Diagnostic Performance of an Expert System for the Interpretation of Myocardial Perfusion SPECT Studies

Ernest V. Garcia, C. David Cooke, Russell D. Folks, Cesar A. Santana, Elzbieta G. Krawczynska, Levien De Braal, and Norberto F. Ezquerra

Emory University School of Medicine, Atlanta; and Georgia Institute of Technology, Atlanta, Georgia

An expert system (PERFEX) developed for the computer-assisted interpretation of myocardial perfusion SPECT studies is now becoming widely available. To date, a systematic validation of the diagnostic performance of this expert system for the interpretation of myocardial perfusion SPECT studies has not been reported. Methods: To validate PERFEX’s ability to detect and locate coronary artery disease (CAD), we analyzed 655 stress/rest myocardial perfusion prospective SPECT studies in patients who also underwent coronary angiography. The patient population comprised CAD patients (n = 480) and healthy volunteers (n = 175) (449 men, 206 women). Data from 461 other patient studies were used to implement and refine 253 heuristic rules that best correlated the presence and location of left ventricular myocardial perfusion defects on SPECT studies with angiographically detected CAD and with human expert visual interpretations. Myocardial perfusion defects were automatically identified as segments with counts below sex-matched normal limits. PERFEX uses the certainty of the location, size, shape, and reversibility of the perfusion defects to infer the certainty of the presence and location of CAD. The visual interpretations of tomograms and polar maps, vessel stenosis from coronary angiography, and PERFEX interpretations were all accessed automatically from databases and were used to automatically generate comparisons between diagnostic approaches. Results: Using the physician’s reading as a gold standard, PERFEX’s sensitivity and specificity levels for detection and localization of disease were, respectively, 83% and 73% for CAD, 76% and 66% for the left anterior descending artery, 90% and 70% for the left circumflex artery, and 74% and 79% for the right coronary artery. These results were extracted from a receiver operating characteristic curve using the average optimal input certainty factor. Conclusion: This study shows that the diagnostic performance of PERFEX for interpreting myocardial perfusion SPECT studies is comparable with that of nuclear medicine experts in detecting and locating CAD.

Key Words: expert systems; artificial intelligence; myocardial perfusion SPECT; quantitative analysis


Received Sep. 6, 2000; revision accepted Jan. 4, 2001.
For correspondence or reprints contact: Ernest V. Garcia, PhD, Emory University Hospital, Room E163, 1364 Clifton Rd. NE, Atlanta, GA 30322.

The use of myocardial perfusion SPECT imaging for the assessment of coronary artery disease (CAD) continues to grow at an unprecedented rate. From 1996 to 1997, procedure volume in the United States increased 17%, with more than 5 million studies performed (1). With the aging of the baby boom generation, an even higher demand for these studies is expected in the next 20 y. Because this increased demand could exceed the supply of expert diagnosticians who perform these studies, these expert diagnosticians may have less time to interpret each study. Increased demand could also lead to the emergence of an increasing number of diagnosticians with limited expertise. Both of these outcomes could result in an increase in interobserver variability and a reduction in accuracy. Hence, it is desirable that tools be implemented to assist physicians in interpreting studies at a faster rate, at a higher level of expertise, or both. Because computers are a necessary part of acquiring and processing these studies, it is reasonable to expect that these tools should be computer based.

