Int J Cardiovasc Imaging (2008) 24:841–848
DOI 10.1007/s10554-008-9329-x

ORIGINAL PAPER

Use of neural networks to improve quality control of interpretations in myocardial perfusion imaging

K. Tägil · J. Marving · M. Lomsky · B. Hesse · L. Edenbrandt

Received: 2 January 2008 / Accepted: 4 June 2008 / Published online: 29 June 2008

Abstract *Background*: The aim of this study was to explore the feasibility of using a technique based on artificial neural networks for quality assurance of image reporting. The networks were used to identify potentially suboptimal or erroneous interpretations of myocardial perfusion scintigrams (MPS). *Methods*: Reversible perfusion defects (ischaemia) in each of five myocardial regions, as interpreted by one experienced nuclear medicine physician during his daily routine of clinical reporting, were assessed by artificial neural networks in 316 consecutive patients undergoing stress/rest 99mTc-sestamibi myocardial perfusion scintigraphy. After a training process, the networks were used to select the 20 cases in each region that were more likely to have a false clinical interpretation. These cases, together with 20 control cases in which the networks detected no likelihood of false clinical interpretation, were presented in random order to a group of three experienced physicians for a consensus re-interpretation; no information regarding clinical or neural network interpretations was provided to the re-evaluation panel. *Results*: The clinical interpretation and the re-evaluation differed in 53 of the 200 cases. Forty-six of the 53 cases (87%) came from the group selected by the neural networks, and only seven (13%) were control cases ($P < 0.001$). The disagreements between clinical routine interpretation by an experienced nuclear medicine expert and artificial networks were related to small and mild perfusion defects and localization of defects. *Conclusion*: The results demonstrate that artificial neural networks can identify those myocardial perfusion scintigrams that may have suboptimal image interpretations. This is a potentially highly cost-effective technique, which could be of great value, both in daily practice as a clinical decision support tool and as a tool in quality assurance.

Keywords Myocardial perfusion scintigraphy · Interpretation · Decision support system · Neural network · Myocardial ischaemia

Introduction

There is an increasing interest in measuring the quality of diagnostic imaging; cardiovascular imaging societies in the US have put this aspect at the top of their agendas [1]. Procedural guidelines are one way that

---

K. Tägil (✉) · L. Edenbrandt
Department of Clinical Sciences Malmö, Lund University
Research Program in Medical Informatics, Malmö
University Hospital, Malmo 205 02, Sweden
e-mail: kristina.tagil@med.lu.se

J. Marving · B. Hesse
Department of Clinical Physiology and Nuclear Medicine,
Rigshospitalet, Copenhagen, Denmark

M. Lomsky · L. Edenbrandt
Department of Clinical Physiology, Sahlgrenska
University Hospital, Gothenburg, Sweden

practice can be improved; the American Society of Nuclear Cardiology, the European Association of Nuclear Medicine, and the European Society of Cardiology have all established such documents for myocardial perfusion imaging [2, 3]. These guidelines, like the recently published documents on Appropriateness Criteria in cardiac imaging [4], generally focus on the equipment, the radiopharmaceutical and imaging protocols, and the image processing, and are much less specific or even silent regarding quality control of interpretation and reporting.

There is, therefore, a need to develop methods for quality control of routine work in the interpretation and reporting of diagnostic images. In myocardial perfusion scintigraphy (MPS), as in many other applications in nuclear medicine, verification of the true diagnosis is not available, so it is not possible to directly measure the accuracy of daily routine work.

To develop a tool that easily can help out with the quality control process is of major concern. Different expert system for the computer-assisted diagnosis of MPS has been developed during the last two decades. Rule-based expert systems for overall detection of coronary artery disease were presented by Garcia et al. [5]. To achieve better and more reliable results a pattern-recognition technique, would be preferable. Artificial intelligence using case-based reasoning and artificial neural networks are techniques that have proved to be of value in pattern recognition tasks and for the interpretation of MPS and in many cases outperforms visual diagnosis [6, 7].

