) that absolute metabolic information for this data did *not* contain important discriminating information. The results also indicate that normalization by division (with the intention of removing multiplicative noise) may have an advantage over mean removal (intended to combat additive noise).

The trends for Scanditronix data had important similarities and important differences compared to those for PETT V data. Performances for Scanditronix lobar data with either mean removal or normalization were *lower* than the performance with simple scaling. These results are intuitively consistent with the plots shown in Figures 1 and 2, which show that the distributions of mean values for

Scanditronix data demonstrate greater separation between classes than is the case for PETT V data, and are also consistent with the separation between classes demonstrated by Kumar et al. (12) for Scanditronix data from a similar subject group. One of the most likely reasons for disparity between databases is the difference in methods for blood collection. "Arterialized" venous blood curves are reportedly less reliable than those from arterial measurements (30–32) and variability in the PETT V blood curves would have translated directly into increased noise in the ROI data. This said, it is interesting to note that the apparent importance of absolute metabolism in the lobar Scanditronix data did not hold true for *lobular* data. For these experiments, the performance for normalized data (containing no information about absolute metabolism) equaled that for data which contained absolute metabolic information.

The results with both PETT V and Scanditronix data-
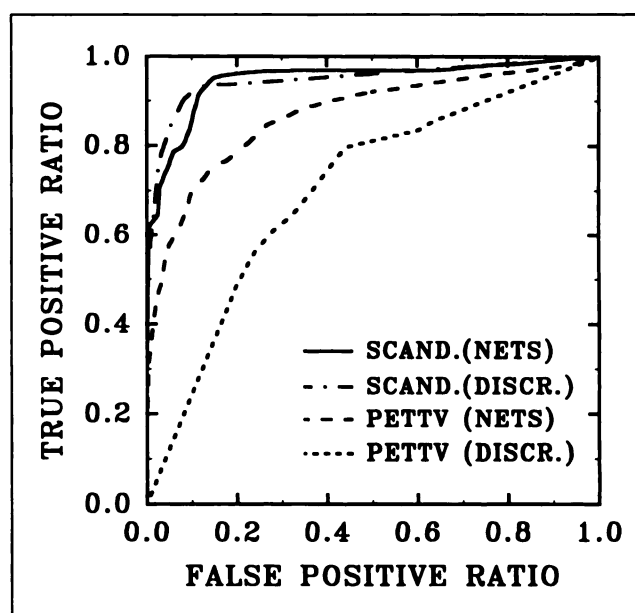
FIGURE 5. ROC curves illustrating and comparing the performances of neural networks and discriminant analysis. For Scanditronix lobular data, the performances of neural networks and discriminant analysis were approximately equal, whereas for the more difficult problem of separating normalized lobular PETT V data, the performance of neural networks was significantly higher than that of discriminant analysis.

TABLE 4

TABLE 4

Summary of Discriminant-Analysis Classification Performances for Various Data Representation Methods Using Two Different PET Databases

Data preparation method data/representation	Simple scaling (Absolute metab.)	Mean removal	Global normalization	Occipital normalization
PETT V/Lobes	0.74	0.75	0.79	0.79
PETT V/Lobules	0.75	0.72	0.74	0.71
Scanditronix/Lobes	0.86	0.83	0.84	0.83
Scanditronix/Lobules	0.84	0.95	0.88	0.95

bases indicate that the lobular representation is probably more appropriate for the detection of dementia. Although classification problems in 25-dimensional space are inherently more difficult than those in 8-dimensional space, the value of this type of representation over the lobar representation was apparently high enough to overcome the increased difficulties associated with classifying patterns in such a higher-dimensional space.

