Automated Interpretation of Myocardial SPECT Perfusion Images Using Artificial Neural Networks

Dan Lindahl[1], MSc, John Palmer[2], PhD, Mattias Ohlsson[3], PhD, Carsten Peterson[3], PhD, Anders Lundin[4], MD, Lars Edenbrandt[1], MD, PhD

Departments of [1]Clinical Physiology, [2]Radiation Physics, [3]Theoretical Physics, and [4]Radiology, Lund University, Lund, Sweden

This study was supported by grants from the Swedish Medical Research Council (B95-14X-09893-04B), the Faculty of Natural Sciences at Lund University, Sweden, Swedish National Board for Industrial and Technical Development and the Swedish Natural Science Research Council.

Address for correspondence and reprints
Lars Edenbrandt
Department of Clinical Physiology
University Hospital
S-221 85 Lund
Sweden

Telephone	+46 46 17 33 05
Fax	+46 46 15 17 69
E-mail	lars.edenbrandt@klinfys.lu.se

Abstract

The purpose of the present study was to develop a computer-based method for automatic detection and localization of coronary artery disease in myocardial bull's-eye scintigrams. Methods: A population of 135 patients who had undergone both myocardial technetium-99m-sestamibi rest-stress scintigraphy and coronary angiography within 3 months was studied. Different image data reduction methods, including pixel averaging and 2-dimensional Fourier transform, were applied to the bull's-eye scintigrams. After a quantitative and qualitative evaluation of these methods, 30 Fourier components were chosen as inputs to multilayer perceptron artificial neural networks. The networks were trained to detect coronary artery disease in two vascular territories using coronary angiography as gold standard. A "leave one out procedure" was used for training and evaluation. The performance of the networks was compared to those of two human experts. Results: One of the experts detected coronary artery disease in one of two vascular territories with a sensitivity of 54.4% at a specificity of 70.5%. The sensitivity of the networks was significantly higher at that level of specificity (77.2%, p=0.0022). The other expert had a sensitivity of 63.2% at a specificity of 61.5%. The networks had a sensitivity of 77.2% (p=0.038) also at this specificity. The differences in sensitivity between experts and networks for the other vascular territory were all less than 6% and not statistically significant. Conclusions: Artificial neural networks can detect coronary artery disease in myocardial bull's-eye scintigrams with such a high accuracy that the application of neural networks as clinical decision support tools appears to have significant potential.

Key words: Diagnosis, computer-assisted; Artificial intelligence; Neural networks, computer; Heart disease, ischemic; Radionuclide imaging

Abbreviations used:

CAD	Coronary Artery Disease;
LAD	Left Anterior Descending artery
LCX	Left Circumflex Artery
RCA	Right Coronary Artery
RMS	Root Mean Square
ROC	Receiver Operating Characteristics
SPECT	Single Photon Emission Computed Tomography.

Introduction

Computer-aided interpretation of diagnostic images has gained much interest in the fields of radiology, nuclear medicine and magnetic resonance imaging (1, 2). Computer technology can be used to support non-experts with a preliminary interpretation in situations where experts are not present. Interpretation of diagnostic images is a pattern recognition task, the result of which generally cannot be encapsulated in a set of criteria. Consequently conventional rule-based expert systems have achieved only limited success in this area. Artificial neural networks represent a computer-based decision method that has proved to be of special value in pattern recognition tasks (3, 4). It is therefore of interest to evaluate the feasibility of using artificial neural networks for interpretation of diagnostic images. However, the large data content in diagnostic images causes problem in training the neural networks.

Artificial neural networks learn by example. The number of examples needed for the network training depends on the size of the network. A large network fed with many input variables needs many examples to train properly. Images, especially in radiology but also in nuclear medicine and magnetic resonance imaging, contain large numbers of pixels. A commonly used image matrix in a scintigram is 256×256, i.e. 65,536 pixels. If all these pixel values were fed to a neural network thousands of examples would be needed. The number of images available for training is typically in the order of 100. A substantial data reduction must therefore be performed without losing the relevant information, prior to the network training.

The purpose of the present study was to develop a computer-based method to classify myocardial perfusion bull's-eye images. Different methods to reduce the data volume of the images were studied. Thereafter, artificial neural networks were trained to detect coronary artery disease (CAD). The performance of the networks was compared to that of two human experts using coronary angiography as gold standard.