The aim of this study was to explore the feasibility of using a technique based on artificial neural networks for quality assurance of image reporting. The networks were used to identify potentially suboptimal or erroneous interpretations of MPS images.

Materials and methodology

Patients

Participants were recruited consecutively from among patients undergoing a stress/rest MPS at Sahlgrenska University Hospital in Gothenburg during the 6 month period from September 15, 2004 to March 14, 2005. Only patients with a complete set of technically sufficient images were included. One stress/rest examination was included per patient. Of the 326 eligible patient studies, 10 were excluded because they could not be processed with the Exini heart software package (EXINI Diagnostics AB, Lund, Sweden). The final study group comprised 316 patients (153 men and 163 women), with a mean age of 62 years (range 21–88 years). All patients were referred to a MPS for the diagnosis (257) or management of coronary artery disease (57). Fifty-nine of the patients were smokers, prior myocardial infarction was found in 51 patients, chest pain in 233, family history of ischemic heart disease in 93, diabetes in 51, peripheral vascular disease in 18, hypertension in 158, hyperlipidemia in 144, prior percutanerous transluminal coronary angioplasty in 30 and prior coronary artery bypass graft surgery in 33 patients. The study was approved by the Research Ethics Committee at Gothenburg University. The patient material was also included in a previous validation of a gated-SPECT quantification algorithm [8].

Radionuclide imaging

The patients were stressed using either maximal exercise as a symptom-limited ergometry test (53%) or a vasodilator test with the patient at rest receiving adenosine infusion over 6 min (47%). The exercise or pharmacological stress was continued for at least 2 min after injection of the tracer. The stress and rest studies were performed in a 2-day protocol using 99mTc-sestamibi. Image acquisition began approximately 60 min after injection of 600 MBq 99mTc-sestamibi. Images were acquired with two different dual-head SPECT cameras (Infinia or Hawkeye, General Electric Medical Systems, Milwaukee, WI, USA) equipped with low energy, high resolution collimators. The planar projection images were acquired in the step and shoot mode using a circular rotation with the patient in a supine position. Acquisition was performed in a 64×64 matrix applying either 64 projections over 180 degrees for 40 s per projection or 68 projections over 204 degrees for 40 s per projection. In patients weighing over 90 kg, the acquisition time per projection was increased to 55 s. A zoom factor of 1.28 was used. The rest study was acquired in ECG-gated mode using 8 frames and an RR interval acceptance window of +20%. Tomographic reconstruction was performed using filtered back-projection (Butterworth filter with critical frequency 0.52 cycles/cm and order 5 in non-gated images, and with critical frequency 0.40 cycles/cm and order 10 in gated images). An automatic motion-correction program was applied where patient motion during acquisition was apparent [9]. No attenuation/scatter correction was used.

Artificial neural networks

The Exini heart software package (EXINI Diagnostics AB, Sweden) was used to process and display the MPS images for visual interpretation as well as automatically extract the image features used as input to the artificial neural networks. The image processing included active shape algorithms for creation of a heart shaped model, as well as algorithms for quantification of left ventricular volumes and regional perfusion at stress and rest [10]. Eight variables were calculated automatically from each of the five regions (anterior, septal, inferior, lateral, and apical); two from the gated images, and six from the summed rest and stress images. Artificial neural networks were used for the automated interpretation of the stress/rest myocardial perfusion. The automatically extracted image features from each of the 316 stress/rest studies were used as inputs to different neural networks for the interpretation of myocardial ischaemia in each of five myocardial regions. The clinical image interpretations of the studies (see below) were used as the desired interpretations during the training of the neural networks.