Altogether, the results serve to elucidate the relative contributions of two components of the discriminating information obtained in these PET studies, i.e., the overall absolute metabolism, and the multivariate patterns of relative regional metabolism. The performance using overall metabolism without pattern information (A = 0.90) was higher than that achieved with normalized Scanditronix lobar patterns (A = 0.86), and performance for lobar pattern data that included absolute metabolic information was higher (A = 0.91) than that for normalized data, indicating that absolute metabolism was able to contribute important discriminating information at the *lobar level*. However, the performance using overall metabolism alone was lower than that for normalized Scanditronix lobular patterns (A = 0.95), and performance for lobular data that included absolute metabolic information was not higher (A = 0.95) than that for normalized data, which indicates that absolute metabolism did not contribute any additional discriminating information at the lobular level. The net implication is that reduction of variability was of overriding importance. The reduction of variability in ROI data due to the higher resolution of the Scanditronix camera, combined with that resulting from normalization of the lobular data, apparently
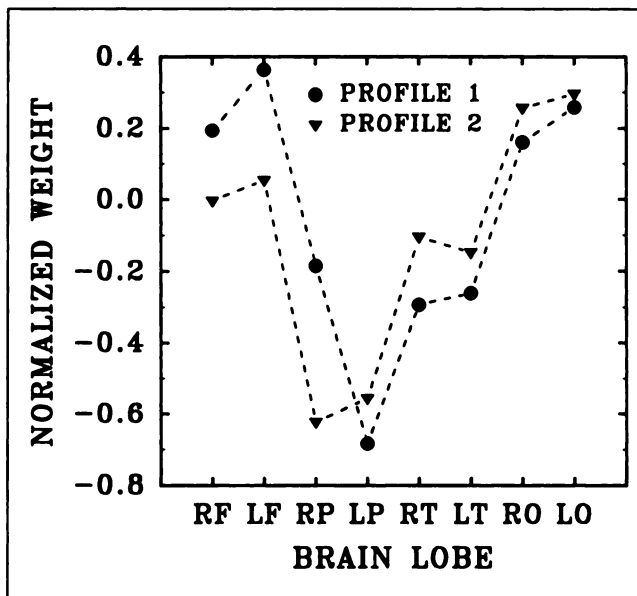
improved the representations to the point that, for a multivariate approach, they contained as much discriminating information as did representations which contained absolute metabolic information.

The finding that absolute quantification may not necessarily be vital for the detection of disease appears to be compatible with the findings of Strother et al. (*6*), who reached a similar conclusion after examining measures of group discrimination for patients with acquired immune deficiency syndrome dementia complex. These results imply that, with regard to the detection of memory disorders, efforts to improve image resolution may be at least as deserving of attention as those focusing on absolute quantification of metabolism. Such considerations may have

FIGURE 6. The most important and generalizable discriminating profiles learned by neural networks during lobe-level training with subjects from both PET databases. Coefficients are plotted for right and left frontal, parietal, temporal and occipital lobes. The two profiles are similar in that they both have relatively low values for parietal and temporal lobes, particularly for parietal lobes. The first profile more strongly emphasizes low values in the left-parietal region, while the second gives approximately equal weights to both parietal lobes. Although these representations admittedly oversimplify the role of these profiles in the neural network classification process, they do provide an intuitive indication of the basis on which classification "decisions" were made, which show that neural networks incorporated and combined some features "typical" of Alzheimer's disease.
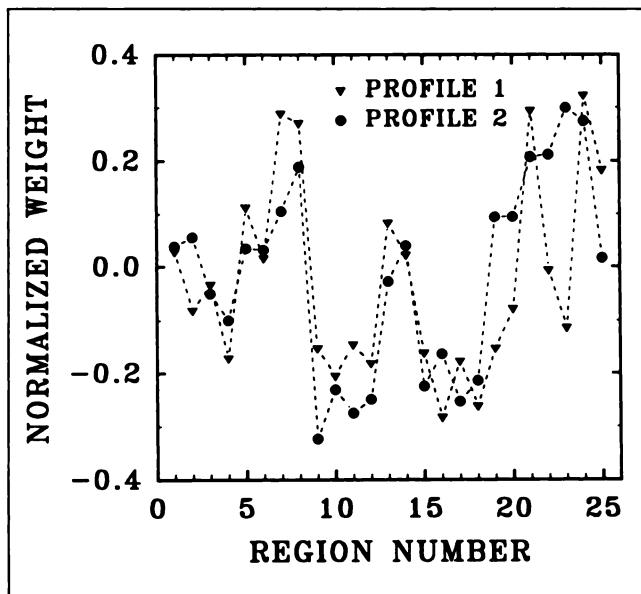
TABLE 5

Summary of Neural-Network Classification Performances for Between-Database Cross-Validation

Testing set	Representation Training set	
	Lobe	Lobule
PETT V/Scand.	0.85	0.90
PETT V + Scand./Scand.	0.84	0.94
Scand./PETT V	0.80	0.70
PETT V + Scand./PETT V	0.81	0.76

All tests were conducted with occipital-normalized data.

FIGURE 7. The most important and generalizable discriminating profiles learned by neural networks during lobule-level training with subjects from both PET databases. The major features of these profiles are: generally low values in the parietal (regions 9–12) and temporal (regions 15–18) regions combined with consistently higher values in the motor-sensory (regions 7 and 8) and occipital regions (regions 13 and 14). The first profile shows consistent asymmetries (right-side coefficients higher than left-side coefficients) in the frontal, temporal and parietal regions, whereas the second profile shows asymmetries in the opposite direction (left-side coefficients higher than right-side coefficients) in the parietal and temporal regions. The trained networks thus identified either type of asymmetry as an indicator of abnormality.

implications for the application of quantitative classification techniques to both current and future metabolic imaging modalities.

