DIGITAL IMAGING BASICS

The Use of Computer-Assisted Diagnosis in Cardiac Perfusion Nuclear Medicine Studies: A Review (Part 3)

Frederick L. Datz, Charles Rosenberg, Frank V. Gabor, Paul E. Christian,
Grant T. Gullberg, Raj Ahluwalia, and Kathryn A. Morton

Computer-assisted diagnosis (CAID) is commonly used to evaluate cardiac nuclear medicine studies such as thallium perfusion scans. Part 1 of this series (*Journal of Digital Imaging*, 5:209-222, 1992) reviewed the basic theory underlying CAID in nuclear medicine and its use in planar thallium imaging. Part 2 discussed the application of CAID to SPECT perfusion studies (*Journal of Digital Imaging*, 6:1-15, 1993). This article reviews new variations of CAID programs for SPECT imaging and the application of expert systems and neural networks to CAID of nuclear medicine perfusion studies.
Copyright © 1993 by W.B. Saunders Company

COMPUTER-ASSISTED diagnosis (CAID) of cardiac nuclear medicine studies has become widespread.[1,2] Part 1 of this series reviewed the basic theory of CAID in nuclear medicine and its application to planar perfusion imaging.[3] Part 2 discussed the modification of planar CAID techniques for single-photon emission computed tomography (SPECT).[4] This article will review additions to the basic SPECT CAID perfusion programs and the application of expert systems and neural networks to the evaluation of cardiac nuclear medicine perfusion studies.

DIFFERENTIATION OF INFARCT FROM ISCHEMIA

Until recently, CAID programs for evaluating SPECT perfusion studies were only able to diagnose the presence or absence of coronary artery disease; they were not able to differentiate infarction from ischemia. Klein et al and Garcia et al have created a CAID program that can make this important distinction.[5,6]

These investigators developed a program that uses the same logic to differentiate ischemia from infarction that is used in the visual analysis of thallium-201 (^{201}Tl) scans. Ischemia is diagnosed visually by the appearance of relatively decreased uptake on stress images that normalizes, "reverses," or "fills in" on redistribution images. Infarction, on the other hand, shows no change between stress and redistribution images. To assist in visually determining if an area reverses, the relative activity in the area of concern is mentally compared with a normal area. If the qualitative visual "ratio" of activity is fixed, the defect is diagnosed as a scar. However, if the activity in the defect increases relative to the normal area, the defect is diagnosed as ischemia. Defects that partially redistribute are believed to contain viable but ischemic tissue that is mixed with scar.

Technique

The reversibility algorithm is a modification of Garcia's bull's-eye program (see Part 2).[7,8] Briefly, Garcia's standard program quantifies uptake by extracting maximal count circumferential profiles on each short axis tomographic slice in 9° arcs. This yields 40 data points per slice. The data are then interpolated to the

equivalent of 15 slices and stored in two 15 × 40 arrays, one for stress and the other for redistribution data. The reversibility bull's-eye program searches the 15 × 40 stress data array and identifies the 5 × 5 area in it that has the highest number of counts.[5] This area is assumed to represent normal myocardium. The array is normalized so that the 5 × 5 area is assigned an arbitrary maximal count of 1,000. The same reference area is identified in the delayed array and the entire delayed array is normalized so that the reference area is equal to 1,000.

The count data for each point in the normalized stress array are then subtracted from the corresponding point in the normalized redistribution array; the resulting data are stored in a third data structure, the reversibility array. The data in the reversibility array represent the relative change in counts between stress and redistribution in each of the 600 points that make up the bull's-eye.

To determine the normal variation in counts between stress and redistribution, the program was applied to a preexisting bull's-eye normal file. These were 20 males and 16 females who were considered to have a less than 5% likelihood of coronary artery disease based on the bayesian analysis of age, sex, symptoms, and risk factors. All had normal resting electrocardiograms (ECGs) and did not develop chest pain or ECG changes when stressed to greater than 85% of age-predicted maximal heart rate. The mean reversibility of this group was then calculated using the reversibility bull's-eye program.

In the 20 male normals, only one showed areas of reversibility greater than 1.5 standard deviations (SDs) beyond the mean in the same area that the standard bull's-eye identified as having counts 2.5 or more SDs less than normal. Four patients had areas of reversibility greater than 1.5 SD in areas that were less than 2.5 SDs less than mean counts (ie, not necessarily abnormal) on the stress images. Three of these patients' reversibility abnormalities were in the extreme apex or base. Based on these results, it was empirically decided to use an increase in relative counts (reversibility) of greater than 1.5 SD more than the mean as abnormal.

Next, five experienced observers read the routine bull's-eye plots of 42 male patients' [201]Tl

scans. Each study was scored: 0, for fixed; 1, for minimal reversal; 2, for near complete reversal; and 3, for complete reversal. Differences in observer readings were handled by consensus. Areas graded as 0 or 1 were considered fixed; defects graded 2 or 3 were considered reversible. The observers identified 17 reversible and 31 fixed defects. Based on these findings, the investigators refined their criteria based on empiric observation. The final criteria are outlined in Table 1.

The 15% size criterion was chosen because, in general, most defects had to show reversibility in at least 15% of a segment for an observer to visually diagnose reversibility. However, some defects that showed reversibility in only 5% of the pixels were perceived as having partial reversibility by experts.

To display the reversibility array data, a standard polar plot is used. In addition, a whiteout map and an SD map are produced. The whiteout bull's-eye duplicates the stress blackout display by setting points that fall more than 2.5 SDs less than the mean to black. Regions that have significantly reversed between stress and redistribution are whited-out, based on the criteria in Table 1. The SD reversibility bull's-eye indicates the number of SDs from the mean that pixel counts change between stress and redistribution; the results are displayed using a color code. A severity score is also assigned using an extraction algorithm that identifies contiguous blacked-out and whited-out pixels that form an individual defect. Using these data, the sum of the number of SDs from the mean of pixels that have been

Table 1. Criteria for a Stress Defect to be Diagnosed as Reversible

The stress study must show a defect more than 2.5 SD less than the mean.

The reversibility counts should be 1.5 SD or greater than the mean reversibility of normals in the region of a stress defect.

The size of the area showing reversibility must be greater than 15% of the area of the defect identified in the original stress bull's-eye.

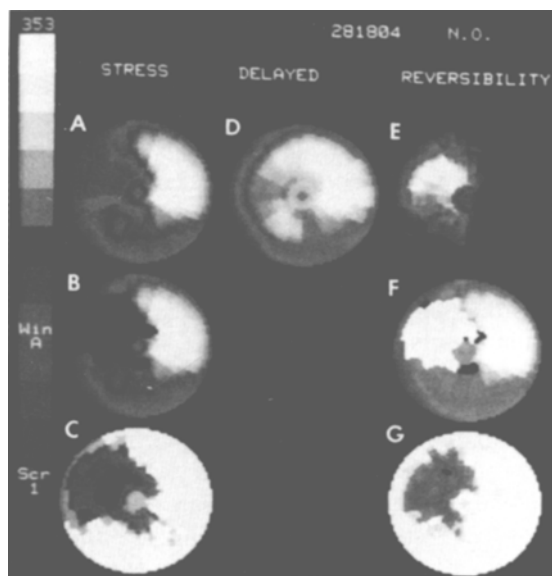The reversible area must not be in the extreme apex or base.

