Using an Artificial Neural Network to Diagnose Hepatic Masses

Philip S. Maclin and Jack Dempsey

Using abdominal ultrasonographic data and laboratory tests, radiologists often find differential diagnoses of hepatic masses difficult. A computerized second opinion would be especially helpful for clinicians in diagnosing liver cancer because of the difficulty of such diagnoses. A backpropagation neural network was designed to diagnose five classifications of hepatic masses: hepatoma, metastic carcinoma, abscess, cavernous hemangioma, and cirrhosis. The network input consisted of 35 numbers per patient case that represented ultrasonographic data and laboratory tests. The network architecture had 35 elements in the input layer, two hidden layers of 35 elements each, and 5 elements in the output layer. After being trained to a learning tolerance of 1%, the network classified hepatic masses correctly in 48 of 64 cases. An accuracy of 75% is higher than the 50% scored by the average radiology resident in training but lower than the 90% scored by the typical board-certified radiologist. When sufficiently sophisticated, a neural network may significantly improve the analysis of hepatic-mass radiographs.

INTRODUCTION

Diagnoses of liver cancers are based primarily on abdominal ultrasonography, a safe, routine, simple procedure that is continually undergoing technological improvement.¹ Our research deals with the specific problem of how to improve the diagnoses of hepatic masses using artificial neural networks trained on ultrasound data and laboratory test results.

Artificial neural networks are solving problems that previous technologies have been unable to resolve satisfactorily, especially in the diagnosis of disease and other applications involving pattern recognition.² An artificial neural network is an informationprocessing technique or a computer-based simulation of a living nervous system, with characteristics that come from a structure of many interconnected elements operating in parallel.

With increased curative surgical techniques for primary and secondary hepatic neo-

From the University of Tennessee, Memphis, Department of Health Informatics, Educational Computing 8 S. Dunlap St., Memphis, Tennessee 38163; and Veteran's Administration Medical Center, Department of Radiology, 1030 Jefferson Ave., Memphis, Tennessee 38104.

plasms, clinical imaging tasks have become more exacting.^{3,4} Neural networks can help solve that problem in radiology. Among the most significant therapeutic advances in the management of primary and metastatic hepatic neoplasms is the continued evolution of aggressive surgical techniques.^{5,6} Tumor detection, differential diagnosis of individual nodules, and mapping of the anatomic extensions of malignant disease are now routinely required. Related and unrelated liver substrate abnormalities such as cavernous hemangioma and focal fatty deposits are often discovered and must be differentiated from metastatic deposits. Also, modern imaging methods such as magnetic resonance imaging (MRI) frequently display tiny nodules (<1 cm) that often prove difficult to adequately characterize (micrometastases versus other types of lesions).⁷

Although no results have been announced on the specific use of a neural network in the diagnosis of hepatic masses, they have been published on related projects. Although no application was attempted, Boone, Gross and Greco-Hunt discussed the possibility of using neural networks in radiologic diagnosis.⁸ They concluded that "developments in this area may ultimately affect radiology."

Gross, Boone, Greco-Hunt, and Greenberg trained a neural network to choose 1 or more from 12 possible diagnoses based on 21 observations of neonatal chest radiographs.⁹ They concluded that a trained neural network has the potential to diagnose neonatal cardiopulmonary disorders "with a consistency approximately equivalent to that of pediatric radiologists."

A prototype neural network was used to diagnose renal cancer in 52 kidney cases.¹⁰ While the approach to the renal cancer diagnosis was similar to our hepatic mass research, diagnosing renal masses is considered less demanding than diagnosing hepatic masses. Furthermore, no laboratory test data was used in the renal research.

Mulsant¹¹ developed a backpropagation neural network that became increasingly proficient at clinical diagnosis of dementia, or irreversible deterioration of intellectual faculties from organic brain disorder. Mulsant concluded that neural computing can help diagnose difficult problems with its ability to learn from experience and to generalize.

Other successful neural networks include emulation of biological system responses,^{12,13} forms of memory and REM sleep,¹⁴ nonlinear Bayesian estimation from sparse data,¹⁵ speech recognition,^{16,17} hypertension,¹⁸ and pathologic diagnoses based on 110 radiographs of bone tumors.¹⁹ In chemistry, various networks predict adverse drug reactions, physicochemical properties characteristic of a safe drug, product distributions from chemical reactions, and three-dimensional structures of proteins from amino acid sequences.²⁰

In summary, biologically inspired neural networks learn from experience, generalize from previous examples to new ones, and abstract essential characteristics from inputs containing irrelevant data. Once trained, a network's response can be partially insensitive to minor variations in input. This ability to see through noise and distortion to the pattern that lies within is vital to pattern recognition in a real-world environment. The network generalizes automatically as a result of its structure.

