EDITORIAL

Detection and Comparison of Patterns in Images

Fundamentally, all medical imaging involves comparisons. When used for diagnosis, images are frequently compared with mental pictures (the interpreter's concepts) of normality and different disease states. For prognosis or treatment monitoring, comparisons often focus on changes in serial studies in the same patient. These comparisons are based on the patterns seen in the images. In nuclear medicine, these patterns represent spatial and temporal arrangements and rearrangements of the physiological or biochemical processes under investigation. How are these patterns best detected and compared?

In this issue of the *Journal*, Kippenhan and coworkers present an extension (*1*) of their previously reported approach (*2*) to classification of normal and Alzheimer's disease subjects based on neural-network analysis of FDG-PET studies. Their approach utilizes the "pattern" formed by organizing region of interest (ROI) values from selected brain areas into (mathematical) vectors (or "profiles"), which provide the input for a neural-network based classifier. With this approach, the investigators found that the network could change a pre-FDG-PET probability of disease of 50% (i.e., that of a subject with the greatest pre-FDG-PET uncertainty) to a post-test probability of either 90% for a positive classification or 10% for a negative classification.

The output of the neural network (i.e., classification of the pattern as either "normal" or "Alzheimer's disease") can be considered equivalent to the interpretation of a diagnostic test as either "normal" or "abnormal." In this fashion, one could apply the concepts of "sensitivity" and "specificity" so often used to express diagnostic performance. It is noteworthy that the investigators chose to use receiver- (or relative-) operating-characteristic (ROC) analysis to study the performance of their classifier. This approach (*3*), which characterizes the accuracy of a test over the full operating range of normal/abnormal decision (or classification) thresholds, has finally achieved the widespread use it deserves in medical imaging. This approach does not require a priori selection of a single decision threshold to use with a new test, and facilitates a posteriori selection of the optimum threshold prior to routine clinical use.

In those clinical situations in which the relevant distinction is between the normal and abnormal, ROC analysis provides a rigorous assessment of a test's diagnostic accuracy. However, with the complex differential diagnoses so frequently encountered with neurologic and psychiatric disorders, the distinction (or classification) that frequently must be made is not between normal and abnormal, but between multiple diagnoses (e.g., Alzheimer's disease versus multi-infarct dementia in a patient with an abrupt change in clinical state). Furthermore, the distinction between normal and abnormal must often be made in the face of subtle disease (e.g., normal aging versus early dementia in a patient with short-term memory loss). In such settings, normal and abnormal (and sensitivity and specificity) take on less straightforward meanings. In this regard, one of the limitations of the current study (*1*) is that it does not provide data indicating the classifier's ability to provide these more difficult, yet more clinically useful, distinctions.

The spatial resolution of PET tomographs continues to improve. While it is easy to show the relationship between resolution and quantitative accuracy (*4*), and therefore the effect of resolution on measured FDG-PET metabolic rate (*5*), increases in diagnostic accuracy with improved resolution are harder to document. In this regard, the current work offers tantalizing evidence that such increases occur in that the classification performance for the better resolution tomograph tended to be better. It should be noted that the use of an objective, quantitative classification scheme explicitly links quantitative and diagnostic accuracy.

The patterns seen in FDG-PET studies arise from a combination of biologic and technical factors; these also lead to the variations seen within a population at a given time and in a given subject over time (*6,7*). Many investigators favor normalizing data by a reference value (e.g., global metabolic rate or that in a reference brain area, such as cerebellum or occipital cortex) to better uncover regional differences (*8*). In this regard, it is interesting to note that, in the current study, occipital-normalized patterns produced the highest diagnostic accuracy (as measured by the area under the ROC curve) for the poorer resolution, but not the better resolution, tomograph, and that normalization improved diagnostic accuracy for the smaller ("lobular"), but not the larger ("lobar"), regions with the better resolution tomograph (*1*). A recent multicenter report suggests that a standardized FDG-PET protocol utilizing ROIs representing larger brain areas and ratios rather than absolute metabolic rates can provide comparable data in spite of differences in tomograph resolution (*9*). Thus, whether or not normalization helps or hurts seems to be determined by a number of factors, including ROI size, tomograph resolution, variation (both biologic and technical) in the data and the type of normalization. In any event, the use of vectors or profiles such as those formed in the current work, to express patterns facilitates both normalization and statistical analyses, as

Received Sept. 24, 1993; accepted Sept. 30, 1993.

For reprints contact: Jonathan M. Links, Radiation Health Sciences, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205-2179.

demonstrated by others (*10*). The term profile when used in this context does not necessarily imply a sequence of ROI values from anatomically contiguous brain areas.

The use of computer-assisted diagnosis is, of course, not new. Common approaches include statistical (e.g., cluster analysis, discriminate analysis) and rule-based designs, such as those based on "artificial intelligence." All approaches are heavily influenced by the data used to initially design the analysis algorithm. In this regard, a very important aspect of the current study was the use of explicitly separate training and test sets, because it demonstrated that the classifier was applicable to a broader population than that with which it was initially derived. It is also clear that the general approach Kippenhan and coworkers have described has much broader applicability than just FDG-PET. The extension to SPECT and planar imaging, and to other diseases, is easy to envision. Such an extension might provide detection of more subtle patterns than the usually visually apparent pattern of probable Alzheimer's disease. In any event, these approaches are clearly consistent with the quantitative, pattern-based character of nuclear medicine, and their continued development should be strongly supported.

Jonathan M. Links
The Johns Hopkins University
Baltimore, Maryland

Michael D. Devous, Sr.
UT Southwestern Medical Center
Dallas, Texas

REFERENCES

1. Kippenhan JS, Barker WW, Nagel J, Grady C, Duara R. Neural-network classification of normal and Alzheimer's disease subjects using high and low resolution PET cameras. *J Nucl Med*, 1993;35:7–15.
2. 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.
3. Metz CE, Goodenough DJ, Rossmann K. Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiology. *Radiology* 1973;109:297–303.
4. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr* 1979;3:299–308.
5. Grady CL, Berg G, Carson RE, Daube-Witherspoon ME, Friedland RP, Rapoport SI. Quantitative comparison of cerebral glucose metabolic rates from two positron emission tomographs. *J Nucl Med* 1989;30:1386–1392.
6. Tyler JL, Strother SC, Zatorre RJ, et al. Stability of regional cerebral glucose metabolism in the normal brain measured by positron emission tomography. *J Nucl Med* 1988;29:631–642.
7. Szabo Z, Camargo EE, Sostre S, et al. Factor analysis of regional cerebral glucose metabolic rates in healthy men. *Eur J Nucl Med* 1992;19:469–475.
8. Friston KJ, Frith CD, Liddle PF, et al. The relationship between global and local changes in PET scans. *J Cereb Blood Flow Metab* 1990;10:458–466.
9. Herholz K, Perani D, Salmon E, et al. Comparability of FDG PET studies in probable Alzheimer's disease. *J Nucl Med* 1993;34:1460–1466.
10. Moeller JR, Strother SC, Sidtis JJ, Rottenberg DA. Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. *J Cereb Blood Flow Metab* 1987;7:649–658.