Fig 1. (A) Stress bull's-eye showing an anterolateral defect. (B) Blackout stress bull's-eye showing pixels less than 2.5 SD of mean as blacked-out area. (C) Scaled display showing the number of SD pixels are from the mean. (D) Delayed bull's-eye (similar to A), showing reversibility in the anterolateral defect. (E) Reversibility bull's-eye. (F) Whiteout bull's-eye display. (G) SD reversibility bull's-eye. (Reprinted with permission.[5])

blacked-out on the stress map and whited-out on the reversibility map is calculated. Examples of the various display types are shown in Fig 1.

Results

The technique has been applied to the same pilot group of patients that were used to help determine the criteria for reversibility.[5] In this group of 42 patients, 48 defects were present. The program correctly classified 81% (25/31) of the fixed defects and 82% (14/17) of the reversible defects. However, the agreement between the program and visual analysis would be expected to be good using this group.

To better judge the efficacy of their program, Garcia et al performed a multicenter trial of the reversibility program.[6] One hundred forty patients from four centers were studied. Gender-matched normal files were used for comparison, as above. Both 5% and 15% limits for the size of the defect that showed reversibility were tested. Results are summarized in Table 2: the overall correlation between the reversibility bull's-eye and human experts for the diagnosis of ischemia versus infarction was less than for the initial

pilot study. Of the 91 defects correctly identified as reversible in the multicenter study, 15 would have been missed if the lesser limit for reversible defect extent were 15% rather than 5%. Almost all of these 15 defects were classified by experts as being only partially reversible. The accuracy of the technique also varied with the vascular territory involved. The left circumflex artery (LCX) territory had significantly better overall agreement for both fixed and reversible defects than the right coronary artery (RCA) or left anterior descending (LAD) artery territories. The portion of the anterior wall perfused by the LAD also showed better agreement with the experts than the septal wall, which is also supplied by the LAD, or the RCA. The finding of better agreement in the LCX territory is consistent with previous bull's-eye testing; in these studies, the program's highest sensitivity for detecting the presence of coronary artery disease was in the RCA-LCX distribution.

The reversibility algorithm has also been applied to patients undergoing ^{201}Tl imaging before revascularization to determine if the program can predict which lesions would or would not improve with surgery.[9] In a group of 25 patients, the sensitivity was 64% for detecting ischemia (tissue that showed improvement or normalization following revascularization). The specificity for detecting scar was 91%.

There are several problems with the above studies. One problem is the standard that was used to develop the program. The readings of experts are purely subjective; no objective confirmation of the correctness of the classification was possible. Because the program's criteria for reversibility is partially derived by comparison with human interpretations, the criteria may be

Table 2. Differentiation of Infarction from Ischemia Using the Reversibility Bull's-eye

	Ischemia (%)	Scar (%)
LAD	78	67
RCA	60	84
LCX	88	89
Overall	73	80

Abbreviations: LAD, left anterior descending artery; RCA, right coronary artery; LCX, left circumflex artery.

erroneous; certainly they are not objective. Another problem is interobserver variability in the classification of ischemia versus infarction. One of the five experts disagreed with the remaining observers in classifying infarct or ischemia in 25% of defects.[5] In 10%, two of the five experts disagreed with the majority. Further evidence of interobserver variability was found in the multicenter trial. One of the four centers' experts had a statistically significant higher correlation with the reversibility bull's-eye program than the others.[6] The imaging protocol used with the program also presents a problem. Studies indicate that up to half of patients with reversible defects scanned with ^{201}Tl and imaged at 4 hours postinjection may not show evidence of ischemia. Delaying imaging for 24 hours, or reinjecting the patient with additional ^{201}Tl just before 4-hour images are obtained significantly improves the differentiation of ischemia versus infarction.[1]

TECHNETIUM-99M SESTAMIBI

Garcia et al have also developed CAID techniques for the new perfusion imaging agent, sestamibi.[10] This agent has several advantages over ^{201}Tl because it is labeled with technetium-99m (^{99m}Tc).[1,11,12] Technetium has a photon energy of 140 keV which is ideal for imaging with today's gamma cameras. In addition, the higher energy photon is less attenuated in the body than [201]Tl. Finally, the shorter half-life and different biodistribution of ^{99m}Tc means up to ten times higher doses can be used compared to ^{201}Tl, significantly improving image quality.

As a first step for adapting their ^{201}Tl program to sestamibi, Garcia et al determined the optimal acquisition and reconstruction parameters for sestamibi.[10] This allows the highest quality data possible to be inputted into the program. As noted in Part 2 of this series, the same acquisition and reconstruction protocol used to produce the normal file must be followed for patient studies so that the count comparisons are accurate.

A new technique for generating count profiles has also been developed that reduces the effects of partial-volume sampling (see Part 2 for a discussion of this problem). It has the additional advantage of reducing the amount of

observer interaction necessary for the program, eliminating another potential source of error.

In this technique, a two-part, three-dimensional sampling scheme of stacked short-axis slices is used. The apical cap is selected by identifying stacked short-axis slices that form a hemisphere; ie, the radius of the top slice of the stack is equal to the depth of the stack. The apical cap is then sampled using a spherical coordinate system in which the search for maximum counts is along a radius perpendicular to the myocardial wall. Each point in the profile represents the maximum counts per pixel for each phi and theta angular sample. The second part of the program assumes a cylindrical shape for the remainder of the myocardium and samples each short-axis slice using a cylindrical coordinate system.

The radii that run from the center of reference to the extracted maximum counts are then filtered in both depth and angle using a 3×3 lowpass filter. The filtered radii are assembled as circles and are transformed into Cartesian coordinates. It is these Cartesian coordinates and their associated maximum counts that are used to generate the polar plots.

Two new two-dimensional displays, distance-weighted and volume-weighted polar maps, have been developed to more accurately represent the location and extent of lesions (Fig 2). The distance-weighted polar map is similar to the maps currently used in the Cedars-Sinai ^{201}Tl program (see Part 2). The width of each displayed profile in the map is the same width. The profile width in an individual study is equal to the radius of the map divided by the total number of both spherical segments and cylindrical slices generated by the sampling program.