Materials and Methods

Patient Population

All patients at Lund University Hospital, who during the period from November 1992 to October 1994 had undergone both rest-stress myocardial perfusion scintigraphy and coronary angiography within three months, were studied retrospectively. The total material consisted of 166 patients and 31 of these patients were excluded because they had undergone angioplasty or coronary artery bypass surgery or had signs of progressive CAD between the scintigraphy and angiography. The population studied consisted of 135 patients (94 males and 41 females, age mean 56.7, range 21-77 years). A contrast left ventriculogram was performed in 106 patients. The ejection fraction was normal in 65 patients (61%), slightly to moderately reduced in 28 (26%) and severely reduced in 13 patients (12%).

Coronary Angiography

Coronary angiography was used as gold standard. The patients were examined using the standard Judkins technique. Angiograms were performed and interpreted by experienced cardiac radiologists. Each coronary artery was examined in 4-6 projections, of which at least two were orthogonal. Significant CAD was defined as 75% or more lumen area reduction in a major coronary artery. The severity of a coronary stenosis was determined by visual assessment. The total myocardial perfusion bed was divided into two vascular territories. One territory was assigned to the vascular bed of the left anterior descendent artery (LAD); the other territory was assigned to the vascular bed of the left circumflex artery (LCX) and the right coronary artery (RCA) together. Each territory was studied separately regarding presence or absence of CAD. In 41 patients CAD was found in both the LAD and the RCA/LCX territories and in 46 patients CAD was found in one territory only (15 LAD and 31 RCA/LCX). Coronary angiography did not reveal significant CAD in the remaining 48 patients.

Myocardial Scintigraphy

Rest and stress studies were performed in a one-day technetium-99m-sestamibi protocol, using 300 MBq at rest, and after a delay of about three hours, 900 MBq at stress. The time period between injection and imaging was one to two hours for the rest studies and 30 minutes to one hour for the stress studies. Out of the 135 patients studied, 131 underwent exercise on a bicycle ergometer and four were injected with dipyridamole (0.56 mg/kg). Exercise was symptom-limited (anginal pain, severe dyspnea or severe fatigue) unless malignant arrhythmia or exercise hypotension occurred (>10 mmHg drop between exercise stages). Workload was increased in a stepwise manner with 10 W/min. Exercise was continued for 1 minute after injection.

The scintigraphy data were acquired in continuous single photon emission computed tomography (SPECT) over 180 degrees during 20 minutes with a 51 $\times$ 37 cm gamma camera (Toshiba GCA-901A/SA), using a low-energy, high resolution collimator (full width at half maximum 9.7 mm at 15 cm) and a zoom factor of 1.5 (pixel size 5.33 mm). Data were binned to 45 projection angles in a 64 $\times$ 64 matrix. Reconstruction was undertaken with standard techniques, using a 2-dimensional Butterworth pre-reconstruction filter (5)

$$f(\omega) = 1/(1+(\omega/\omega_c)^m),$$

with $m = 6$ and $\omega_c = 0.25$ cycles per pixel, followed by a ramp reconstruction filter.

A short-axis slice set was generated for entry into a bull's-eye (6) program, in which careful 3-dimensional alignment of the rest and stress images were performed on an interactive basis by experienced clinical operators. This procedure included translations along the left ventricle axis and small corrections to the initially selected direction of this axis. Rotations about the axis were not made. After slice selection and positioning, data were sampled for the myocardial maximum in 64 radial directions for each slice, including the apex. The sampled data were then linearly interpolated to exactly 17 slices, and organized in a matrix map of size 17 slices $\times$ 64 angles. The rest matrix, corrected for decay of technetium-99m during the interval from rest to stress, was subtracted from the stress matrix as background. The maps were then normalized as follows: the rest map was scaled such that the average value of the region above 90% of its maximum was put to a fixed value, the same for all patients. The stress map was then scaled such that its average value, in the region above 90% of its maximum, was made equal to the average value of the rest map in a geometrically identical region. This normalization makes the rest and stress data equal in a region which is likely to be least ischemic and makes the absolute normalization dependent on the rest data only. The polar representation of these data was only used for visual classification; the neural networks worked exclusively on Cartesian data.