Over the past several years, artificial intelligence methods have been investigated as a way to develop such a tool. Examples include neural networks (2–4) and case-based reasoning techniques (5) to provide computer-assisted diagnosis of myocardial perfusion planar and SPECT studies. Most of these approaches have used the polar map output of well-established database myocardial perfusion quantification programs (6–8) as input to the decision-making process. In the artificial neural network approach, the concept is to try to emulate how human neurons perform pattern recognition tasks. Repeated recognition trials are run using sample perfusion data as input and using corresponding coronary angiography results as output to modify the strength between the input and output nodes. In this manner, the network is trained and the input data eventually predict the output. In the case-based reasoning approach, the algorithm searches a library of patient cases to find the ones that best match those of the patient study being analyzed. The common findings from these cases, such as coronary angiography results, are then used to assist the diagnostican’s interpretation.
Another artificial intelligence approach that has been investigated for this purpose is the knowledge-based expert system. In expert systems, a knowledge base of heuristic rules is obtained from human experts, capturing how they make their interpretations. These rules are usually expressed in the form of “if/then” expressions. Expert systems have been investigated in nuclear medicine to assist in the interpretation of perfusion–ventilation lung studies (9) and hexamethylpropyleneamine oxime brain SPECT studies (10). Expert systems have also been used in cardiology for assessment of acute myocardial infarction from electrocardiography (ECG) analysis (11), for echocardiography analysis (12), and for the management of ventricular tachycardia (13).

Since 1985, we have been developing an expert system called PERFEX (an abbreviation of “perfusion expert”; Syntermed, Atlanta, GA) as a tool for the computer-assisted diagnosis of stress/rest myocardial perfusion SPECT studies (14,15). The purpose of this study was to investigate how computer-assisted interpretations suggested by PERFEX compare with those of human experts. These investigations were designed to show the overall performance of PERFEX before its dissemination and distribution. We have chosen 2 sets of gold standards: the interpretation by human experts and the results of coronary angiography studies. Although the goal of expert systems is to match the interpretation of human experts, we used the results of coronary angiography to resolve differences between the expert system and the human experts.

**MATERIALS AND METHODS**

**Patients**

All studies used for this retrospective evaluation were obtained from the cardiac database of patients referred to our nuclear medicine service for myocardial perfusion SPECT imaging from September 1989 to February 1997. Patients were selected who had undergone any of the following conventional myocardial perfusion SPECT protocols (16): stress/redistribution $^{201}$TI ($n = 376$), low-dose rest/high-dose stress $^{99m}$Tc-sestamibi ($n = 138$), or rest $^{201}$TI/stress $^{99m}$Tc-sestamibi ($n = 141$). Stress was performed using exercise ($n = 419$), dipyridamole ($n = 221$), adenosine ($n = 3$), or dobutamine ($n = 12$). These consecutive patients were selected because they had undergone both myocardial perfusion SPECT imaging and coronary angiography within 2 wk of each other with no intervention between the 2 imaging procedures. Patients who had previous coronary artery bypass surgery were excluded. Patients were also excluded because of data retrieval problems or incomplete SPECT studies. This selection process resulted in 655 patients without intervention between the 2 imaging procedures. Patients who were assigned to 1 or more vascular territory combinations: left anterior descending artery (LAD), left circumflex artery (LCX), right coronary artery (RCA), LAD or LCX, LAD or RCA, or LCX or RCA. Once a region was determined to be hypoperfused, it was assigned to the territory in which the majority of the region fell. If a defect or reversibility region fell between 2 territories, it was assigned to the “or” of the 2 territories, as was done in previous studies (17).

Independent diagnosis of CAD was based on coronary angiography during cardiac catheterization. Diagnosis of CAD was based on 1 or more of the major coronary vessels’ having at least 1 stenosis with $\geq 50\%$ luminal narrowing or diffuse disease. Luminal narrowing and diffuse disease were qualitatively assessed by an experienced attending cardiologist. The application of these criteria to the 655-patient population resulted in 175 patients without CAD and 480 patients with CAD. The breakdown of disease by vascular territory in the 480 CAD patients was as follows: 346 LAD, 256 LCX, and 281 RCA. These included 194 patients with single-vessel disease, 169 with double-vessel disease, and 117 with triple-vessel disease.

**Data Analysis and Expert System Interpretation**

All SPECT patient studies were reconstructed and reoriented into oblique-axis tomograms using conventional techniques (16). The studies were then submitted to a well-established method of database quantification (18). This method identified hypoperfused regions as those with normalized count distributions falling below a predetermined number of SDs below the mean pattern of sex-matched normal response for the specific myocardial perfusion SPECT protocol used. The program also identified ischemic regions as stress perfusion defects that improve at rest. This was determined quantitatively as a predetermined number of SDs above the mean normalized difference between the stress and rest distributions.