The second new display is a volume-weighted polar plot. This map was developed to correct for spatial distortions caused by displaying three-dimensional data in a two-dimensional format. Using this technique, the area a defect occupies on the map more accurately reflects the true three-dimensional volume of the lesion in the myocardium. To generate the display, the volume of the apical cap is first calculated. The proportion of the total myocardium that the apex represents is then determined. The apex is mapped onto the polar display such that the area it occupies on the plot is proportional to its
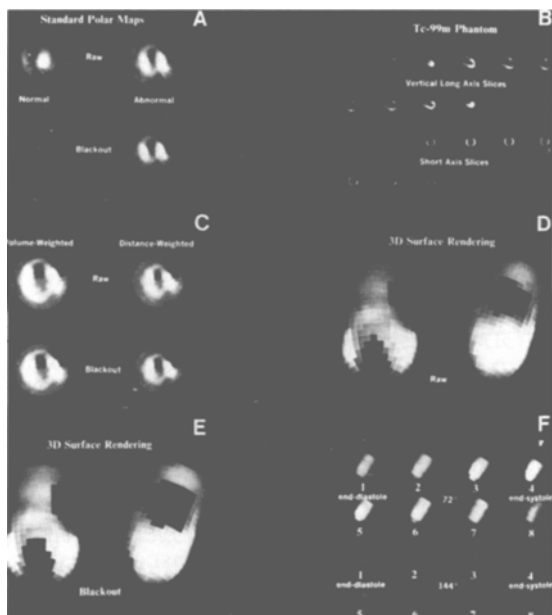
Fig 2. Examples of different displays: A is standard polar displays. B shows tomographic slices. C are volume- and distance weighted displays. D and E are three-dimensional surface renderings. F are gated images. (Reprinted with permission.[10])

true volume. In the apical cap region, each display ring is the same width. However, in the remainder of the myocardium, the area of each ring is proportional to the volume of that slice.

Both types of displays can be generated as raw displays or as binary or blackout displays. In the latter, any pixels whose values are less than a set lower limit are assigned values of 0; all other pixels are assigned values of 1. When displayed, pixels with a value of 0 are blacked-out.

Phantom and patient studies comparing the new types of displays to the conventional polar plot have shown the distance-weighted display to best indicate the location of lesions and the volume-weighted polar plot to best represent lesion size.[10] Both maps were judged equally adept at assessing the apex. The standard polar display ranked below both new displays in all three areas. However, even though these displays are significant improvements over the standard polar plot, there is no technique for mapping three-dimensional data onto two-dimensional displays that can completely prevent distortions in the size, shape, and location of perfusion defects.

EXPERT SYSTEMS

Expert systems have become fairly common for nonimaging applications.[13-15] These systems are designed to emulate the knowledge and decision-making process of a human expert in a very narrowly defined domain. The heart of expert systems is a set of heuristically derived rules that a human expert uses to make decisions in a specific area.

Expert systems are made up of three parts: a knowledge base, a set of decision rules, and an inference engine.[13-15] The knowledge base is a special database that contains the accumulated body of knowledge, the ideas and concepts, that human experts in a particular field would possess. The decision rules are the methods or rules used to make the decisions. The inference engine is the processing module of the program. It weighs input against facts in the knowledge base and rules in the decision module to derive inferences (conclusions) on which the expert system then acts. Early expert systems used hard decision rules with firm antecedents and certain consequences. However, much human decision making in areas such as medicine are based on incomplete information. For this reason, "fuzzy" logic has been applied to newer expert systems to allow the use of probabilistic rules and observations.

Technique

There have not been many fully implemented expert systems developed for nuclear medicine.[16,17] However, Ezquerra et al have constructed an expert system for the diagnosis of ^{201}Tl scans.[18-21] The program is written in LISP and consists of three modules: (1) a knowledge base contains both information about evaluating ^{201}Tl studies and the decision rules for diagnosis; (2) an inference engine makes decisions using backward chaining (see below); and (3) features extraction algorithm that examines the bull's-eye data arrays and copies the appropriate data. The architecture of program is shown in Fig 3.

The knowledge base was constructed similar to the EMYCIN expert system originally developed by Shortliffe and Buchanan at Stanford University.[22,23] It differs in two ways: (1) meta-level rules are used to help structure and control the use of knowledge; and (2) symbolic
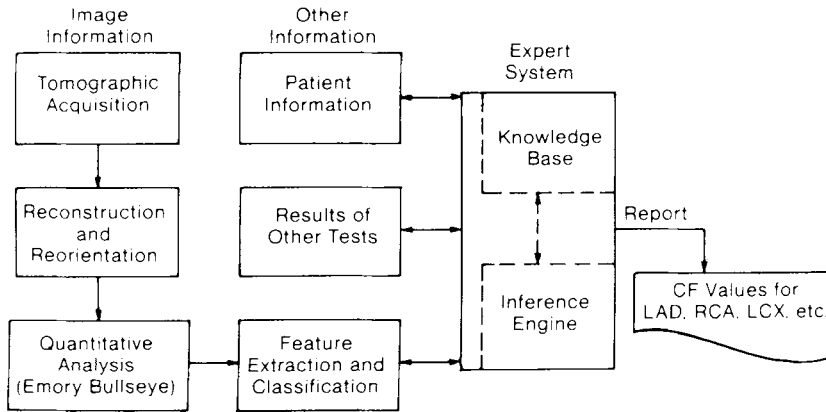
Image
Information

Other
Information

Expert
System

Tomographic
Acquisition

Patient
Information

Knowledge
Base

Reconstruction
and
Reorientation

Results of
Other Tests

Report

Quantitative
Analysis
(Emory Bullseye)

Feature
Extraction and
Classification

Inference
Engine

CF Values for
LAD, RCA, LCX, etc

Fig 3. Flow chart of the expert system for evaluating ^{201}Tl images. (Reprinted with permission.[2])

representations and descriptions of image information are created and used.

The core components of the knowledge base are frames, parameters, rules, certainty factors, and goals.[19-21] Frames are a technique to organize knowledge into manageable units by dividing problems into smaller subproblems. Each frame has its own set of rules and handles one of the subproblems. Ten frames were defined: (1) PATIENT, in which patient specific data is extracted. This includes information that can be used to increase or decrease the probability of coronary artery disease being present, such as the patient's age and sex; (2) DEFECTS, which obtains descriptions and representations of defects on the bull's-eye plot; (3) DEFECT SHAPE and (4) DEFECT LOCATION, which determine the shape and location of each perfusion defect; (5) ARTIFACT, which uses the

