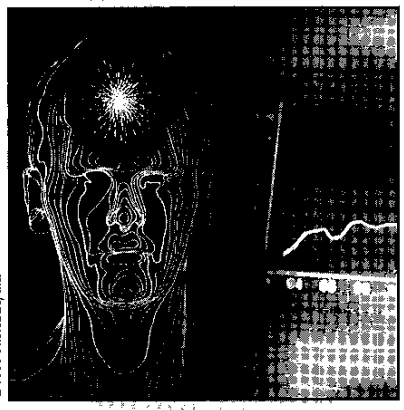


Issues in Automating Cardiac SPECT Diagnosis

Dealing with the Computational Complexity Involved in Knowledge Discovery when Working with Images

Jaroslaw P. Sacha[1], Krzysztof J. Cios[1], Lucy S. Goodenday[2]
[1]University of Toledo, Ohio
[2]Medical College of Ohio, Toledo

This article addresses the problem of automating diagnostic imaging, using the approach of data mining and knowledge discovery. We used information from cardiac single-photon emission computed tomography (SPECT) to diagnose patterns of myocardial perfusion. We started with the creation of a database consisting of 613 patient SPECT studies collected under conditions of rest and stress, which provided more than 90,000 two-dimensional (2-D) images. The general knowledge discovery process then proceeded through data preparation and image processing, discovery of patterns in the data, and interpretation.

Overview

The goal of the knowledge discovery process is to reveal some new and useful information from the data. In our case, the data are cardiac SPECT images. The knowledge we want to reveal is in the form of diagnostic rules, so that in the future a computer can semi-automatically diagnose patients.

Our data consisted of primary data (with no influence on its collection) in the form of patient records and "raw" images. Since the data were not suitable for direct use, we created secondary data by refactoring patient records and by extracting features from raw images. This re-creation of data relates to creating a nested knowledge discovery process structure. Careful preprocessing of the initial data is essential to ensure that useful knowledge could be derived from the data.

Knowledge discovery is a highly iterative process; it is often necessary to repeatedly backtrack to previous tasks and repeat certain actions. Some of these tasks may be mini knowledge discovery processes themselves. A fundamental task for a knowledge discovery process is creation of a database and efficient methods of access to the contained data. This database contains the initial data and data generated at each step of the knowledge discovery process. Our database, besides structured categorical and numerical information, contains other modes of data, in particular 2-D and three-dimensional (3-D) images. This type of a database is very typical in health-care applications, where test results may have a large variety of formats such as a category number, a 2-D image or signal time sequence (e.g., EEG, ECG), a 3-D SPECT or CT image, or four or more dimensional data like MRI or gated SPECT.

For humans, working with images is an easier task than working with large quantities of numbers. This is just the opposite for computers. Humans excel in qualitative tasks, computers in quantitative. Automatic discovery of information in images is a complex task. It is very difficult to model the way humans process images, primarily because mental operations to interpret images (finding objects and assessing their features) are mostly performed below the threshold of consciousness [1]. On the other hand, image operations are computationally very intensive. It is difficult to have, for example, production rules triggered by raw images, because images have complex secondary characteristics, where extraction of desired characteristics is not well defined. In the approach that is presented in this article, we try to balance computational complexity and mimicking human experts in interpreting the images.

The goal of this investigation is to use the data mining and knowledge discovery approach to semi-automate the process of using cardiac SPECT images for diagnosis. The knowledge discovery process is still more of an art than science, since there is no theory that can guide those involved in data mining activities toward

 0739-5175/00/$10.00©2000IEEE July/August 2000

discovering new and useful knowledge. Because of this, researchers have resorted to heuristic approaches, outlining a series of steps to be followed in a general discovery process. In this work, we will follow the six-step knowledge discovery process, consisting of: Understanding the Problem Domain, Understanding the Data, Preparation of the Data, Data Mining, Evaluation of the Discovered Knowledge, and Using the Discovered Knowledge [2, 3]. Below, we describe in detail all issues involved in each of steps one through five. Before the discovered knowledge can be used, however, it needs to undergo clinical testing; such testing is not a part of this work.

Understanding the Medical Problem Domain

The goal of this step is to obtain a clear understanding of the medical problem domain and determine boundaries within which to operate. In our case, the intended user of the final system will be a cardiologist, to whom it will serve as a second-opinion agent. Other clients will be cardiology fellows and bioengineering students working on related projects. However, in this project, we concentrate mainly on the building blocks of such a system, since it will take some time before the final system is ready for testing.