Image Data Reduction

In order to shrink the dimensionality of the artificial neural network input space, different image data preprocessing methods were evaluated. In Figure 1 an original rest image is presented in polar form together with five images reproduced using different data reduction methods. The original 17×64 image matrix was coarsened to 1×8 and 4×4 pixels respectively as shown in Figure 1 b-c. The resulting numbers of variables for a rest and a stress study together were 16 and 32. Figure 1 d-f shows reconstructed rest images using 2-dimensional Fourier transform technique with inverse transformation of 18, 30 and 72 components for the rest and stress studies together.

The following technique for extracting Fourier components from the rest and stress images was used. Each of the two 17×64 images was expanded by mirroring about row 17, and then discarding the last row (i.e. the first row of the succeeding Fourier period), to produce 32×64 matrices. By this construction, only local data were in effect convoluted by the extraction of a few of all the available Fourier components, irrespective of slice location. The two 32×64 matrices were input as the real and imaginary parts of a complex 32×64 matrix in a fast Fourier transform (7) The spatial low-frequency components describing the rest and stress images are found near the origin in the spatial frequency plane, as mapped by the transformed complex matrix. Selections of up to 36 complex low frequency components (72 real values) were made for further image analysis. Since the relative importance of axial and radial information is a priori unknown, different sets of components were tried.

The number of components needed to reproduce the original rest and stress images in a satisfactory way was analyzed both quantitatively and qualitatively. The differences in pixel values of the original images and the corresponding values of the inverse-transformed data were studied using root mean square (RMS) error calculations. A visual inspection of original and inverse-transformed images was also performed as in Figure 1. The best method, with consideration taken both to amount of image reduction and differences between original and inverse-transformed images, was then used to calculate input variables to the artificial neural networks.

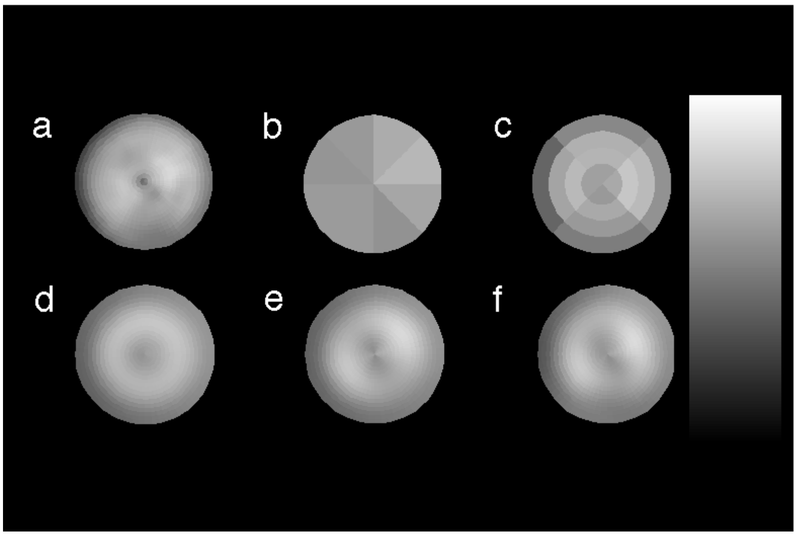


Figure 1. Comparison between (a) original bull's-eye image and the corresponding images reconstructed with different data reduction methods; b-c) coarsened images obtained by pixel averaging in 8 and 16 sectors; d-f) images reconstructed from 18, 30 and 72 Fourier components.

Neural Network

A multilayer perceptron neural network architecture (8) was used. The networks consisted of one input layer, one hidden layer and one output layer. The number of neurons in the input layer was equal to the number of input variables (30). The hidden layer contained three neurons and the output layer contained one neuron which encoded whether CAD was present or not. Three different sets of networks were studied, one that determines whether CAD was present or absent regardless of the location, one that detects CAD in the LAD territory and one that detects CAD in the RCA/LCX territory. The same type of network architecture and training parameters were used for the different networks.

During the training process the connection weights between the neurons were adjusted using the back-propagation algorithm. The sigmoid activation function was used. The learning rate had a start value of 0.3. During the training it was decreased between epochs. The momentum was set to 0.7. Updating occurred after each twenty patterns. The network weights were initiated with random numbers between -0.03 and 0.03. Training was set to stop at a training error of 0.36. All calculations were done using the JETNET 3.0 package (9).