patient gender and the defect's location and shape to determine if it is an artifact. An example is decreased anterior wall activity in a female that could represent breast attenuation. Another artifact identified is a defect that forms a complete ring at the base of the bull's-eye plot. These defects are not classified as lesions because they are most likely due to improper slice selection during bull's-eye reconstruction; (6) LAD, (7) LCX, and (8) RCA, which tabulate incremental evidence for coronary artery disease in each of their vascular distributions; (9) UPDATE CORONARIES, which tabulates incremental evidence of the overall presence or absence of coronary artery disease; (10) PATIENT CONDITION, which determines the overall condition of the patient and generates the final printed report. The organization of these frames is shown in Fig 4.
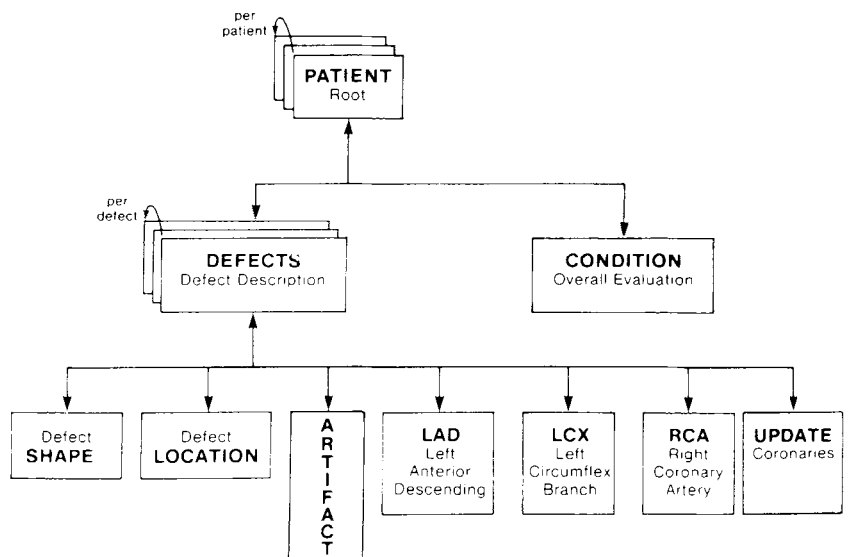
per
patient

PATIENT
Root

per
defect

DEFECTS
Defect Description

CONDITION
Overall Evaluation

Defect
SHAPE

Defect
LOCATION

A
R
T
I
F
A
C
T

LAD
Left
Anterior
Descending

LCX
Left
Circumflex
Branch

RCA
Right
Coronary
Artery

UPDATE
Coronaries

Fig 4. Frames used in expert system. (Reprinted with permission.[2])

Parameters are the names of facts or database items (eg, inferoapical, which describes that region of the myocardium).

Rules are the tools for making decisions. They are frequently heuristically derived, "rules of thumb" that human experts have stated are useful in their decision making. Rules usually consist of a premise, the "if" clause, and a predicate, the "then" clause, which determines the conclusion or action to be drawn.[13-15] An example of a rule used in the program's ARTIFACTS frame is: If a defect is ring-shaped and if it is confined to the base of the bull's-eye, then it is not a true defect. The rules used in the program were determined by reviewing studies from 291 patients who had coronary artery disease documented by angiography.[19,20] Using this data, heuristic rules were determined that best correlated the presence and location of perfusion defects on the bull's-eye plot with documented anatomic lesions. Thirty to 100 rules were generated.[21]

Certainty factors are numeric measures of confidence that facilitate the use of fuzzy logic.[19,20] They allow better manipulation of incomplete or conflicting data. An example is a score for the presence of coronary artery disease that ranges from -100 for absolute confidence of no disease, through $+100$, absolute confidence that disease is present. In the final report this is expressed in the context of sentences: "there is strong evidence (80) of right coronary artery disease."

Goals are the objectives the program actively works towards. In Ezquerra's et al program, subgoals are associated with each frame; the completion of these goals moves the program towards its overall objective, the diagnosis of coronary artery disease. For example, the goal of the ARTIFACTS frame is to determine whether defects are artifacts or true abnormalities.[19,20]

The inference engine applies the facts and rules in the knowledge base to make inferences or conclusions using backward chaining. This is a form of goal-oriented problem solving in which the program starts with a statement and a set of rules leading to the statement, and then works backward, matching the rules with information from the knowledge database until the statement can either be verified or proven wrong. For example, a rule states, "If A then B." If a part of the program triggered interest in B, the program would turn its attention backward to A and determine if A were true.[13-15]

Results

A model of the above program that used 30 rules was tested on a group of 50 patients. The results of the expert system were compared with the results of angiography and to human interpretations.[19-21] The program had a sensitivity of 98% (41/42) for detecting coronary artery disease. Specificity was 88% (7/8). The vascular territory involved was correctly identified in 80% of defects (120/150). When compared with human experts, results were similar except the agreement for vascular territories was closer, 92% (138/150).

A sample output from the program is as follows: "DEFECT SHAPE is wraparound left anterior descending (0.4); POSSIBLE ARTIFACT is none; WALL LOCATIONS are anteroseptal (0.83), septoanterior (0.77), septoinferior (0.55), and anterolateral (0.27); PATIENT CONDITION is abnormal (0.95); and DISEASED CORONARIES are left anterior descending (0.94), right coronary artery (0.27), and left circumflex (0.11)."[21]

NEURAL NETWORKS

Introduction

Neural networks are powerful multivariate statistical methods that have been applied to a wide range of problems in pattern recognition and classification.[24] Neural networks consist of computational neurons, or "units," interconnected by weighted links. Learning takes place in neural networks through adjustments of the values of the links, mimicking synaptic changes in real biological systems.

Some of the advantages of neural networks are their straightforward learning algorithms, their ability to continue to learn from "experience" after initial training, and their computation speed once trained which makes them useful for real time applications. Additionally, neural networks do not suffer the same memory and attention deficits people do.

Specific advantages of neural networks for medical imaging applications include their ability to deal with the lack of hard-and-fast rules,

to integrate medical data of many different types, and to handle the physiological variations that occur among patients.

How Neural Networks Function

The activation of a unit in a neural network is generally taken to correspond roughly to the firing rate of a biological neuron, and can vary from zero (resting potential) to one (maximal firing rate).[24] The connection strengths or "weights" correspond to the synaptic efficacy, and can have any value, positive or negative. Each unit's activation value is computed based on a weighted sum of unit activations times the weights from units feeding into it. Typically, the actual activation is a nonlinear function of this sum that approaches zero for large negative inputs and monotonically approaches one for large positive inputs. The most commonly used activation function is:

$$O_j = f(net_j) = \frac{1}{1 + e^{-net_j}} \qquad (1)$$

where O_j is the activation value of the jth unit and the total input net_j at the unit is

$$net_j = \sum_i w_{ij} O_i \qquad (2)$$

where wij is the weight from the ith to the jth unit.