MATERIALS AND METHODS

Most artificial neural networks are developed in software since conventional computers can implement most ANNs using an appropriate programming language.²¹ After

An Artificial Neural Network

success is proven, some networks are converted to hardware implementations (neurocomputers) to gain processing speed.² The majority of applications are either with neural network shells²² or the C programming language for portability to various computers.²³ The most successful algorithm in solving clinical diagnosis programs so far has been backpropagation.^{24,25}

BrainMaker Professional version 2.02, a software shell designed for neural network applications,²⁶ was used to diagnose hepatic masses in our research. BrainMaker uses standard backpropagation,^{27,28} which is the most widespread learning algorithm for multilayered, straightforward connection. In considering what paradigm would be best for our hepatic mass application, standard backpropagation was chosen because of its successes in clinical diagnoses. Attributes of the hepatic mass application closely matched capabilities of the backpropagation paradigm.²¹ Backpropagation has been applied to a wide variety of research applications, including medical diagnostics.²⁹

Interpretation by a radiologist involves several steps. First, the ultrasound images from the patient's file are viewed and a list of any abnormalities present (referred to as the radiographic findings) is compiled. Next, numeric data of laboratory tests from computer printouts are read. Then the radiologist uses a cognitive process in which conclusions are sought as to the possible diagnosis or diagnoses based on the radiographic findings.

Our research was based on unpublished raw data from hospital patients. At least two radiologists interpreted the characteristics found in the ultrasound images and agreed on their interpretations. The patient usually had only one ultrasound exam, but often had the same type of laboratory test repeated several times over a period of days. If a laboratory test was repeated several times, the radiologists relied more on the test with the date closest to that of the ultrasound reading. Consequently, our research used only the laboratory test with the closest date.

The network was trained 16 different times on 60 cases, 93.75% of the data, and tested on the remaining 4 cases, which were selected at random. Since the number of cases in the data set was limited to 64, the "leave out k" approach was used.³⁰ The "leave out k" method is a process that holds out a different group of facts each time the training is done and then tests with the holdout group. To verify that the network was properly tested on all facts, this procedure was replicated 16 times, and each time a different set of 4 was withheld for test purposes. This procedure resulted in 64 test cases, a number significant enough to find the robustness of generalization, or a global minimum. This approach was used because of the limited number in the data set. The limited data set was the result of having to use only patient histories in which autopsies or surgery had confirmed the diagnosis.

The neural network had 35 elements in the input layer. Some data attributes were represented as binary to the input processing elements. For example, ultrasound is reflected or transmitted from the liver mass so that the area appears white (hyperechoic), gray (hypoechoic), or very black (anechoic); if the mass appeared white, a binary 1 was submitted to the hyperechoic node and a binary 0 to the hypoechoic and anechoic nodes. Each laboratory test was represented as a numeric value to an input element and then normalized. If the laboratory test was not taken or unavailable, a binary 1 was submitted to an input element representing that test as missing. One problem was that laboratory data were often incomplete since few patients were given every possible laboratory test.

The output layer contained the five possible predominant classifications of the he-

patic mass: HCC, metastatic carcinoma, abscess, cavernous hemangioma, or cirrhosis. Since the 64 patient cases used in this research had known histological or cytological results, the output classification was not arbitrary. Five output elements were used with a binary 1 or 0 for each possible predominant outcome.

For this hepatic mass application, the most successful network architecture had 35 input elements in the input layer, 35 elements in each of two hidden layers, and five elements in the output layer. Since the hidden layer architecture is a critical variable, four combinations commonly found in similar research were tested: a ratio of 1:1:1, two hidden layers of 35 elements each; a ratio of 1:2, one hidden layer of 70 elements; a ratio of 1:4, one hidden layer of 140 elements; and a ratio of 1:5, one hidden layer of 175 elements.

The subjects were patients at Veterans Administration Medical Center in Memphis, Tennessee, who had been diagnosed as having liver lesions and later had the diagnosis verified by either surgical pathology or autopsy. Only cases in which the outcomes were certain were submitted to the neural network. Our research used only cases where the hepatic masses were abnormal, the outcome was confirmed by surgical pathology or an autopsy, and some supplemental laboratory test data were available. Since physicians seldom order every available laboratory test, a patient was commonly missing one or more possible tests.

