EDITORIAL
Artificial Neural Networks: Better Than the Real Thing?

Artificial neural networks (ANNs), a relatively new approach to the automated interpretation of medical images (1,2), presents us with the problem of comparing this new technique with accepted forms of image interpretation. Comparing the usefulness and validity of one imaging technique to another is difficult, and these problems are well illustrated when Porenta et al. (3) use ANN to interpret planar ^{201}Tl-dipyridamole stress-redistribution scintigrams. They used several techniques to compare ANN interpretations with human experts (HE) and with coronary angiography (CAG), which included receiver operational characteristic (ROC) curve analysis, concordance rates from a 3 × 3 classification matrix and a comparison of positive and negative predictive values.

ANNs are computer programs which simulate a functional and structural model of the brain (4–6). Various parameters of medical images are defined, and this information is provided to the computer, along with established diagnoses. After applying a training set of data, ANNs can provide image interpretation. ANNs also have the ability to learn from new images with known diagnoses. ANNs have been used for a variety of pattern recognition and decision-making problems. In nuclear medicine, these include: the automated interpretation of cerebral perfusion imaging (7), ventilation-perfusion lung scans (8,9) and myocardial perfusion stress studies (10).

Image interpretation is a complex task that involves not only the simple task of detecting an abnormality in an image, but also other qualities of these abnormalities such as the number of abnormalities present, as well as the location, size, shape and texture of these abnormalities. It is these param-

eters of image abnormalities that are used in HE interpretation that can be included in a method such as ANNs. This would make ANNs a very powerful technique to the HE aid in the interpretation of medical images or to do the interpretation without human input.

As a first approach, Porenta et al. (3) used ROC curves to compare the three procedures. They took the unusual step of using two "gold standards," that is, the establishment of the patient's "true" diagnosis, by first considering the CAG to define the patient's medical condition, and then using the HE as the basis for "truth." ROC analysis has been shown to provide useful comparisons when evaluating the ability of an observer to identify the presence or absence of a single, well defined signal (11–17). ROC curves were applied to the analysis of different types of radiographic film for diagnosing tuberculosis almost 50 yr ago (18). This technique did not catch on until the 1970s and then became a very popular way to compare medical imaging systems in the 1980s (19).

Unfortunately, ROC analysis does not take into consideration all of the parameters of a medical image. A simple example of this problem is the situation where an observer incorrectly identifies an abnormality in one part of the image and misses an abnormality in another part of the image. Is this interpretation a false-negative or a false-positive? Another example is when an observer correctly identifies an abnormality in one area of the image but misses an abnormality in another part of the image. Is this reading a true-positive or a false-negative interpretation?

And these are the easy problems! The problem with multiple signals and their location has been considered and solutions using LROC analysis (20) and, more recently, FROC analysis (21–25) have been applied. However,

this still excludes what has been called the classification or recognition problem (26).

Chesters (11) states that: "surprisingly few attempts have been made to quantify the recognition of signals. ... Models of recognition are few and incomplete in comparison with the rather well developed models of detection based on signal detection theory. As a result, the comparison of the performance of imaging systems is usually made on the basis of detectability of signals and an important question is whether the ranking of performance on the basis of recognizability is the same as on the basis of detectability. It is usually agreed that detection is a necessary condition for recognition, but certain experimental evidence suggests that it may not be sufficient." In medical imaging, recognizing and classifying the signal can be more important than detecting the presence or absence of a signal.

Suppose an observer identifies an abnormality in the correct location, but fails to give the correct impression of the size, shape or texture of the abnormality. In the binary task of ROC analysis to decide if a signal is present or not present, this interpretation would be considered true-positive. ROC analysis would be unable to indicate how much of the information of the "fine structure" of the abnormality in the image is being transmitted to the reader. This "fine structure" information can provide clues to discriminate between a benign and a malignant lesion.

Porenta et al. (3) use ROC curves to test the ability of the techniques to detect coronary artery disease. They then extend their comparison of HE and ANN by investigating the ability of the planar thallium images to provide information on two other aspects of the images: the severity and the localization of the disease. They used concordance rates from a 3 × 3 classification matrix to classify the sever-

ity of the disease, and positive and negative predictive values to localize the areas involved. While the approach that the authors took to study severity and localization do not appear to be independent tasks, these analyses do demonstrate that different information can be generated on the usefulness of one technique over the other.

Porenta et al. (3) suggest that "the information content of the 45 integer numbers resulting from segmental analysis (used in their ANN analysis) is considerably less than the information that is present in the scintigraphic images which were used by the human expert to derive his diagnostic classifications. Thus, ANN that use the entire images as input may possibly achieve a better diagnostic performance."