The neural networks had a multilayer perception architecture with one input layer, one hidden layer, and one output layer. A general presentation of neural networks may be found in the work of Bishop [11]. The neural networks were trained using a back-propagation algorithm [12], and a weight elimination technique was used to avoid "over-training" [13]. The input layer contained one unit for each of the eight input variables, while the hidden layer contained between two and six units, depending on the myocardial segment that was being trained. The output layer consisted of a single unit that encoded the probability of myocardial ischaemia with a continuous value between 0 and 1. During training, the output was set to 0, representing no myocardial ischaemia, and 1, representing myocardial ischaemia.

A six-fold cross validation procedure was used in order to obtain reliable performance. In this procedure, the data set was randomly divided into six equal parts; training was performed on five parts, while the remaining part was used for testing. The procedure was repeated six times, so that each part was used once as the test set. To make the training even more reliable not only 6 networks but an ensemble of networks was used. In this ensemble 30 separate neural networks were trained for each task and the mean of these 30 networks was used as the network ensemble output. This procedure ensured that patient studies used for network testing were not used for training of that particular network model. When training and testing the network it showed a performance resulting in a average sensitivity and specificity of 83 and 78%, respectively.

Image interpretations

Clinical interpretations of the stress/rest MPS images were performed at the time of reporting the studies by one nuclear medicine physician with long experience in nuclear cardiology. In the clinical interpretation 90 patients out of the 316 (28%) were regarded as having coronary artery disease.

Interpretations were based on complete sets of images (raw projection data, short axis and vertical and horizontal long axes slice images, polar plots, and quantification results from ECToolbox Cequal quantification), and supported by patient history, risk factors, results from earlier examinations, data from the stress test, and, if present, coronary angiography. Five myocardial regions were used for localization of reversible defects, as described above. The clinical interpretations were coded into a data file and used as the desired interpretation during training of the neural networks.

A subset of all patient studies, selected by the networks as described below, was re-evaluated as consensus reading by three experienced nuclear medicine physicians all with many years of nuclear cardiology experience. Their interpretation regarding the presence of ischaemia was based only on image data, that is, without any additional information other than patient age and gender, and with no knowledge of either the clinical interpretation or the neural network results. The analysis also included assessment of technical image quality, for example, the presence of high extracardiac activity, using reconstructed images (short and long axis slices). The

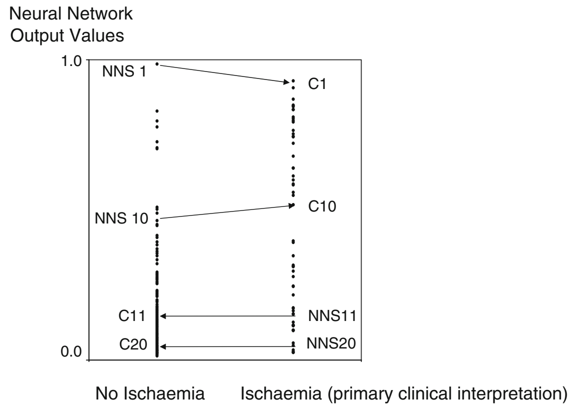


Fig. 1 The procedure for selection of images to the visual re-evaluation in the inferior myocardial segment. The 10 cases with highest NN output values in the no ischaemia group were selected; NNSelected = NNS 1–10. The 10 cases with lowest NN output values in the ischaemia group were selected; NN Selected = NNS 11–20. A group of 20 control cases was obtained by for each of the NNS 1–20 cases selecting the case with neural network output value closest to each of the previously selected cases. The control cases were selected from the opposite group (ischaemia vs. no ischaemia) as the NN selected cases

selection of cases for this re-evaluation was based on the neural network output values: 20 cases in each of the five myocardial regions were selected as those more likely to have a false clinical interpretation regarding ischaemia (defined as a reversible perfusion defect). These cases comprised the 10 cases (or regions) among those with a clinical interpretation of ischaemia that had the lowest neural network output values (i.e. neural networks suggested no ischemia), together with the 10 cases among those with a clinical interpretation of no ischaemia that had the highest neural network output values (i.e. neural networks suggested ischaemia). These 20 cases were defined as "neural network cases". In order to blind the reevaluation procedure, 20 control cases were selected for each of the five regions. The control cases were those with neural network output values closest to each of the "neural network cases", but in agreement with the primary clinical interpretation of ischaemia or no ischaemia, respectively (Fig. 1).