An automatic feature-extraction program then described the location and severity of each defect and corresponding reversibility (19). The location was expressed in the form of 32 possible descriptors (20). These descriptors were defined as coordinates of both depth (basal, medial, distal apical, and proximal apical) and angular location (8 subsets of the septal, inferior, lateral, and anterior myocardial walls). The severity was expressed in terms of certainty factors ranging from $-1$ to $+1$ for the pixel in that descriptor with the most severe finding ($-1$ means there is definitely no disease, $+1$ means there is definitely disease, and the range from $-0.2$ to $+0.2$ means the presence of disease is equivocal or indeterminate). Certainty factors are heuristically defined numeric estimates of evidence for or against a particular hypothesis. The certainty factor model is a well-known approach to uncertainty reasoning used in artificial intelligence (21). Initially, the certainty factor of each abnormal descriptor was allowed to vary between 0.2 and 0.99 in linear proportion to the number of SDs below the mean normal response. In this representation, a certainty factor of $+0.2$ corresponds to the threshold (in number of SDs below the mean) for just detecting disease and $+0.99$ corresponds to $\approx 8$ SDs below the mean (very sure that findings are abnormal). Descriptors with all pixels above the normal response were set to a certainty factor of $-1$.

The description of how PERFEX works has been provided in detail elsewhere (14,15,20). The architecture of PERFEX was inspired by that of MYCIN (21), a pioneering rule-based expert system developed in the 1970s to assist physicians with the decisions involved in the selection of appropriate therapy for patients with infections. To create the PERFEX knowledge base, a study
RESULTS

A plot of percentage agreement (accuracy) between PERFEX and the human experts as a function of CF shift level showed that the 0.10 level generated the best overall agreements for localizing disease to the LAD and RCA vascular territories (Fig. 1). The 0.20 level generated the best overall agreement for the LCX territory (Fig. 1). The plot for agreement in the detection of CAD is not shown because that factor is dependent on the certainty of disease in each of the 3 vascular territories. ROCs corresponding to the agreement of localizing disease to these vascular territories provide the sensitivity and specificity for each of the CF shift levels using the expert as the gold standard (Fig. 2) or using coronary angiography as the gold standard (Fig. 3).

Using the 0.15 CF shift level as an average optimal input certainty factor, we obtained the sensitivity and specificity results shown in Figure 2. The sensitivity ranged from a high of 90% for localization in the LCX vascular territory to a low of 74% for the RCA region. The sensitivities for overall detection of CAD and for localization in the LAD region were 83% and 77%, respectively. The specificity ranged from a high of 79% for the RCA vascular territory to a low of 66% for the LAD region. The specificities for overall detection of the absence of CAD and for localization in the LCX region were 73% and 70%, respectively.

Using coronary angiography as the gold standard, analysis of the PERFEX results at the 0.15 CF shift level generated the following statistics (Fig. 3). The sensitivity ranged from a high of 80% for detection of CAD to a low of 65% for the RCA region. The sensitivities for localization in the LAD vascular territory and the LCX region were 69% and 68%, respectively. The specificity ranged from a high of 65% for the RCA vascular territory to a low of 42% for detecting the absence of CAD. The specificities for the LAD and LCX regions were 54% and 56%, respectively.

Analysis of the nuclear medicine physicians’ interpretation of the SPECT perfusion studies using coronary angiography...
raphy as the gold standard generated the following results (Figs. 3 and 4). The sensitivities for detection of CAD and for localization to LAD, LCX, and RCA vascular territories were 87%, 69%, 61%, and 73%, respectively. The specificities for detecting the absence of CAD overall and in the LAD, LCX, and RCA vascular territories were 21%, 59%, 88%, and 71%, respectively.