The best known neural network learning procedure is back propagation.[25-27] Back propagation is a method for changing the weights or link values in a network model to iteratively allow a feed-forward network to implement a desired input-to-output mapping. Because the network developer must provide a target output pattern that represents the correct answer for each input pattern, back propagation is a supervised learning procedure. Back propagation is used primarily in networks in which the units are organized in layers, and are connected only to other units in adjacent layers. The technique is a generalization of earlier, less-powerful learning algorithms capable of learning only a single layer of weights.[28,29]

An alternative to the sigmoidal units generally used in backpropagation are radial basis function (RBF) units. RBF units are a special class of neural network unit which have local "receptive fields," as opposed to the more

common units that partition the input space in a linear fashion described above.[30,31] RBF units are most often trained in an unsupervised way, without regard to any (supervised) visual scores assigned to them. They can also be combined with a second supervised layer to map these learned encodings to desired training targets.[32] The decision as to which unit-type to use in a given situation is currently largely an experimental question.

Neural Networks in Nuclear Medicine

Although neural networks are used in a significant number of clinical and decision-making applications in ECG processing,[33,34] they have thus far found only limited application in the field of nuclear medicine. Noncardiac imaging applications include the evaluation of ventilation-perfusion scans[35] and the discrimination of normal versus Alzheimer's disease on positron emission tomography scans.[36]

Rosenberg et al have applied neural networks to the evaluation of planar ^{201}Tl images. One network consisting of a layer of gaussian RBF units was trained on cases in an unsupervised fashion to discover features in circumferential profiles of planar thallium images.[37] Then a second network was trained in a supervised way to map these features to the physician's visual interpretations of the same images using the delta rule.[28] This architecture was previously found to compare favorably with other network learning algorithms (two-layer back propagation and single-layer networks) on this task.[27,28,37]

In these experiments, all the input vectors representing single views were first normalized to unit length.

$$\vec{V} = \frac{\vec{I}}{\|I\|} \qquad (3)$$

The activation value of a gaussian unit, O_j, is then given by:

$$net_j = \sum_i (w_{ij} - v_i)^2 \qquad (4)$$

$$O_j = \exp\left(-\frac{net_j}{\sigma}\right) \qquad (5)$$

where j is an index to a gaussian unit and i is an

input unit index. The width of the gaussian, given by w was fixed at 0.25 for all units.

$$\Delta w_{i,winner} = \eta(v_i - w_{i,winner}) \qquad (6)$$

The gaussian units were trained using a competitive learning rule that moves the center of the unit closest to the current input pattern ($0max$, ie, the "winner") closer to the input pattern (see Figs 5, 6).

Technique

Scintigraphic images were acquired for each of three views: anterior, left lateral oblique, and left lateral for each patient case. Acquisition was performed twice, once immediately following a standard exercise test and once following a delay of 4 hours. Each image was preprocessed to produce a circumferential profile[38,39] in which maximum pixel counts within each of 60 contiguous segmental regions are plotted as a function of angle.[40] Preprocessing involved positioning of the region-of-interest (ROI), interpolative background subtraction, smoothing, and rotational alignment to the heart's apex[40] (See Part 1 of this series).

Cases were preselected based on the following criteria[41]: (1) Insufficient exercise—Cases in which the heart rate was less than 130 beats/min were eliminated, as this level of stress is

Fig 6. The network architecture. The first layer encoded the three circumferential profiles representing the three views. The second layer consisted of RBF units. The third layer are semilinear units trained in a supervised fashion. The outputs of the network corresponded to the visual scores as given by the expert observer. An additional unit per view encoded the scaling factor of the input patterns lost as a result of input normalization. ●, severe; ◉, moderate; ◐, mild; ○, normal.

generally deemed insufficient to accurately distinguish normal from abnormal conditions; (2) Positional abnormalities—In a few cases, the ROI was not positioned nor aligned correctly by the technician; (3) Breast artifacts—Women with prominent breast attenuation were not used; and (4) Increased lung uptake—In cases of multivessel disease, a significant proportion of the thallium activity often appears in the lungs; this makes evaluating the heart more difficult because of the partially overlapping activity from the lungs.

Cases were selected at random during a 21-month period. Approximately one third of the cases were eliminated because of insufficient heart rate, 5% because of breast artifacts, 4% because of lung uptake, and 2% because of positional abnormalities. A set of 100 usable cases remained.

Visual Interpretations Used for Training

Each case was visually scored by a single expert observer for each of nine anatomical regions: Septal (proximal and distal); Anterior (proximal and distal); Apex; Inferior (proximal and distal); and Posterior-lateral (proximal and distal). Scoring for each region was from normal (1) to severe (4), indicating the level of the observed perfusion deficit.
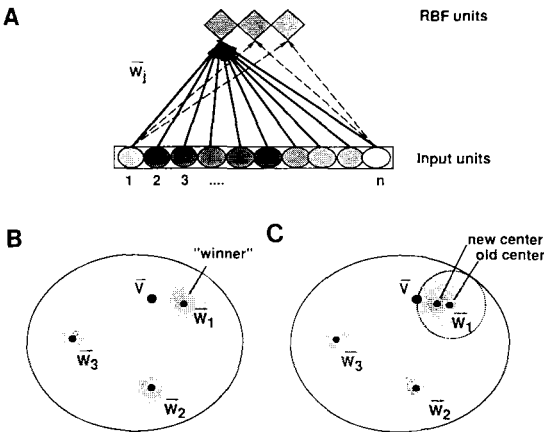

Fig 5. (A) One set of RBF units and their connections. (B and C) Geometrical interpretation of the weight update procedure for the RBF units. The shading indicates RBF activation in three units produced in response to the input pattern. Activation falls off as an exponential (gaussian) function of the euclidean distance from the unit centers. The closest W to V is the winner. (C) The winners weight vector is brought closer to V resulting in a larger activation value on the next presentation of the same input pattern. The other W's are also modified slightly.
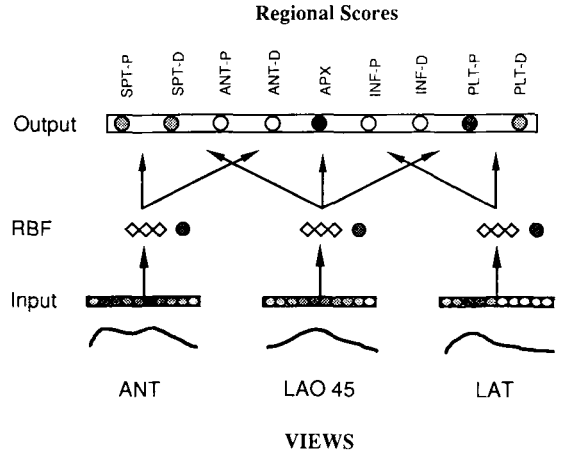
Intraobserver variability was examined by having the observer reinterpret 17 of the cases a second time. The observer was unable to remember the cases from the first reading and could not refer to the previous scores.

Exact matches were obtained on 91.5% of the regions in the visual readings; only eight of the 153 total regions (5%) were labeled as a defect (mild, moderate, or severe) on one occasion and not on the other. All differences, when they occurred, were of a single rating level.