Statistical methods

The differences in binary proportions regarding number of cases with different interpretation between the clinical interpretation and the re-evaluation were tested using Fisher's exact test. *P*-values less than 0.05 were considered statistically significant.

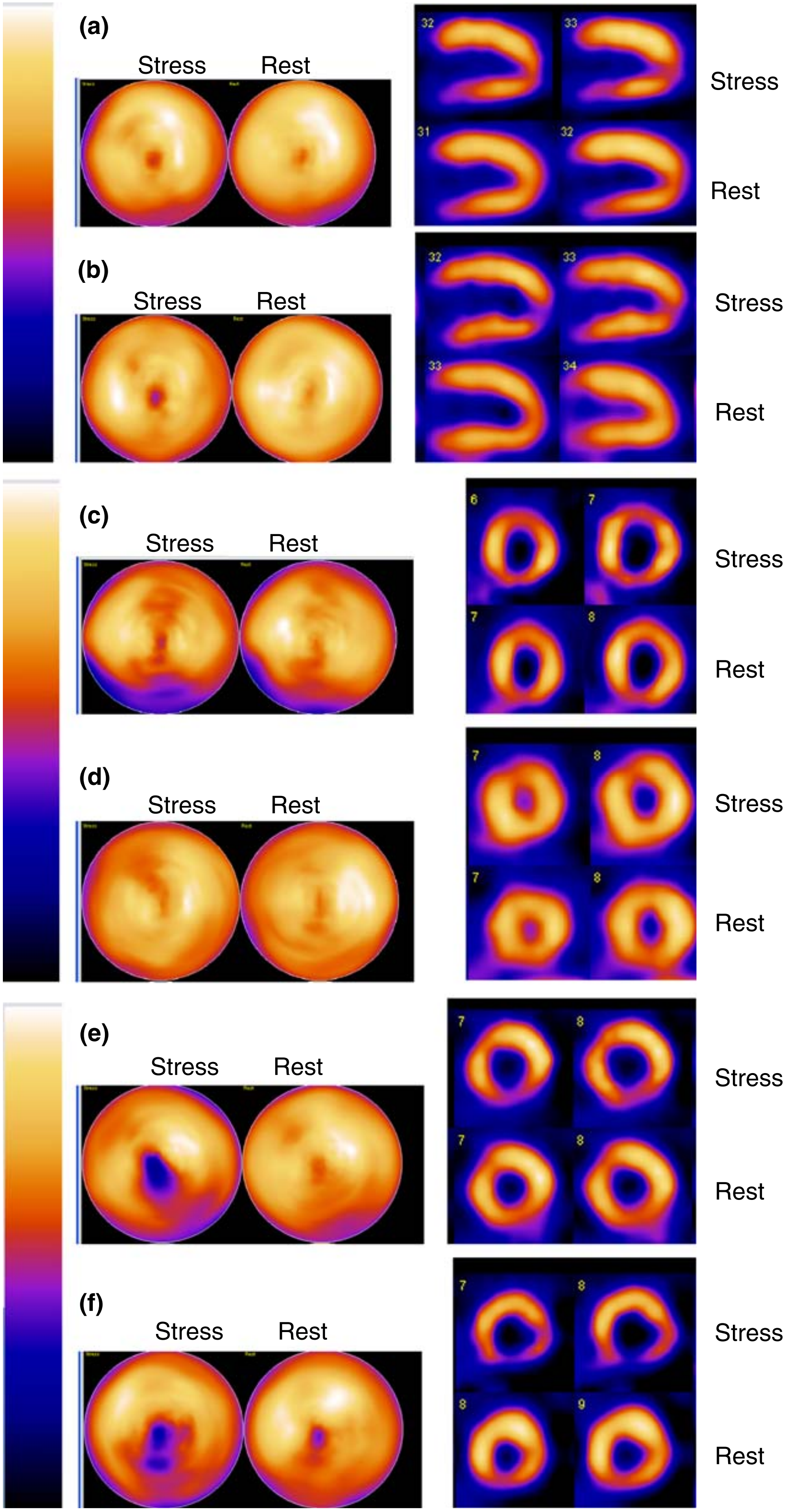
Results

The clinical interpretation and the re-evaluation of the 200 segments differed in 53 cases (coming from 46 patients) (Table 1). Of these 53 discrepant cases, 46 (87%) belonged to the group selected by the neural networks and only seven (13%) were control cases ($P < 0.001$). Disagreements between the interpretations involved all myocardial regions, but were most common in the apical region and occurred most rarely in the lateral wall, where the difference did not attain statistical significance ($P = 0.053$). The disagreements related either to small and mild perfusion defects, which are difficult to interpret or, in case of more obvious defects, to the exact localization in one or two myocardial segments. It can be difficult to decide whether a defect is best described as one localized in the anterior segment or as an anterior and apical defect. No clearly false interpretation was found; that is, the quality of the clinical interpretations was good. Figure 2 shows rest and stress images from six cases with differing interpretations at the clinical interpretation and the re-evaluation regarding

Table 1 Result of the visual re-evaluation

Myocardial region	Technically deficient	Disagreement between 1st interpretation Re-evaluation			*n*
		NN selected	Control cases	*P*	
Anterior	1	11/19	2/20	0.002	40
Septal	2	8/19	0/19	0.003	40
Inferior	1	9/19	2/20	0.01	40
Lateral	3	5/20	0/17	0.053	40
Apical	3	13/17	3/20	<0.001	40
Total	10	46/94	7/96	<0.001	200

Fig. 2 Polar plots from 6 patients, 5 neural network selected patients and one control patient (2D) with different interpretations at the first visual interpretation and the re-evaluation



presence and absence of ischemia Fig. 2a–d but also extent of ischemia Fig. 2e and f. Five of these six cases were selected by the neural networks and one was a control case (Fig. 2d). See Fig. 2.