The results highlighted in Figure 5 appear to confirm that the ability of neural networks to employ multiple discriminating surfaces allowed it to outperform discriminant analysis for the difficult problem of separating groups for normalized lobular PETT V data. The separation of groups for Scanditronix data was apparently accomplished with a single discriminating surface, allowing discriminant analysis to perform at the level of neural networks.

It is apparent from the results of Table 5 that it is possible to successfully combine normalized data from different databases. The major exception to this was training with Scanditronix data and testing with PETT V data at the lobule level. The PETT V lobular data apparently served as a fairly good training set for testing on Scanditronix lobular data, but the reverse was not true. Presumably, this occurred because the PETT V data was the noisier of the two sets of data; training with this data would have yielded an "expanded" decision region compared to the decision region resulting from training with the Scanditronix data. Better performance would be expected by training with noisier data and testing on the less noisy data than by performing the reverse experiment. It should be noted that

any recommendations based on these experiments apply only to normalized data from the two databases and not to data based on absolute metabolism.

The classification methods shown here can be used to influence a diagnostic decision based on Bayesian theory. For example, classification at the point of maximum information on the ROC curve for normalized Scanditronix lobular data resulted in a sensitivity of 90% with a corresponding specificity of 89%. Given a patient with mild cognitive impairment, where the diagnosis of AD was uncertain, corresponding to a prior probability of 50%, then a positive classification at this point on the ROC curve would yield a post-test probability of disease of 90% (strongly confirming the presence of abnormality), while a negative classification would give a post-test probability of 10% (strongly ruling out abnormality). The corresponding post-test probability values for normalized PETT V lobular data were 87% (for a positive classification) and 24% (for a negative classification).

Although the representations of discriminating profiles in Figures 6 and 7 admittedly oversimplify the role of these profiles in the neural network classification process, they do serve to provide an intuitive indication of the basis on which classification "decisions" were made. They show that the neural networks incorporated and combined features thought to be "typical" of Alzheimer's disease. Interested readers may contact the primary author to obtain complete quantitative descriptions of networks trained with subjects from both databases.

These profiles represent discriminating "surfaces" which roughly define directions (in n-dimensional pattern space) of increasing severity of dementia. This interpretation of the profiles is consistent with results shown by Kumar et al. (12) for normalized data. They found mainly parietal deficiencies in "mild" dementia; parietal, temporal and premotor deficiencies in "moderate" cases; and widespread involvement in severe cases. In addition, the profiles emphasizing "left-side-low" asymmetry are consistent with profiles from our previous work (8) and with published reports of predominantly left-side deficits (10) for similar subject groups. The fact that the two profiles in Figure 7 incorporated asymmetries of opposite directions in the parietal and temporal regions indicates that the trained network identified either type of asymmetry as an indication of abnormality. It should be noted that typical analyses comparing normal to abnormal groups based on mean trends would have great difficulty recognizing such a situation, since the presence of both types of asymmetry in one group would cancel each other out on a mean basis.

In summary, analyses of these results indicate that the combination of PET and neural networks performed well in discriminating normal from AD subjects. The increased spatial resolution of the Scanditronix camera apparently allowed a net increase in signal-to-noise ratio of the lobule-level ROI data, which resulted in improved classification performance. Absolute quantification of metabolism was shown to contain discriminating information, but classifi-

cation performance for lobule-level data was not improved by including this information. These results indicate that the maturation of PET technology over the last few years has improved its clinical research value in the diagnosis of dementia. The results also indicate that PET data obtained with "first generation" cameras may have value as training data, despite (or perhaps even because of) the fact that they may contain greater variability. The above findings could not have been confirmed intuitively. They provide directions for future investigations employing PET in the study of dementia. These results show that it should be possible to share metabolic data from different scanners and institutions to develop an extensive "knowledge" base of metabolic patterns. This would greatly enhance the utility of metabolic neuroimaging by using it to predict disease probabilities based on a metabolic profile obtained in a given clinical situation.