Below are the 35 different fields of data that were input to separate elements on the neural net:

- 01 Size of the hepatic mass in millimeters. Values ranged from 0 to 150 mm and were normalized to range from 0 to 1.
- 02 Binary 1 if no hepatic mass was detected; otherwise 0.
- 03 Binary 1 if single mass was detected; otherwise 0.
- 04 Binary 1 if multiple mass was detected; otherwise 0.
- 05 Binary 1 if hyperechoic (sound was reflected, so the area appeared white); otherwise 0.
- 06 Binary 1 if hypoechoic (no sound was reflected, so the area appeared gray); otherwise 0.
- 07 Binary 1 if anechoic (sound was absorbed, so the area appeared very black); otherwise 0.
- 08 Binary 1 if internal echo was detected (the area appeared white); otherwise 0.
- 09 Binary 1 if no internal echo was detected (the area appeared black); otherwise 0.
- 10 Binary 1 if sound was transmitted; otherwise 0.
- 11 Binary 1 if no sound was transmitted; otherwise 0.
- 12 Binary 1 if air was detected (often indicating abscess); otherwise 0.
- 13 Binary 1 if no air was detected; otherwise 0.
- 14 Binary 1 if computed tomography (CT) enhancement was detected; otherwise 0.
- 15 Binary 1 if no CT enhancement was detected; otherwise 0.
- 16 Binary 1 if CT fat was detected; otherwise 0.
- 17 Binary 1 if no CT fat was detected; otherwise 0.
- 18 Binary 1 if homogeneous (same texture) was detected; otherwise 0.
- 19 Binary 1 if not homogeneous (mixed texture); otherwise 0.

- 20 Amount of total bilirubin that was detected by test. Values ranged from 0.2 to 18.9 and were normalized to range from 0 to 1.
- 21 Binary 1 if no test was taken for total bilirubin.
- 22 Amount of direct bilirubin that was detected by test. Values ranged from 0.1 to 18.0 and were normalized to range from 0 to 1.
- 23 Binary 1 if no test was taken for direct bilirubin.
- 24 Amount of alkaline phosphatase that was detected by test. Values ranged from 66 to 1664 and were normalized to range from 0 to 1.
- 25 Binary 1 if no test was taken for alkaline phosphatase.
- 26 Amount of aspartate amino transferase (AST). Older terminology is glutamicoxaloacetic transaminase. Values ranged from 10 to 903 and were normalized to range from 0 to 1.
- 27 Binary 1 if no test was taken for AST.
- 28 Amount of lactate dehydrogenase that was detected. Values ranged from 78 to 2035 and were normalized to range from 0 to 1.
- 29 Binary 1 if no test was taken for LDH.
- 30 Amount of gamma glutamyl transferase (gamma GT) detected. Values ranged from 29 to 9214 and were normalized to range from 0 to 1.
- 31 Binary 1 if no test was taken for gamma GT.
- 32 Amount of albumin that was detected. Values ranged from 1.0 to 4.7 and were normalized to range from 0 to 1.
- 33 Binary 1 if no test was taken for albumin.
- 34 Amount of white blood count that was detected. Values ranged from 2.5 to 21.0 and were normalized to range from 0 to 1.
- 35 Binary 1 if no test was taken for white blood count.

Spreadsheet column numbers for the output elements and the five possible diagnoses are as follows:

- 36 Binary 1 if HCC; otherwise 0.
- 37 Binary 1 if metastatic carcinoma; otherwise 0.
- 38 Binary 1 if abscess; otherwise 0.
- 39 Binary 1 if cavernous hemangioma; otherwise 0.
- 40 Binary 1 if cirrhosis; otherwise 0.

In the data, there were 26 cases of HCC, 9 cases of metastatic carcinoma, 4 cases of abscess, 16 cases of cavernous hemangioma, and 9 cases of cirrhosis.

RESULTS

The neural network shell software (BrainMaker Professional) was trained using the data on an 80386/25 microcomputer running MS-DOS 5.00. A spreadsheet was used to input and organize the raw data and print a text file to the disk for importing into the neural network software, which randomized the 64 facts and defined the maximum and minimum parameters for normalizing the inputs. The definition file and fact file in text

(ASCII) were created. The software reads the definition file and fact files in the training phase. The neuron transfer function was sigmoid with a low value of 0, a high value of 1.0, a center of 0.5, and a gain of 1.0. Symmetry, blurring and noise were not used. Training tolerance was 0.01 (or 1%); testing tolerance was 0.2 (or 20%). Backpropagation learning rate was 1.0, and the backpropagation smoothing factor was 0.9.