In 33 cases, the clinical interpretation was that ischaemia was present, while the re-evaluation concluded that there was no ischaemia. Only 6 cases differed in the opposite direction. The re-evaluation concluded that the images were technically deficient in 10 of the 200 cases (seven patients, three of them contributing images from two different myocardial regions). Five patients had high extracardiac activity, and in two patients processing was not correctly performed by the Exini software program. Six of the 10 cases were selected by the neural network and the remaining four were control cases.

Discussion

The present study demonstrates that the artificial neural network based technique can be used as an effective tool to identify those MPS images more likely to have a false clinical interpretation regarding ischaemia, even minor uncertainties. The consensus re-evaluation by the three physicians disagreed with the routine clinical interpretation 6–7 times more often in cases where the neural networks disagreed with the primary clinical interpretation than in the control cases where the neural networks and primary clinical interpretation agreed. The cases where different classifications were made at the clinical interpretation and the reevaluation at least partly reflect inter-observer variability. Lindahl et al. [14] reported a considerably higher variability of 28% for three physicians who independently classified myocardial perfusion images using a four-grade scale. A binary classification was used in the present study, thus reducing the number of possible discrepant classifications. It should be remembered that these disagreements were present between interpretations by very experienced nuclear cardiologists. The existence of disagreement does not necessarily mean that the clinical routine evaluation was false and the re-evaluation true—nor vice versa. It only proves that they did not agree. When disagreement occurred, it was in relation to either small and mild perfusion defects or to the localization of defects. There were no obviously erroneous interpretations related to large and severe perfusion defects, indicating that in spite of the fact that the quality of the clinical interpretations, performed by an experienced nuclear medicine expert, was good, the more likely erroneous interpretations were detected.