The output values for test cases were in the range from 0 to 1. A threshold in this interval was used above which all values were regarded as consistent with CAD. By varying this threshold a receiver operating characteristic (ROC) curve was obtained.

”Leave One Out” Procedure

In order to get as reliable performance as possible a ”leave one out” validation procedure was used. One patient study was used as test case while the remaining 134 cases were used for training. This procedure was repeated 135 times such that each case in the data set once was used as a test case. The test results of the 135 different networks were then concatenated and the resulting list was used in the calculations of neural network performance.

Human Expert Classification

The performance of the networks was compared with those of two human experts. The bull's-eye images were presented to the experts in random order. Neither clinical data nor the results from angiography, neural networks or the classification of the other expert were available during the classification procedure. The experts had to rely on four bull’s-eye images per patient study only: rest image, stress image, difference image and ratio image. They did not view short or long axis images. The experts classified each patient study and each vascular territory for the presence of CAD using a 4-grade scale; 'definitely not CAD', 'probably not CAD', 'probably CAD' and 'definitely CAD'.

Statistical methods

Sensitivity and specificity for the experts in assessing CAD/no CAD was calculated and plotted together with the ROC curves for the networks. Areas under the ROC curves were calculated for the networks as a measure of performance. The 4-grade scale used by the experts made it possible to calculate three sensitivity/specificity pairs for each expert. The comparisons between networks and experts were performed as follows. The threshold applied to the network outputs was chosen such that the specificity of the neural networks was the same as that of the expert. Thereafter the corresponding sensitivity of the networks was compared to the sensitivity of the expert and the significance of the difference in sensitivity was tested paying attention to the fact that the same scintigrams were used, i.e. a McNemar type of statistic was used.

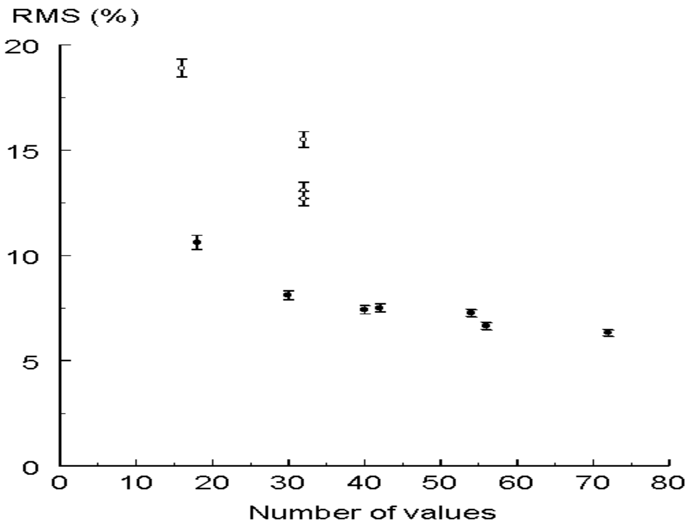
Results

Image Data Reduction

The results of the quantitative comparison between the different image data reduction methods are presented in Figure 2. The RMS values of the Fourier transformed and inverse transformed images were lower than those of the images which were coarsened with pixel averaging, compared at the same number of variables. An increase in the number of variables used to describe an image reduced the RMS error irrespectively of the method used. In the images based on

9 complex Fourier components (18 values), small but clinically important perfusion defects are not reproduced satisfactory. Most of the relevant information of the original images is contained in the images based on both 15 (30 values) and 36 complex components (72 values) (Figure 1) but the RMS is not reduced in proportion to the increased number of components. For the 72 values the number of neurons in the network would exceed the number of training examples and, as a few preliminary runs indicated, the results did not improve. Therefore, 15 complex Fourier components were used as inputs to the artificial neural networks.

Figure 2. RMS values for reconstructed images using different image data



reduction methods. Non-filled circles denote pixel averaging methods. Filled circles denote Fourier transform methods. The values shown are mean and standard error of the mean.