In contrast, measured interobserver variability was much higher. A set of 13 cases was interpreted individually by three expert observers in a previous experiment.[42] Percent agreement (exact matches) between the observers was 82% (288/351). Of the 63 mismatches, 5, or approximately 8% of the regions, were of two levels of severity. There were no differences of three levels of severity. Approximately two thirds of the disagreements were between normal and mild regions. These results indicate the single observer data used in the study were more reliable than the mixed consensus and individual scores used previously.

The Network Model

The input units of the network were divided into three groups, each representing the circumferential profile for a single view. Encoding of the input data was done by setting the activations of the 60 units in each input group to the normalized values in the corresponding circumferential profile representing that view.

A set of three RBF units was assigned and trained for each view. Then a second layer of weights was trained using the delta rule[28] to reproduce the target visual scores assigned by the expert observer. The categorical visual scores were translated to numerical values to make the data suitable for network learning: normal, 0.0; mild defect, 0.3; moderate defect, 0.7; and severe defect, 1.0.

Training of a single network required only a few minutes of computer time. Once trained, results would be generated nearly immediately from the circumferential profiles.

To make efficient use of the available data, 100 identical networks were trained; each network was trained on a subset of 99 of the 100 cases and tested on the remaining one. This procedure, sometimes referred to as the "leave-

one-out" or the "jack-knife" method, allowed the generalization performance for each case to be determined. This procedure was followed for both the RBF and the delta rule training.

Each of the 100 networks was initialized and trained in the same way. RBF-to-output unit weights were initialized to small random values between 0.5 and −0.5. Input-to-RBF unit weights were first randomized and then normalized so the weight vectors to each RBF unit were of unit length.[43]

Unsupervised, competitive training of the RBF units continued for 100 epochs or complete sweeps through the set of 99 cases; 20 epochs with a learning rate of 0.1 followed by 80 epochs at 0.01 without momentum.

Supervised training using a learning rate of 0.05 and momentum 0.9 was terminated, based on cross-validation testing after 200 epochs. Further training led to overtraining and poorer generalization.

Results

Because of the larger numbers of confusions between normal and mild regions in both the interobserver and intraobserver scores, disease was defined as moderate or severe defects. The threshold value dividing the output values of the network into these two sets was varied from 0 to 1 in 0.01 step increments. The number of agreements between the expert observer and the network were computed for each threshold value. The resulting scores, accumulated over all threshold values, were plotted as a receiver operating characteristic (ROC) curve (Fig 7).

Best performance (percent correct) was achieved with a threshold value of 0.28 which yielded an overall accuracy of 88.7% (798/900 regions) on the stress data. However, this value of the threshold heavily favored specificity over sensitivity because of the preponderance of normal regions in the data. Using the decision threshold which maximized the sum of sensitivity and specificity, 0.10, accuracy dropped to 84.9% (764/900) but sensitivity improved to 0.771 (121/157), and specificity was 0.865 (643/743) (Fig 8).

In addition to diagnosing the presence or absence of coronary artery disease, neural networks were also applied to the problem of differentiating infarct from ischemia. To take into account the delayed distribution as well as
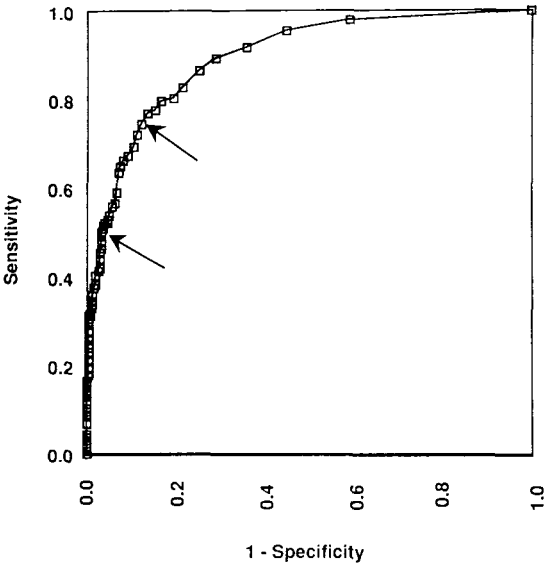
Fig 7. ROC curve for the network. The top arrow indicates the optimal value for the sum of sensitivity and specificity; the bottom arrow the optimal value for accuracy.

the stress set of images, the network was essentially duplicated: one network processed the stress data and the other processed the redistribution data.[43]

The combined network exhibited only a limited ability to distinguish between scar and ischemia. Performance on scar detection was good (sensitivity 0.728 [75/103], specificity 0.878 [700/797]), but the sensitivity of the network for ischemia detection was only 0.185 (10/54). This result may be explained, at least in part, by the much smaller number of ischemic regions in-

cluded in the data set as compared with scars (54 versus 103).

Problems

One major problem in the above study is due to limitations in defect sampling. For a statistical system (networks or otherwise) to generalize well to new cases, the data used in training must be representative of the full population of data likely to be sampled. This is unlikely to happen when the number of positive cases is on the order of 50, as was the case with ischemia, because each possible defect location, plus all the possible combinations of locations, must be included.

A variant of back propagation, called competitive back propagation, has recently been developed which is claimed to generalize appropriately in the presence of multiple defects.[44] Weights in this network are constrained to take on positive values so diagnoses made by the system add constructively. In a standard back-propagation network, multiple diseases can cancel each other out, due to complex interactions of both positive and negative connection strengths. This learning algorithm may improve the detection of ischemia.

Other possible improvements include
1. Elicit confidence ratings—Expert visual interpretations could be augmented by degree of confidence ratings. Highly ambiguous cases could be reduced in importance or eliminated. The ratings could also be
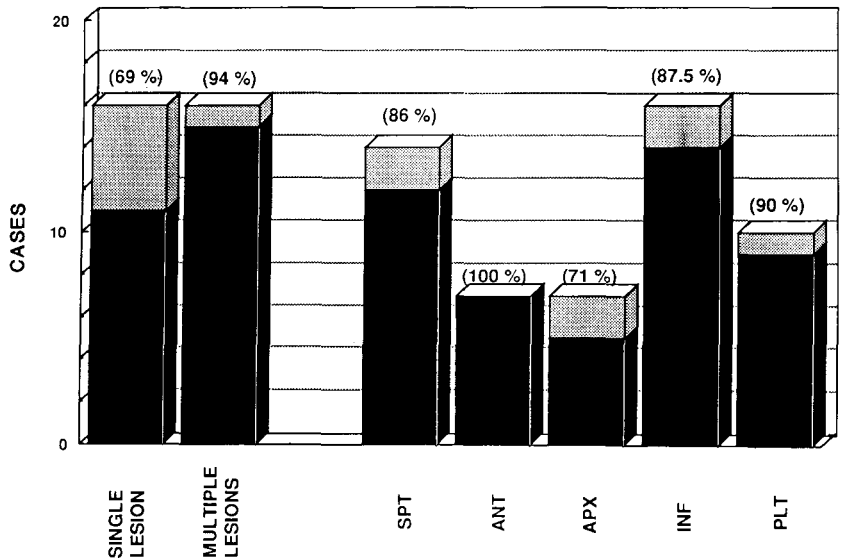
Fig 8. Number of stress defects correctly identified by the network (black) in comparison to the total for each region (gray). Proximal and distal subregions have been merged. Single lesions were defined as lesions in one of the five areas: Septal (SPT), Anterior (ANT), Apex (APX), Inferior (INF), or Posterior-lateral (PLT). Multiple lesions were cases in which moderate and severe defects occurred in more than one area.

used as additional targets for the network; cases indicated by the network with low levels of confidence would require closer inspection by a physician. Initial results are promising in this regard.