The disagreements could to some degree be explained by the fact that clinical data was available at the clinical interpretation, while the re-evaluation was blinded to clinical data. In the majority (38/53) of cases with disagreement, the clinical interpretation was that ischaemia was present, while the re-evaluation concluded that there was no ischaemia. In some cases, the presence of additional clinical data may lead to more correct interpretation, while in other cases it may lead to the opposite; for example, in studies where the result of visual assessment lies between intermediate and high likelihood for ischaemia, the addition of pre-test likelihood (patient history, risk factors, exercise ECG, etc.) is probably helpful in arriving at the correct diagnosis. Discrepant interpretations included different estimations of the severity of a perfusion defect, as well as uncertainty over the extent or exact localization of a defect, for example in cases with an anterior or an antero-apical defect.

The control cases were selected based on the neural network output values, and not on the basis of similarity in appearance of the images to a experienced observer. This kind of selection process could result in control cases that are more or less easy to classify compared with the typical case in the database not selected by the neural network. In this study, no clear mistakes were found; that is, no cases with very high or very low output values were selected by the neural network. The result of this was that there were no control cases with very high or very low output values, and so the control cases were probably more of borderline type than clear cut, easily interpretable cases. The seven cases of disagreement in the control group would probably have been fewer if more of these easy cases had been selected.

Technical problems could also be detected with the neural network technique. One group of discrepant cases included technically suboptimal scintigraphic studies, a well-known factor for misinterpretation [15]. Another group could be correctly interpreted but falsely reported due to administrative mistakes.

Clinical implications

In large centres, especially academic ones, it is often the case that two or more physicians are available for a discussion regarding interpretation of more difficult patient studies. However, in the daily clinical setting at busy small laboratories or private cardiology clinics, interpretation of myocardial SPECT images is frequently based on the interpretation of only one physician [16].

It has been considered a strength in clinical medicine to obtain a second opinion where necessary. Compared to humans, a computer yields more precise and standardized interpretations [17], although these interpretations are not necessarily more accurate. Computer interpretation will depend on many factors, including an optimal gold standard for the computer training. Optimal quality is therefore required of the databases on which the decision support systems and subsequent interpretation algorithms are based. By using the neural network to identify possible suboptimal or erroneous interpretations in the database, these cases can be checked in order to assure high quality of the data. In our study, 100 segments were selected out of a total of 1,580 segments, reducing the number of cases where double checking was necessary.

Limitations of the study

The processed images used at the clinical interpretation and the re-evaluation were nearly identical, but there were slight differences due to variations in the image processing of the ECToolbox, Cequal, and Exini software programs, respectively. A second confounding factor is related to the lack of clinical information other than age and gender at the visual reevaluation, compared to the extensive clinical information—or bias—at the clinical interpretation. This different level of pre-test knowledge probably had some influence on the interpretation, as discussed above.

Conclusions

This study demonstrates that a neural network can be used to identify scintigrams with potentially erroneous image interpretations. This allows re-evaluation efforts to be focused, resulting in a better cost-benefit ratio. The technique used might also highly relevant in many other fields of imaging where correct visual interpretation is based on long training.

Acknowledgements This study was supported by a scholarship from Medicon Valley Association. Conflicts of interests: Kristina Tägil and Lars Edenbrandt are shareholders in EXINI Diagnostics AB, which owns the Exini heart software used in the study.