Image Classification

The ROC curve of the artificial neural networks detecting CAD in the LAD territory is presented in Figure 3 together with the specificity / sensitivity value pairs of the human experts. One of the experts had a sensitivity of 54.4% at a specificity of 70.5%. The sensitivity of the networks was 77.2% at that level of specificity. This difference in sensitivity of 22.8% was statistically significant (p=0.0022). The other expert had a sensitivity of 63.2% at a specificity of 61.5%. At this specificity the networks had a sensitivity of 77.2%, and this difference of

14.0% was also statistically significant (p=0.038). No other differences in sensitivity between experts and networks were statistically significant.

The ROC curve of the networks detecting CAD in the RCA/LCX territory is presented in Figure 4. The area under this curve was 0.82 compared to the area under the LAD curve of 0.76. The networks determining CAD regardless of location had an area under the ROC curve of 0.80 (Figure 5). The differences in sensitivity between experts and networks shown in the Figures 4 and 5 were all less than 6% and not statistically significant.

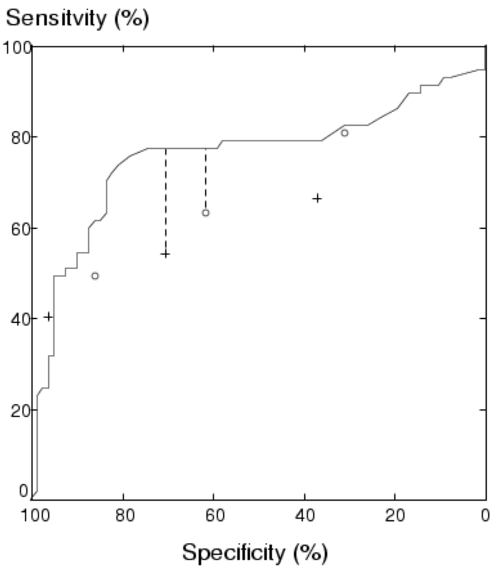


Figure 3. Classification performance for detecting CAD in LAD territory. The neural network performance are presented as an ROC curve and experts' results are shown as 'o' (expert 1) and '+' (expert 2). Statistically significant differences in sensitivity between networks and experts are indicated with vertical lines.

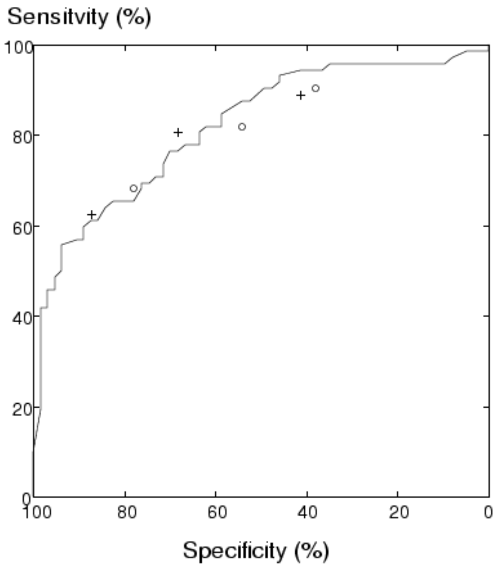
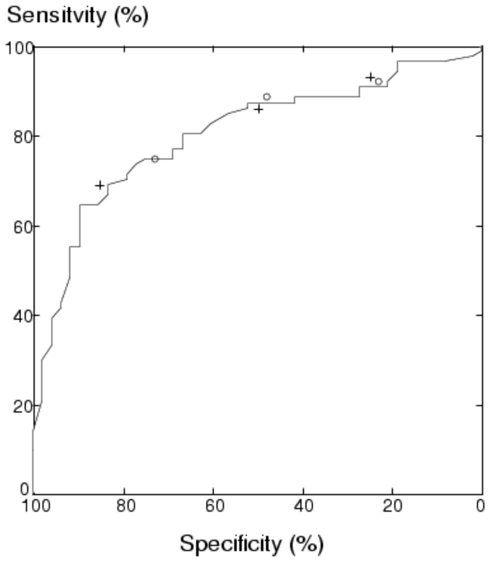


Figure 4. Classification performance for detecting CAD in the RCA/LCX territory. Same notation as in Figure 3.

Figure 5. Classification performance for detecting CAD regardless of localization. Same notation as in Figure 3.