2. Provide additional information—Clinical history, gender, and exercise ECG could be incorporated into the network. Clinical history has been found to have a profound impact on interpretation of radiographs.[45] The inclusion of these variables should allow the network to approximate more closely a complete diagnosis, and boost the utility of the network in the clinical setting.

3. Add constraints—The angles that relate the three views were not used. It may be possible to build these angles in as constraints and thereby cut down on the number of free network parameters.

Related Neural Network Studies

Porenta et al trained a network using back propagation based on expert visual scores and angiography results.[46] Input to the network in both cases was a highly compressed version of the circumferential count profiles for the three planar views (see Part 1). Mean values were computed within five anatomically derived segments from the original 60 in the circumferential profile. The stress, delayed distribution, and washout profiles were all presented as 45 values (5 segments × 3 views × 3 conditions).

Performance of the network was assessed by how well it matched either the physician's visual scores or the angiographic results. The best results were obtained when the network, with a single output unit, was trained simply to predict the presence or absence of coronary artery disease based on human visual scores. The correlation results on the test set ranged from 87% to 91% on this task (93% sensitivity and 83% specificity). The network trained on the angiography results were not as good (67% agreement with angiography). This result is perhaps not surprising, given the fact that the input data, the circumferential profile, reflects a physiological factor, rather than an anatomic

variable. As noted in Part 2, a number of factors in addition to stenosis affect blood flow; therefore, no simple relationship holds between myocardial perfusion on thallium studies and stenosis on angiography. Similar results for stenosis prediction based on angiography data from ^{201}Tl scintigrams have been obtained by Cios et al.[47] They compared back propagation and the Kohonen algorithm,[48] which builds multidimensional topological maps, to a variety of "fuzzy set" rules–based construction techniques. The Kohonen network appeared to perform about as well as the rule-based approaches based on bayesian probabilities and Dempster-Shafer theory of evidence.[49] The back propagation network yielded somewhat worse results but this difference probably was not significant. The application of fuzzy set methods to the interpretation of myocardial scintigraphy also has been addressed by Rosenberg et al.[50]

Two studies have extended the application of neural networks to the interpretation of thallium SPECT data. Fujita et al compressed SPECT images to a 16 × 16 binary matrix of averaged pixel values; they then processed the data using a back propagation network.[51] The training data consisted of eight abnormal cases as judged by the consensus of three expert observers, for each coronary artery and all combinations of vessels, plus another eight cases of no disease (normal). Following training, the network was tested on two new cases for each disease type. Overall performance of the network was 77%, which compared favorably with human experts when tested individually. Cianflone et al[52] used back propagation on SPECT data to determine the abnormal vessel. The accuracy was 100% (16/16) when a single vessel was involved, and 80% (20/24) when multiple vessels were affected.

SUMMARY

CAID of coronary artery disease on nuclear medicine images is a useful technique. Expert systems and neural networks hold the promise of extending the accuracy beyond that of human observers.

REFERENCES

1. Datz FL, Gullberg GT, Gabor FV, et al: SPECT myocardial perfusion imaging update. Semin Ultrasound CT MRI 12:28-44, 1992

2. DePuey EG, Garcia EV, Ezquerra NF: Three-

dimensional techniques and artificial intelligence in Tl-201 cardiac imaging. AJR Am J Roentgenol 1152:1161-1168, 1989

3. Datz FL, Gabor FV, Christian PE, et al: The use of

computer-assisted diagnosis in cardiac-perfusion nuclear medicine studies: A review. J Digit Imaging 5:209-222, 1992

4. Datz FL, Gabor FV, Christian PE, et al: The use of computer-assisted diagnosis in cardiac perfusion nuclear medicine studies: A review (Part 2). J Digit Imaging 6:1-15, 1993

5. Klein JL, Garcia EV, DePuey EG, et al: Reversibility bull's-eye: A new polar bull's-eye map to quantify reversibility of stress-induced SPECT thallium-201 myocardial perfusion defects. J Nucl Med 31:1240-1246, 1990

6. Garcia EV, DePuey EG, Sonnemaker RE, et al: Quantification of the reversibility of stress-induced Tl-201 myocardial perfusion defects: A multicenter trial using bull's-eye polar maps and standard normal limits. J Nucl Med 31:1761-1765, 1990

7. Garcia EV, VanTrain K, Maddahi J, et al: Quantification of rotational Tl-201 myocardial tomography. J Nucl Med 26:17-26, 1985

8. VanTrain KF, Maddahi J, Berman DS, et al: Quantitative analysis of tomographic stress Tl-201 myocardial scintigrams: A multicenter trial. J Nucl Med 31:1168-1179, 1990

9. Diamond G, Forrester J: Analysis of probability as an aid in clinical diagnosis of coronary artery disease. N Engl J Med 300:1350-1358, 1979

10. Garcia EV, Cooke CD, VanTrain KF, et al: Technical aspects of myocardial SPECT imaging with Tc-99m sestamibi. Am J Cardiol 66:23E-31E, 1990

11. Wackers FJT, Berman DS, Maddahi J, et al: Technetium-99m hexakis 2-methoxyisobutyl isonitrile: Human biodistribution, dosimetry, safety, and preliminary comparison to Tl-201 for myocardial perfusion imaging. J Nucl Med 30:301-311, 1989

12. Leppo JA, Meerdink DJ: Comparison of the myocardial uptake of a technetium-labeled isonitrile analogue and thallium. Circulation 65:632-639, 1989

13. Rich E, Knight K: Artificial Intelligence (ed 2). New York, NY, McGraw-Hill, Inc, 1991

14. Levine R, Drang D, Edelsan B: Artificial Intelligence and Expert Systems: A Comprehensive Guide (ed 2). New York, NY, McGraw-Hill, Inc, 1990