Comparison of the sensitivity and specificity of PERFEX (at the 0.15 CF shift level) versus those obtained from the expert reading by nuclear medicine physicians, both using coronary angiography as the gold standard, yielded the following results (Fig. 4). Statistically significant differences were obtained in the detection of the presence and absence of CAD and in the specificity of the LCX region.

FIGURE 2. ROC curves showing agreement between PERFEX and nuclear medicine physicians. ROC curves are plotted for overall detection of CAD (A) and for localization to LAD (B), LCX (C), and RCA (D) vascular territories. Points shown on ROC curves are generated from different CF shift levels. Level corresponding to 0.15 CF shift level was deemed to be average optimal input certainty factor from comparison of accuracy at different levels (Fig. 1).

FIGURE 3. ROC curves showing agreement between PERFEX and coronary angiography. ROC curves are plotted for overall detection of CAD (A) and for localization to LAD (B), LCX (C), and RCA (D) vascular territories. Points shown on ROC curves are generated from different CF shift levels. *Results obtained from nuclear medicine physicians using coronary angiography as gold standard.
these 3 categories, PERFEX obtained better results in 1 category (specificity for detecting the absence of CAD) and the human expert obtained better results in the other 2. There were no statistically significant differences in the RCA region or the LAD region or in the sensitivity of localizing LCX disease. These comparisons were also performed at the 0.10 and 0.20 CF shift levels (Fig. 4).

**DISCUSSION**

The primary goal of this study was to assess the performance of an expert system in detecting and localizing CAD. Because there is no widely accepted approach to evaluating the performance of such computer-based decision-making systems, a statistically defined series of tests that involved stringent demands was used. That is, the results obtained from PERFEX were compared with those of nuclear medicine expert readers and with independent results from coronary angiography. Sensitivity and specificity for detecting and localizing CAD in a large population of 655 patients were used as the primary criteria for comparison.

The results of this study show that PERFEX is almost as accurate as nuclear medicine expert readers in detecting and localizing CAD when coronary angiography is used as the gold standard. These results are remarkable considering that PERFEX reaches its conclusions in <3 s per patient. Furthermore, this version of the expert system does not use most of the relevant clinical and quality control information available to the diagnosticians. This lack of information may partly account for the differences between the expert system recommendations and those of the human experts when using the interpretation of the latter as the gold standard. This knowledge includes information on body habitus and level of tissue attenuation by breast muscles or diaphragm.

Another reason for the apparent superiority of experts over PERFEX for localizing disease to the LCX vascular territory is that the experts tended to assign disease to the posterolateral territory (LCX or RCA) much more frequently than did PERFEX. Experts used this category on 197 patients and PERFEX on only 3 (at the 0.15 CF shift level). PERFEX almost always assigned disease to either the LCX territory or the RCA territory. In our analysis, assignment to the posterolateral territory (LCX or RCA), rather than to the specific LCX territory or RCA territory, generated a better result because correct agreements with coronary angiography were counted whether the patient had LCX or RCA disease. When this analysis was repeated, giving PERFEX the same advantage as the experts, the results yielded no statistical difference between experts and PERFEX for the sensitivity (71% vs. 72%), specificity (57% vs. 55%), or accuracy (65% vs. 65%) in detecting posterolateral disease.

It is difficult to compare the results of PERFEX with those obtained from other artificial intelligence approaches that use artificial neural networks (2–4) or case-based reasoning (5). That is because the criteria used to determine diagnostic accuracy (sensitivity and specificity) are a function of the prevalence of disease and the referral bias of the population. These vary widely between study populations. Thus, the reported sensitivity and specificity of the expert system should be used for comparison with those of the human experts rather than as a measure of the accuracy of the diagnostic performance of the program for all patient groups. Our philosophy is that there is no optimal threshold to interpret all patient populations. In fact, in the present implementation of this program, physicians are allowed to use the expert system to interpret at different points along the ROC curve (Fig. 3), resulting in different sensitivity and specificity results depending on how aggressive or conservative an interpretation is desired.