Discussion

We have demonstrated that artificial neural networks can be used to classify myocardial perfusion images regarding presence and locality of CAD. The performances of the networks were similar or better than those of two human experts. These results indicate that neural networks could be used to assist clinicians in achieving a correct interpretation and thereby improve the diagnostic accuracy of medical imaging. An important ingredient in the computer-based method to classify bull's-eye images is the 2-dimensional Fourier transform.

The networks detected CAD with better performance in the RCA/LCX territory than in the LAD territory. Also the experts performed best in the RCA/LCX territory. Neither networks nor experts could achieve a very high sensitivity even at a low specificity in the LAD territory as shown in Figure 3. These findings indicate that they cannot find scintigraphic abnormalities in the LAD territory in a number of cases with CAD according to coronary angiography. This could be due to an angiographic overestimation of lesion severity or functionally important coronary collaterals. However, the same disagreement between angiographic and scintigraphic findings were not found in the RCA/LCX territory.

Why Image Data Reduction?

When using artificial neural networks two things have to be considered regarding the input data. On the one hand it is an advantage to present as much information as possible to the network. On the other hand, a large number of inputs results in a network with many connection weights which requires a large training database in order to generalize well. This is because the number of connection weights or fitted parameters depends strongly on the number of input neurons. The number of input neurons in a network is equal to the number of variables which are presented to the network. In general the majority of weights in a neural network are connecting the input neurons and the hidden neurons.

Image data reduction with 2-dimensional Fourier transform reduces the need for a large number of input neurons. The 30 variables used in this study contain information sufficient to reproduce the scintigrams in a satisfactory way as shown in Figure 1. The method preserves more of the original image structure than methods where the matrix for each of the rest and stress images are coarsened to distinct predefined regions using pixel averaging. This superior performance is not surprising since the Fourier transform does not favor fixed spatial locations as a pre-defined subdivision might do.

The Fourier transform is well suited for the analysis of myocardial bull’s-eye images. However, in applications where the information content of the images lie both in the high and low spatial frequency domains, the method may need some additional modification or may be less suitable.

Another general approach that has been used is the 'preclassification method'. Human observers analyze the images and make classifications into a number of predefined classes which then are used as input data (10, 11). One problem with this method is the intra- and interobserver variability. The network could produce different output values for the same image if it is trained with different preclassifications. There is also a considerable risk that information missed by the observers, is lost.

Requirements on the training material

The performance of an artificial neural network largely depends on the size and composition of the training database. In this study a database of 135 patients who had undergone both scintigraphy and coronary angiography was used. The training sets contained 134 patients and the networks had 97 connection weights. The resulting ratio between the number of examples in the training set and the number of connection weights in the network was above one. Conversely, the group of Fujita (12) used 58 myocardial perfusion images to train networks that contained 26,500 connection weights. We find it remarkable that they designed a study around networks with so many weights and so few examples. Surprisingly, their networks correctly classified many of the cases in the test sets. One reason for this is probably the selection of cases in the database; 74 relatively typical cases from SPECT bull’s-eye examinations were utilized. Cases where the diagnoses obtained by angiography and scintigraphy differed were excluded.

Porenta and coworkers (13) trained neural networks containing more than 700 connection weights with less than 40 examples from myocardial scintigraphy with thallium-201. Their networks performed significantly worse than a human expert. We think this is due to the small number of examples relative to the number of connection weights.

The composition of the training set is as important as the number of examples. The examples must be representative of patient studies found in a clinical setting. Networks trained with a group of selected typical cases, or cases with purely synthesized defects will not be useful in clinical practice. Because of the choice of clinical cases in our study, the results should give an accurate hint of the usefulness of neural networks in a clinical environment.

Expert systems

Computer-aided decision support systems can be based on different methods from the field of artificial intelligence. The most widely used decision support systems, interpretation programs in computerized electrocardiographs, are generally based on conventional expert systems. Artificial neural networks have only recently been implemented in these programs. Expert systems have also been developed for interpretation of myocardial bull's-eye images (14-16), in which criteria for the detection of CAD were developed using separate normal limits in different predefined territories of the images. However, the visual interpretation of an image is generally not dependent on simple criteria but to a great extent a pattern recognition task. Artificial neural networks have shown to be well suited for this type of tasks. Networks have outperformed very complex expert systems, for example electrocardiographic interpretation programs (17). Therefore, it was of interest to apply neural networks for the interpretation of bull's-eye images. One advantage with the neural network approach is the possibility to use Fourier transform components as inputs. These components constitute an unbiased description of the image in contrast to the predefined parameters commonly used in expert systems.