15. Taylor W: What Every Engineer Should Know About Artificial Intelligence. Cambridge, Massachusetts, MIT, 1988

16. Gabor FV, Datz FL, Christian PE, et al: The use of computer-assisted diagnosis for pulmonary embolism on ventilation-perfusion lung scanning. J Nucl Med 33:1003, 1992 (abstr)

17. Gabor FV, Datz FL, Christian PE, et al: An expert system for the interpretation of radionuclide ventilation-perfusion lung scans. SPIE Proc (in press)

18. Garcia E, Ezquerra N, DePuey EG, et al: An artificial intelligence approach to interpreting Tl-201 3-dimensional myocardial distributions. J Nucl Med 27:1005, 1986 (abstr)

19. Ezquerra NF, Garcia EV, DePuey EG, et al: Development of an expert system for interpreting medical images, in: Proc IEEE Int Conference on Systems, Man and Cybernetics, vol 1. New York, NY, IEEE 1986, pp 205-210

20. Ezquerra NF, Shapiro S, Garcia EV, et al: A knowledge-based system for interpreting cardiovascular nuclear medicine images, in Computers in Cardiology. New York, NY, IEEE Computer Society, CH 2476-0:3-8, 1987

21. DePuey EG, Garcia EV, Ezquerra NF: Three-dimensional techniques and artificial intelligence in Tl-201 cardiac imaging. AJR Am J Roentgenol 152:1161-1168, 1989

22. Shortliffe EH, Buchanan BG: A model of inexact reasoning in medicine. Mathematical Biosciences 23:351-379, 1975

23. VanMelle W, Scott AC, Bennett JS, et al: The EMYCIN manual technical report, Heuristic Programming Project, Stanford University, CA 1981

24. Khanna T: Foundations of Neural Networks. Reading, MA, Addison-Wesley, 1990

25. Rumelhart DE, McClelland JL: Parallel Distributed Processing: Explorations in the Microstructure of Cognition (vols 1 and 2) Cambridge, MA, MIT, 1986

26. Cun YL: Une procedure d'apprentissage pour reseau a seuil assymetrique, in Cognitiva 85: A la Frontiere de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences, Paris, 1985, CESTA

27. Werbos P: Beyond regression: New tools for prediction and analysis in the behavioral sciences. Doctoral thesis, Harvard University, 1974

28. Widrow B, Hoff ME: Adaptive switching circuits, in 1960 IRE WESCON Convention Record 4:96-104, New York, NY, 1960

29. Rosenblatt F: Principles of Neurodynamics. New York, NY, Spartan, 1962

30. Poggio T, Girosi F: Regularization algorithms for learning that are equivalent to multilayer networks. Science 247:978-982, 1990

31. Broomhead DS, Lowe D: Multivariable functional interpolation and adaptive networks. Complex Systems 2:321-355, 1988

32. Moody J, Darken C: Fast learning in networks of locally-tuned processing units. Neural Computation, 1:281-294, 1989

33. Baxt WG: Use of an artificial neural network for diagnosis of myocardial infarction. Ann Int Med 115:843-848, 1991

34. Baxt WG: Use of an artificial neural network for data analysis in clinical decision-making: The diagnosis of acute coronary occlusion. Neural Computation 2:480-489, 1991

35. Gabor FV, Datz FL, Christian PE, et al: Use of computer-assisted-diagnosis for pulmonary embolism on ventilation-perfusion lung imaging. J Nucl Med 33:1003, 1992

36. Kippenhan JS, Barker W, Pascal S, et al: A neural-network classifier applied to PET scans of normal and Alzheimer's disease patients, in Proceedings of 37th Annual Meeting Society of Nuclear Medicine 31, Washington, DC, May 1990

37. Rosenberg CR, Erel J, Atlan H: A neural network that learns to interpret myocardial planar thallium scintigrams, in Giles CL, Hanson SJ, Cowan JD (eds): Advances in Neural Information Processing Systems 5. Palo Alto, CA, Morgan Kaufmann, 1993

38. Garcia EV, Maddahi J, Berman DS, et al: Space-time quantitation of thallium-201 myocardial scintigraphy. J Nucl Med 22:309-317, 1981

39. Francisco DA, Collins SM, Go RT, et al: Tomographic thallium-201 myocardial perfusion scintigrams after maximal coronary artery vasodilation with intravenous dipyr-

idamole: Comparison of qualitative and quantitative approaches. Circulation 66:370-379, 1982

40. 40. Garcia EV: Physics and instrumentation of radionuclide imaging, in Marcus ML, Schelbert HR, Skorton DJ, et al (eds): Cardiac Imaging. Philadelphia, PA, Saunders, 1991

41. Beller GA: Myocardial perfusion imaging with thallium-201, in Marcus ML, Schelbert HR, Skorton DJ, et al (eds): Cardiac Imaging. Philadelphia, PA, Saunders, 1991

42. Rosenberg CR, Erel J, Atlan H: A neural network that learns to interpret myocardial planar thallium scintigrams. Neural Computation, 1993 (in press)

43. Erel J, Rosenberg CR, Atlan H: Automatic interpretation of myocardial planar thallium scintigrams using neural networks. Presented at the American College of Cardiology 42nd Annual Scientific Session, March 14-18, 1993, Anaheim, CA.

44. Cho S, Reggia JA: Multiple disorder diagnosis with adaptive competitive neural networks. Artificial Intelligence in Medicine, 1993 (in press)

45. Doubilet P, Herman PG: Interpretation of radiographs: Effect of clinical history. AJR Am J Roentgenol 137:1055-1058, 1981

46. Porenta G, Kundrat S, Dorffner G, et al: Computer-based image interpretations of thallium-201 scintigrams: Assessment of coronary artery disease using the parallel distributed processing approach, in Proceedings of the 37th Annual Meeting of the Society of Nuclear Medicine, May 1990, p 825

47. Cios KJ, Goodenday LS, Merhi M, et al: Neural networks in detection of coronary artery disease, in Computers in Cardiology Conference, Jerusalem, Israel, IEEE Computer Society Press, September 1989, pp 33-37

48. Kohonen T: Self-organization and Associative Memory (ed 3). Berlin, Germany, Springer-Verlag, 1989

49. Cios KJ, Shin I, Goodenday LS: Using fuzzy sets to diagnose coronary artery stenosis. Computer, March, 57-63, 1991

50. Rosenberg S, Itti R, Benjelloun L: Symbolic reasoning about myocardial scintigrams in prolog. Eur J Nucl Med 12:65-68, 1986

51. Fujita H, Katafuchi T, Uehara T, et al: Application of artificial neural network to computer-aided diagnosis of coronary artery disease in myocardial bull's-eye images. J Nucl Med 33:272-276, 1992

52. Cianflone D, Carandente O, Fragasso G, et al: A neural network-based model of predicting the probability of coronary lesion from myocardial perfusion SPECT data, in Proceedings of the 37th Annual Meeting of the Society of Nuclear Medicine, May 1990, p 797