Determination of Medical Objectives

The ultimate goal is to semi-automate evaluation of cardiac SPECT images. Information will be delivered in the form of an advisory system to a cardiologist, so that the diagnostic process can be enhanced, made more consistent, and easier. To achieve this main goal, we need to solve several subgoals. First, we need to organize patient SPECT data and diagnostic results into an easily accessible form of relational database.

Assessment of the Current Situation

Cardiac SPECT provides a clinician with a set of 3-D images visualizing distribution of radioactive counts within the myocardium and surrounding structures [4]. Images represent radioactive count densities within the heart muscle that are proportional to muscle perfusion; in particular of the left ventricle (LV), which is normally thicker than other cardiac structures. Two studies are performed after a patient is injected with a tracer, one at rest (rest image) and one after injection during maximal stress (stress image). The studies are represented by two 3-D density images. Clinicians compare the two images in order to detect abnormalities in the distribution of blood flow within the left ventricular myocardium.

Visualization of the SPECT images is complicated by the fact that 3-D density images cannot be directly presented using contemporary display devices that produce 2-D pictures; some kind of transformation has to be performed. This transformation introduces reduction of information. There are two alternatives: 2-D density images or 3-D surface rendering [5]. The first preserves most of the intensity information; however, 3-D relations are only implicit. The second alternative makes 3-D information explicit; however, density is represented indirectly through the shape of the 3-D surface and/or its color [6, 7].

Typically, the LV is visualized as a series of 2-D intensity slices. When sliced perpendicular to the long axis of the LV, the view is termed "short axis." Slices parallel to the long axis of the LV are called "vertical long axis," and "horizontal long axis" views (see Fig. 1). 3-D relations are only implicit in the views; it is left to the interpreting physician to mentally reconstruct them as a 3-D object.

Another family of 2-D visualization methods is based on projections in non-Cartesian coordinate systems. The 3-D LV is *unwrapped* on a 2-D plane by radial projection into spherical coordinates [8], or a combination of spherical and cylindrical coordinates [9]. These are generally referred to as *bull's-eye* methods, since they produce pairs, rest/stress, of round images (see Fig. 6).

A number of 3-D surface-rendering methods exist. They are occasionally used in association with *gated blood-pool SPECT* [10] to produce motion sequences of left ventricular function. This is a very interesting research area; however, in this article, we concentrate on static SPECT images of myocardial perfusion rather than chamber geometry.

A number of techniques, most of them concerned with visualization, have been developed to aid in the classification of the images. However, it has been shown that interpretation of images by strictly visual techniques is fraught with error and inconsistency [11]. For this reason, assistance in diagnosis has been sought through the use of computer-derived image display and quantitation. Such quantitation has demonstrably decreased the variability in image interpretation [12].

One of the few examples of automatic interpretation of SPECT images is the PERFEX expert system [13]. This system infers the extent and severity of coronary artery disease from perfusion distribution.

To summarize, we observe that in most cases:

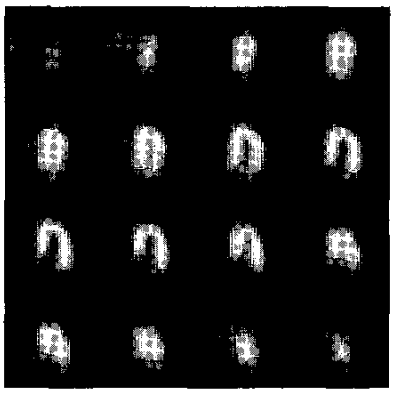
- Diagnosis is performed visually by a physician.
- Most methods are highly interactive and subjective (the diagnosis varies greatly from one physician to another).
- Available software packages are expensive; e.g., proprietary 3-D quantitative programs such as CEQUAL and QGS.
- There are no feature-extraction algorithm libraries available for SPECT images.

Data Access

All patient data available for the project were originally stored in a single Microsoft Excel file with 184 fields and 4275 records. Each record corresponds to a single visit—a SPECT procedure. Data include personal patient information such as age, sex, height; information about the procedure; and the nuclear cardiologist's interpretation of the SPECT images [by regions of interest (ROIs)] and perfusion classification (i.e., diagnosis).

There were 6817 SPECT image files available (after cleaning), corresponding to about 613 cases. Each set was stored in a separate directory, the name of which was a combination of the patient's hospital number and the date of study. The following information was available:

- Patient identification number.
- Raw data for rest and stress study.



1. Horizontal long-axis view of the left ventricle.

- 3-D images corresponding to short, vertical long, and horizontal long axis views of the heart.