To determine a suitable number of elements in the hidden layers, four different architectures were fully trained and tested. The best results (75% accuracy) were obtained with 2 hidden layers of 35 elements each (Table 1). That network was trained on 64 facts (cases) in 2220 iterations.

Table 2 shows the number of correct diagnoses, number of wrong diagnoses, total diagnoses and percentage of accuracy by disease category using 35:35:35:5 architecture (two hidden layers).

DISCUSSION

Artificial neural networks are especially useful in diagnosing hepatic masses for several reasons. Since radiologists frequently use abdominal ultrasonagraphic data and laboratory tests to perform differential diagnoses of hepatic masses, enough data are available for computer-assisted diagnoses. Analysis of hepatic masses involves a number of health care personnel with varied training, experience, and skill in image interpretation. The ability to suggest a specific diagnosis from the hepatic mass varies considerably among physicians. Based on the results of our research, a properly trained neural network could provide a consistently high level of accuracy in suggesting appropriate differential diagnostic considerations. Classification tasks are a strong forte for artificial neural networks. Such a tool would help relieve the workloads of radiologists.

In this application, the artificial neural network was 75.0% accurate in correctly diagnosing disease in the hepatic mass. This accuracy is higher than the 50% scored by the typical radiology resident in training on these same cases at the Memphis hospital, but is lower than the 90% scored by a board-certified radiologist on the same cases. These results are similar to previous studies on neural network diagnosis.^{19,31}

For future improvement, an obvious step is to add more cases to the knowledge base and retrain the neural network. For example, only four abscess cases were in the training set and the network needs more to fully capture the range of readings on a typical abscess. Likewise, if a radiologist who had seen the images and laboratory data for only four abscesses would have difficulty in correctly diagnosing abscesses.

Several technical techniques in programming neural networks merit further investi-

| Number elements on each layer | Number of correct diagnoses | Accuracy of network | Number iterations needed for training |
|----------------------------------|-----------------------------------|------------------------|---------------------------------------|
| 35:35:35:5 | 48 | 75.0 | 2220 |
| 35:70:5 | 41 | 64.1 | 4140 |
| 35:140:5 | 40 | 62.5 | 2488 |
| 35:175:5 | 39 | 60.9 | 2146 |

Table 1. Results of the Four Architectures Used

| | Correct diagnoses | Wrong diagnoses | Total diagnoses | Percentage accuracy |
|------------|----------------------|--------------------|--------------------|------------------------|
| НСС | 4 | 5 | 9 | 44.4 |
| Carcinoma | 19 | 7 | 26 | 73.1 |
| Abscess | 4 | 0 | 4 | 100.0 |
| Hemangioma | 13 | 3 | 16 | 81.3 |
| Cirrhosis | 8 | 1 | 9 | 88.9 |
| Totals | 48 | 16 | 64 | 75.0 |

Table 2. Diagnostic Accuracy by Category Using 35:35:35:5 Architecture^a

^a Note: Correct diagnoses reflect the number of times the network correctly diagnosed the fact (patient case).

gation to improve the performance of the neural network in diagnosing hepatic masses correctly. The standard backpropagation network (Rumelhard *et al.*, 1986) is only algorithm that has been attempted in this hepatic mass problem. Other network algorithms might be investigated to see if they perform more accurate diagnoses. In addition, further research should be conducted to address the choice of training feature network elements, radiographic findings, different laboratory tests, and the possibility of eliminating bad examples from the training set. If radiologists edit the training cases for a more balanced representation of possible diagnoses, the performance of the network probably would improve.

APPENDIX A: Raw Data

In processing the raw data so that it could be submitted to the neural network for training, the first field (patient number) was removed. Otherwise, the data were presented in the column order as listed below with each two lines being a fact (patient case):

```
Mass Hyp. Echo So'd Air Ench Fat Homo
```

An Artificial Neural Network

REFERENCES

 Bryan, P.J., Dinn, W.M., Grossman, Z.D., Wistow, B.W., McAfee, J.G., Kieffer, S.A., Correlation of computed tomography, gray scale ultrasonography, and radionuclide imaging of the liver in detecting space-occupying processes. *Radiology* 124:387–393, 1977.