ANN is a valuable technique that could, but does not necessarily have to, replace physician readers. Certainly, ANN could add valuable insights into how human experts interpret images and what parameters are important to the human expert. Then, once these parameters are defined, imaging equipment and computer software can be developed to enhance these parameters.

Are artificial neural networks better than the real thing (28)? We don't know, but we believe that, theoretically, they could be. We definitely feel that investigations should continue into the use of artificial neural networks to aid human experts in the interpretations of medical images, as well as to demonstrate how human experts interpret medical images (29, 30). We need better analytic techniques to evaluate an observer's ability to not only detect a signal in a medical image, but also to recognize and classify that signal.

Alan B. Ashare
St. Elizabeth's Medical Center of Boston
Boston, Massachusetts

Dev P. Chakraborty
Hospital of the University
of Pennsylvania
Philadelphia, Pennsylvania

REFERENCES

1. Boone JM, Sigillito AG, Shaber GS. Neural networks in radiology: an introduction and evaluation in a signal detection task. Med Phys 1990;17:234–241.
2. Boone JM, Gross GM, Greco-Hunt V. Neural networks in radiologic diagnosis: I. Introduction and illustration. Invest Radiol 1990;25:1012–1016.
3. Porenta G, Dorffner G, Kundrat S, Petta P, Duit J, Sochor. Automated interpretation of planar thallium-201 dipyridamole stress-redistribution scintigrams using artificial neural networks. J Nucl Med 1994;35:000–000.
4. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 1982;79:2554–2558.
5. Floyd CE Jr, Tourassi GD. An artificial neural network for lesion detection on single-photon emission computed tomographic images. Invest Radiol 1992;27:667–672.
6. Tourassi GD, Floyd CE Jr. Artificial neural networks for single photon emission computed tomography: a study of cold lesion detection and localization. Invest Radiol 1993;28:617–677.
7. Chan KH, Johnson KA, Becker JA, et al. A neural network classifier for cerebral perfusion imaging. J Nucl Med 1994;35:771–774.
8. Tourassi GD, Floyd CE Jr, Sostman HD, Coleman RE. Acute pulmonary embolism: artificial neural network approach for diagnosis. Radiology 1993;189:555–558.
9. Scott JA, Palmer EL. Neural network analysis of ventilation-perfusion lung scans. Radiology 1993;186:661–664.
10. Fujita H, Katafuchi T, Uehara T, Nishimura T. Application of artificial neural network to computer-aided diagnosis of coronary artery disease in myocardial SPECT bull's-eye images. J Nucl Med 1992;33:272–276.
11. Chesters MS. Human visual perception and ROC methodology in medical imaging. Phys Med Biol 1992;37:1433–1476.
12. Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. CRC Critical Reviews Diag Imag 1989;29:307–335.
13. Swets JA, Pickett RM. Evaluation of diagnostic systems; methods from signal detection theory. New York: Academic Press; 1982.
14. Swets JA, Pickett RM, Whitehead SF, et al. Assessment of diagnostic technologies. Science 1979;205:753–759.
15. Green DA, Swets JA. Signal detection theory and psychophysics. New York: Wiley; 1966.
16. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986;21:720–733.
17. Metz CE. Some practical issues of experimental design and data analysis in radiologic ROC studies. Invest Radiol 1989;24:234–245.
18. Yerushalmy J. Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray technics. Public Health Reports 1947;62:1432–1449.
19. Doubilet PM. Statistical techniques for medical decision making: applications to diagnostic radiology. Amer J Roent 1988;150:745–750.
20. Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. Radiology 1975;116:533–538.
21. Egan JP, Greenberg GZ, Schulman AL. Operating characteristics, signal detectability, and the method of free response. J Account Soc Amer 1961;33:993–1007.
22. Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free-response approach to the measurement and characterization of radiographic-observer performance. J Appl Photogr Eng 1978;4:166–171.
23. Chakraborty DP. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med Phys 1989;16:561–568.
24. Chakraborty DP, Winter LH. Free-response methodology: alternative analysis and a new observer-performance experiment. Radiology 1990;174:873–881.
25. Herman C, Buhr E, Hoeschen D, Fan SY. Comparison of ROC and AFC methods in a visual detection task. Med Phys 1993;20:805–811.
26. Patton DD. The abnormal brain scan: specificity of descriptive parameters. CRC Crit Rev in Clin Radiol Nucl Med 1976;7:339–425.
27. Phillips WC Jr, Scott JA. Medical decision making: practical points for practicing radiologists. Amer J Roent 1990;154:1149–1155.
28. Boone JM. Neural networks at the crossroads. Radiology 1993;189:357–359.
29. Worthington BS. Human observer performance; a new set of problems to solve. Invest Radiol 1993;28:S160–S162.
30. Kundel HL. Images, image quality and observer performance. Radiology 1979;132:265–271.