ACKNOWLEDGMENTS

REFERENCES

1. Yoshii F, Barker WW, Chang JY, et al. Sensitivity of cerebral gluclose metabolism to age, gender, brain volume, brain atrophy, and cerebrovascular risk factors. *J Cereb Blood Flow Metab* 1988;8:654–661.
2. Horwitz B. Functional interactions in the brain: use of correlations between regional metabolic rates. *J Cereb Blood Flow Metab* 1991;11:A114–A120.
3. Haxby JV. Resting state regional cerebral metabolism in dementia of the Alzheimer type. In: *Positron emission tomography in dementia*. New York: Wiley-Liss; 1990:93–116.
4. 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.
5. Ford I, McColl H, McCormack AG, McCrory SJ. Statistical issues in the analysis of neuroimages. *J Cereb Blood Flow Metab* 1991;11:A89–A95.
6. Strother SC, Liow JS, Moeller JR, Sidtis JJ, Dhawan VJ, Rottenberg DA. Absolute quantitation in neurological PET: do we need it? *J Cereb Blood Flow Metab* 1991;11:A3–A16.
7. Kippenhan JS. Neural-network classification and analysis of positron emission tomography images of subjects with memory disorders. PhD Dissertation, University of Miami, 1991.
8. 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.
9. Ter Pogossian M, Mullani N, Hood J. Design considerations for a positron emission transverse tomograph (PETT V) for imaging of the brain. *J Comput Assist Tomogr* 1978;2:149–154.
10. Loewenstein DA, Barker WW, Chang J, et al. Predominant left hemisphere metabolic dysfunction in dementia. *Arch Neurol* 1989;46:146–152.
11. Daube-Witherspoon ME, Green MV, Holte S. Performance of Scanditronix PC1024-7B PET scanner. *J Nucl Med* 1987;28:607–608.
12. Kumar AK, Schapiro MB, Grady C, et al. High-resolution PET studies in Alzheimer's disease. *Neuropsychopharmacol* 1991;4:35–46.
13. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 1984;34:939–944.
14. Phelps M, Huang S, Hoffmann E, Selin E, Sokoloff L, Kuhl D. Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18) 2-fluoro-2-deoxy-D-glucose: validation of method. *Ann Neurol* 1979;6:371–388.
15. Hutchins GD, Holden JE, Koeppe RA, Halama JR, Gatley SJ, Nickles RJ. Alternative approach to single-scan estimation of cerebral glucose metabolic rate using glucose analogue, with particular application to ischemia. *J Cereb Blood Flow Metab* 1984;4:35–40.
16. Duara R, Margolin R, Robertson-Tchabo E, et al. Cerebral glucose utilization, as measured with positron emission tomography in 21 resting healthy men between the ages of 21 and 83 years. *Brain* 1983;106:761–775.
17. Brooks RA, Hatazawa J, Di Chiro G, Larson SM, Fishbein DS. Human cerebral glucose metabolism determined by positron emission tomography: a revisit. *J Cereb Blood Flow Metab* 1987;7:427–432.
18. Huang SC, Phelps ME, Hoffman EJ, Sideris K, Selin CJ, Kuhl DE. Noninvasive determination of local cerebral metabolic rate for glucose in man. *Am J Physiol* 1980;238:E69–E82.
19. Swets J. Measuring the accuracy of diagnostic systems. *Science* 1988;240: 1285–1293.
20. Metz CE. ROC Methodology in radiologic imaging. *Invest Radiol* 1986; 21(9):720–733.
21. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, PDP Research Group. *Parallel distributed processing, volume 1*. Boston: MIT Press; 1986: 318–364.
22. Pao Y-H. *Adaptive pattern recognition and neural networks*. Reading, MA: Addison-Wesley; 1989.
23. Folstein MF, Folstein SE and McHugh PR. 'Mini-mental state': a practical method for grading the mental state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198.
24. Kippenhan JS, Nagel JH. Optimization and evaluation of a neural-network classifier for PET scans of memory-disorder subjects. *Proc 13th Ann Int Conf IEEE/EMBS* 1991;13:1472–1473.
25. *SAS/STAT guide for personal computers, version 6 edition*. Cary, NC. SAS Institute Inc.; 1985:83–110.
26. Morrison DF. *Multivariate statistical methods*. New York: Mc-Graw Hill; 1976.
27. Metz CE, Goodenough DJ and Rossmann K. Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiography. *Radiology* 1973;109:297–303.
28. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–215.
29. Friedland RP, Horwitz B, Grady C, et al. An ROC analysis of the diagnostic accuracy of PET with FDG and x-ray computed tomography in Alzheimer's disease. *J Cereb Blood Flow Metab* 1989;9(suppl 1):S566.
30. Foster NL, Chase TN, Fedio P, Patronas NJ, Brooks RA, DiChiro G. Alzheimer's disease: focal cortical changes shown by positron emission tomography. *Neurology* 1983;33:961–965.
31. Benson DE, Kuhl DE, Hawkins RA, Phelps ME, Cummings JL, Tsai SY. The fluorodeoxyglucose ^{18}F scan in Alzheimer's disease and multi-infarct dementia. *Arch Neurol* 1983;40:711–714.
32. Friedland RP, Budinger TF, Ganz E, et al. Regional cerebral metabolic alterations in dementia of the Alzheimer type: positron emission tomography with [^{18}]fluorodeoxyglucose. *J Comp Assist Tomogr* 1983;7:590–598.