An Artificial Neural Network

- Nelson, M.M., and Illingworth, W.T., A Practical Guide to Neural Nets. Reading, MA: Addison-Wesley, 1991.
- 3. Sugarbaker, P.H., and Kemeny, N., Management of metastatic liver cancer. Adv. Surgery 22:1-55, 1988.
- 4. Sugarbaker, P.H., Surgical decision making for large bowel cancer metastatic to the liver. *Radiology*, 174:621–626, 1990.
- Adson, M.A., Van Heerden, J.A., Adson, M.H., Wayner, J.S., Ilstrup, D.M., Resection of hepatic metastases from colorectal cancer. Arch. Surgery 119:647–651, 1984.
- 6. Malt, R.A., Current concepts: Surgery for hepatic neoplasms. New Eng. J. Med. 313:1591-1596, 1985.
- 7. Ferrucci, J.T., Liver tumor imaging: current concepts. Amer. J.l Radiol. 155:473-484, 1990.
- Boone, J.M., Gross, G.W., and Greco-Hunt, V., Neural networks in radiologic diagnosis. *Invest. Radiol.* 25:1012–1016, 1990.
- Gross, G.W., Boone, J.M., Greco-Hunt, V., and Greenberg, B., Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs. *Invest. Radiol.* 25:1017–1023, 1990.
- Maclin, P.S., Dempsey, J., Brooks, J., and Rand, J., using neural networks to diagnose renal cancer. J. Med. Syst. (U.S.), 15(1):11-19, 1991.
- 11. Mulsant, B.H., A neural network as an approach to clinical diagnosis. MD Comp 7(1):25-36, 1990.
- 12. Frenger, P., The silicon synapse or neural net computing. Biomed. Sci. Instrum. 25:267-73, 1989.
- Braham, R., and Hamblen, J.O., On the behavior of some associative neural networks. *Biol. Cybernetics* (West Germany), 2:145–151, 1988.
- 14. Crick, F., Neural networks and REM sleep. Biosci. Rep. (England) 8(6):531-535, 1988.
- Shadmehr, R., and D'Argenio, D.Z., A neural network for nonlinear Bayesian estimation in drug therapy. Neural Comput. 2:216–225, 1990.
- 16. Lippmann, R.P., Review of neural networks for speech recognition. Neural Comput. 1:1-38, 1989.
- 17. Elman, J.L., and Zipser, D., Learning the hidden structure of speech. J. Acoustical Soc. Amer. 83(4): 1615-26, 1988.
- Poli, R., Cagnoni, S., Livi, R., Coppini, G., Valli, G., A neural network expert system for diagnosing and treating hypertension. *Computer* 24:3, 64–71, 1991.
- Piraino, D.W., Amartur, S.C., Richmond, B.J., Schils, J.P., Thome, J.M., Belhobek, G.H., and Schlucter, M.D., Application of an artificial neural network in radiographic diagnosis. J. Digital Imaging 4(4):226–232, 1991.
- 20. Borman, S., Neural network applications in chemistry begin to appear. *Chem Engineering News* 67:24–28, 1989.
- 21. Simpson, P.K., Artificial Neural Systems. New York: Pergamon Press, 1990.
- 22. Dyke, R., NeuralWorks Professional II review. Neural Network News. Atlanta: AIWeek Inc., 1(3), 5, 1989.
- 23. Blum, Adam, Bidirectional Associative Memory Systems in C++. Dr. Dobb's J. 15(4), 16-26, 1990.
- Bounds, David G., and Lloyd, Paul J. A multilayer perception network for the diagnosis of low back pain. Proceedings of the IEEE International Conference on Neural Networks: Vol. II. 483–489, 1988.
- Guha, A., Using small multilayer networks to search real hyperspaces. Neural Networks Supplement: INNS Abstracts, 1:337, 1988.
- 26. BrainMaker Users Guide and Reference Manual (4th edition), Sierra Madre, CA: California Scientific Software, 1990.
- Rumelhart, D.E., McClelland, J.L., and PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vols. 1, 2, Cambridge, MA, MIT Press, 1986.
- McClelland, J.L., and Rumelhart, D.E., *Explorations in Parallel Distributed Processing*, Cambridge, MA: MIT Press, 1988.
- 29. Wasserman, P.D., Neural Computing: Theory Pract. New York: Van Nostrand Reinhold, 1989.
- 30. Neurocomputers 4(4):5, 1990. (Newsletter published monthly by Gallifrey Publishing, P.O. Box 155, Vicksburg, MI 49097. Name later changed to Sixth Generation Systems newsletter.)
- Asad, N., Doi, K., MacMahon, H., et al., Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung disease: pilot study. Radiology 177:857–860, 1990.