The choice of gold standard

Coronary angiography is considered to be the appropriate gold standard for the diagnosis of CAD. However, myocardial scintigraphy can provide additional information of clinical importance (18). For example, the extent and severity of ischemia as reflected by the myocardial scintigraphy provide prognostic information in patients with known CAD.

We used coronary angiography as gold standard because this provides an independent reference in the comparison of network and human experts. The results show that artificial neural networks can classify scintigrams in search for CAD with angiography as gold standard. We believe that networks also could classify scintigrams in search for ischemia using human experts as gold standard.

Conclusions

We found it possible to develop a computer-based method to classify myocardial perfusion bull's-eye images with performance as good as or better than a human expert. Artificial neural networks can detect CAD in myocardial bull's-eye scintigrams with such a high accuracy that the application of neural networks as clinical decision support tools appears to have significant potential.

References

1. Miller AS, Blott BH, Hames TK. Review of neural network applications in medical imaging and signal processing. *Med Biol Eng Comput* 1992;30:449-464.
2. Datz FL, Rosenberg C, Gabor FV, et al. The use of computer-assisted diagnosis in cardiac perfusion nuclear medicine studies: a review. *J Dig Imag* 1993;6:67-80.
3. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346:1075-1079.
4. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135-1138.
5. Gonzalez Rc, Wintz P. Digital Image Processing, second edition. United States: Addison Wesley; 1987:170-171.
6. Van Train KF, Garcia EV, Cooke D, Areeda J. Quantitative Analysis of SPECT Myocardial Perfusion. In: DePuey EG, Berman DS, Garcia EV, eds. Cardiac Spect Imaging, New York: Raven Press; 1995:49-74.

7 Press WH, Teukolsky SA, Vetterling WT, Flammery BP. Numerical Recipies in C, Second edition. Cambridge University Press; 1995:521-525.

8. Rumelhart DE, McClelland JL, eds. *Parallell distributed processing.* Volumes 1 & 2. Cambridge, MA: MIT Press; 1986.
9. Peterson C, Rögnvaldsson T, Lönnblad L. JETNET 3.0 - A versatile artificial neural network package. *Comp Phys Comm* 1994;81:185-220.
10. Scott JA, Palmer EL. Neural network analysis of ventilation-perfusion lung scans. *Radiology* 1993;186:661-664.
11. Tourassi GD, Floyd CE, Sostman HD, Coleman RE. Artificial neural network for diagnosis of acute pulmonary embolism: effect of case and observer selection. *Radiology* 1995;194:889-893.
12. 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.
13. Porenta G, Dorffner G, Kundrat S, Petta P, Duit-Schedlmayer J, Sochor H. Automated interpretation of planar thallium-201-dipyridamole stress-redistribution scintigrams using artificial neural networks. *J Nucl Med* 1994;35:2041-2047.
14. Van Train KF, Areeda J, Garcia EV et al. Quantitative same-day rest-stress technetium-99m-sestamibi SPECT: Definition and validation of stress normal limits and criteria for abnormality. *J Nucl Med* 1993; 34:1494-1502.
15. Van Train KF, Garcia EV, Maddahi J et al. Multicenter trial validation for quantitative analysis of same-day rest-stress technetium-99m-Sestamibi Myocardial Tomograms. *J Nucl Med* 1994;35:609-618.
16. Kahn JK, McGhie I, Akers MS et al. Quantitative rotational tomography with 201-Tl and 99m-Tc 2-methoxy-isobutyl-isonitrile: a direct

comparision in normal individuals and patients with coronary artery disease. *Circulation* 1989; 79:1282-1293.

17. Hedén B , Edenbrandt L, Wesley K, Haisty Jr, Pahlm O. Artificial Neural Networks for the Electrocardiographic Diagnosis of Healed Myocardial Infarction. *Am J Cardiol 1994;74:5-8.*
18. Berman DS, Kiat H, Friedman JD, Diamond G. Clinical applications of exercise nuclear cardiology studies in the era of healthcare reform. *Am J Cardiol* 1995;75:3D-13D.