We have preferred the expert system approach to the neural network approach because the heuristic rules that are used in expert systems to reach a conclusion may be traced and linked to each other, thereby providing a mechanism to justify or explain any conclusion reached. By contrast, neural networks do not provide justifications, although they are excellent for pattern recognition tasks. Moreover, neural network systems require a much larger training dataset than...
expert systems to converge on reasonable results for the same task.

There are equally compelling reasons for preferring knowledge-based approaches to case-based reasoning approaches. The main challenge is that a very large library of image cases would be required to create a sufficiently robust system, similar to the large data demands posed by neural network training. Also, accurate measures of image “similarity” would have to be developed to adapt the indexed cases to the patient case under consideration. Knowledge captured as rules, coupled with the uncertainty reasoning model used by PERFEX, seems to overcome these challenges while featuring a quick processing time.

There are 4 limitations to this study. First, all the data used for this evaluation were obtained retrospectively as part of the routine clinical evaluation of the patient and the routine database of report forms that were used for a research protocol. One would expect the entirety of the database to be <100% and the interpretations and assessments both from the myocardial perfusion SPECT study and the coronary angiography study to be less detailed and accurate than if they had been performed under a strict prospective research protocol. Moreover, because these retrospective data were interpreted so long ago, the processing was performed manually and not with the advantage of today’s automatic processing programs. Nevertheless, because the data used in this study reflect the true clinical information used to manage the patients, the results reported herein reflect the true effectiveness of the tests for this population. Second, there is also a referral bias in the routine work-up of the patient that accounts for, among other things, an apparent low specificity for detecting the absence of CAD (22). Third, although the interpretation by the nuclear medicine experts appears to be the ideal gold standard for assessing an expert system, we used coronary angiography to resolve the differences between the experts and PERFEX. The anatomic information extracted from coronary angiography studies is not expected to always coincide with the physiologic information obtained from myocardial perfusion studies. Fourth, although a main advantage of PERFEX over artificial neural networks is that it provides justifications for its conclusions, this function was not validated. We have yet to design an objective approach to validate this function.

We are continuing to investigate how to further improve the diagnostic performance of PERFEX. One approach has been to add heuristic rules that use the patients’ clinical information routinely available to physicians. This includes such information as whether the patient has left bundle branch block or had previous myocardial infarction, the results of stress ECG, the technical quality of the study, and information on body habitus. Another approach, as suggested by our results, is to use different CF shift levels for the different vascular territories. We are also investigating the use of data mining techniques to automatically find associations between the myocardial perfusion quantitative results, clinical variables, and angiographic results. These associations may be used as heuristic rules to enhance the expert system (23).

CONCLUSION

Automatic computer-assisted interpretation of myocardial perfusion SPECT studies by an expert system agrees well with the interpretations of expert nuclear medicine physicians and exhibits diagnostic accuracy consistent with that of the experts when coronary angiography is used as the gold standard.

ACKNOWLEDGMENTS

The authors thank the many diagnosticians who performed and interpreted the cardiac imaging procedures used in this study, and the staff of the Emory Cardiac Data Bank, who assisted in identifying the population used. This study was funded in part by National Library of Medicine grant LM06726. Some of the authors (Ernest V. Garcia, C. David Cooke, and Russell D. Folks) receive royalties from the sale of the application software related to the research described in this article. The terms of this arrangement have been reviewed and approved by Emory University in accordance with its conflict-of-interest practice.

REFERENCES

1. Nuclear medicine procedures up 7%; cardiac perfusion studies up 17% [press release]. Des Plaines, IL: Technology Marketing Group; September 1, 1998.


Diagnostic Performance of an Expert System for the Interpretation of Myocardial Perfusion SPECT Studies

Ernest V. Garcia, C. David Cooke, Russell D. Folks, Cesar A. Santana, Elzbieta G. Krawczynska, Levien De Braal and Norberto F. Ezquerra