References

1. Iglehart JK (2006) The new era of medical imaging—progress and pitfalls. N Engl J Med 354:2822–2828
2. Hesse B, Tagil K, Cuocolo A, Anagnostopoulos C, Bardies M, Bax J et al (2005) EANM/ESC Group EANM/ESC procedural guidelines for myocardial perfusion imaging in nuclear cardiology. Eur J Nucl Med Mol Imaging 32: 855–897
3. Brindis RG, Douglas PS, Hendel RC, Peterson ED, Wolk MJ, Allen JM et al, American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group; American Society of Nuclear Cardiology; American Heart Association. ACCF/ASNC appropriateness criteria for single-photon emission computed tomography myocardial perfusion imaging (SPECT MPI): a report of the American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group and the American Society of Nuclear Cardiology endorsed by the American Heart Association (2005) J Am Coll Cardiol 46:1587–1605. Review. Erratum in: J Am Coll Cardiol 2005, 46:2148–2150
4. Patel MR, Spertus JA, Brindis RG, Hendel RC, Douglas PS, Peterson ED et al (2005) ACCF proposed method for evaluating the appropriateness of cardiovascular imaging. J Am Coll Cardiol 46:1606–1613
5. Garcia EV, David Cooke C, Folks RD, Santana CA, Krawczynska EG, De Braal L, Ezquerra NF (2001) Diagnostic performance of an expert system for the interpretation of myocardial perfusion SPECT studies. J Nucl Med 42:1185–1191
6. Bagher-Ebadian H, Soltanian-Zadeh H, Senior Member, IEEE, Setayeshi S, Smith ST (2004) Neural network and fuzzy clustering approach for automatic diagnosis of coronary artery disease in nuclear medicine. IEEE Trans Nucl Sci 51:184–192
7. Datz FL, Rosenberg C, Gabor FV, Christian PE, Gullberg GT, Ahluwalia R, Morton KA (1993) The use of computer-assisted diagnosis in cardiac perfusion nuclear medicine studies: a review (Part 3). J Digit Imaging 6(2):67–80
8. Lomsky M, Richter J, Johansson L, Høilund-Carlsen PF, Edenbrandt L (2006) Validation of a new automated method for analysis of gated-SPECT images. Clin Physiol Funct Imaging 26:139–145
9. Matsumoto N, Berman DS, Kavanagh PB, Gerlach J, Hayes SW, Lewin HC et al (2001) Quantitative assessment

of motion artifacts and validation of a new motion-correction – program for myocardial perfusion SPECT. J Nucl Med 42:687–694

10. Lomsky M, Richter J, Johansson L, El-Ali H, Åström K, Ljungberg M et al (2005) A new automated method for analysis of gated-SPECT images based on a 3-dimensional heart shaped model. Clin Physiol Funct Imaging 25:234–240
11. Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press
12. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representation by errorpropagation. In: Rumelhart DE, McCleland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition. MIT Press/Bradford Books, Cambridge, Mass, pp 318–363
13. Hanson SJ, Pratt LY (1989) Comparing biases for minimal network construction with backpropagation. In: Touretzky DS (ed) Advances in neural informationprocessing systems. San Mateo, CA: Morgan Kaufmann, p 17185.
14. Lindahl D, Lundin A, Palmer J, Edenbrandt L (1999) Improved classification of myocardial bull's-eye scintigrams with computer-based decision support system. J Nucl Med 40:96–101
15. Piepsz A, Tondeur M, Ham H, De Palma D, Roca I (2007) Interobserver reproducibility in reporting on Tc-99m DMSA scintigraphy. A wide collaborative study. ISCORN 2007 symposium radionuclides in nephro-urology
16. Kelion AD, Anagnostopoulos C, Harbinson M, Underwood SR, Metcalfe M, British Nuclear Cardiology Society (2000) Myocardial perfusion scintigraphy in the UK: insights from the British Nuclear Cardiology Society Survey 2000. Heart 91(Suppl 4):iv2–iv12
17. Fletcher BD, Glicksman AS, Gieser P (1999) Interobserver variability in the detection of cervical-thoracic Hodgkin's disease by computed tomography. J Clin Oncol 17:2153–2159