The format of image files had to be determined by reverse engineering. Image files were spread between different hard drives and archive media. Not all directories contained complete data sets. Combined image data sets occupied over 2 Gbytes of disk space. Due to cost and licensing issues, software for manipulation and visualization of SPECT images had to be written from scratch.

Determination of Data Mining Objectives

The main data mining objective was to come up with diagnostic rules that can be used for evaluation of cardiac SPECT images. The selected data mining algorithms should be able to generate production rules that are understandable to the user. Due to anatomical differences, separate models for males and females are needed. The input variables were features extracted from SPECT images and some variables from patients' records, such as age, weight, height, smoker/nonsmoker, etc.

Extraction of features from SPECT images was probably the most complex part of this project. First, we built a model of a normal LV and then used it for image registration, segmentation, and feature extraction. (Registration, in image processing, refers to a process of scaling, shifting, and rotation of an image so that an object of interest on that image can be transformed to some desired pose.) A number of feature-extraction approaches were explored.

Before the main data mining goal could be pursued, a number of subgoals had to be reached. First was a warehousing effort to facilitate coherent access to existing data as well as a framework for accepting new data. The objective was to use Structured Query Language (SQL) queries to directly prepare numerous variants of needed data sets (SQL is used to query and manipulate contents of a database. Standard SQL is supported by all major database systems). Second was creation of a model of a normal LV. From a data mining point of view, it is was a summarization or information-granular- ization task [3, 14, 15]. Only examples of patients with normal LVs were used for creating the model. The model was validated qualitatively by a cardiologist as well as quantitatively by computing correlations between the model and examples of normal ventricles contained in the database. Parameters of the model were optimized until the desired quality was achieved. The third subgoal was to determine what information about the patient should supplement the features extracted from the SPECT images in order to generate diagnostic rules. Besides the final diagnosis of LV perfusion, diagnosis for a number of predetermined

Table 1. Cases with Complete Sets of SPECT Images and Corresponding Diagnosis

	Male			Female			Total		
	No ART	with ART	**Total**	No ART	with ART	**Total**	No ART	with ART	**Total**
NL	20	17	***37***	28	31	***59***	48	48	***96***
IS	15	4	***19***	15	4	***19***	30	8	***38***
INF	36	3	***39***	16	3	***19***	52	6	***58***
IS-IN	47	2	***49***	8	1	***9***	55	3	***58***
EQ	3	0	***3***	1	0	***1***	4	0	***4***
REV	0	1	i	0	1	***1***	0	2	***2***
LVD	0	0	***0***	2	0	***2***	2	0	***2***
	121	*27*	***148***	*70*	*40*	***110***	*191*	*67*	***258***
IS, IS-IN	1	0	***1***	0	0	***0***	1	0	***1***
IS, REV	2	1	***3***	1	0	***1***	3	1	***4***
IS, LVD	0	0	***0***	0	1	***1***	0	1	***1***
IS-IN, LVD	6	0	***6***	0	0	***0***	6	0	***6***
INF-IS	10	0	***10***	1	0	***1***	11	0	***11***
INF, IS-IN	2	0	***2***	0	0	***0***	2	0	***2***
INF, REV	3	0	***3***	0	0	***0***	3	0	***3***
INF, LVD	8	0	***8***	1	0	***1***	9	0	***9***
LVD, REV	1	0	***1***	0	0	***0***	1	0	***1***
	33	1	***34***	3	1	***4***	36	2	***38***
Total	154	28	***182***	73	41	***114***	227	69	***296***

NL - normal, IS - ischemia, INF - infarct, IS-IN - ischemia and infarct, EQ - equivocal, REV - reverse redistribution, LVD - left ventricular dysfunction, ART - artifact)

ROIs within SPECT images was also recorded in the patient database. This information was used to check whether the correct features from the SPECT images were extracted. This was done in order to eliminate feature-extraction errors as well as to help in evaluating the quality of the available data.

Understanding the SPECT Data

The data collection process was initiated at the Medical College of Ohio (MCO) in 1992. Data were recorded first on paper worksheets, then entered manually into an Excel spreadsheet. Each row corresponds to a single patient visit. We recorded 184 parameters. SPECT images were stored in a proprietary format, without database organization. Archiving of images was not systematic, due to significant storage requirements. Earliest available SPECT images date back only to 1996. As a result, from over 4000 visits recorded since 1992, only 613 image sets were available at the end of 1998, and only about half of them had complete diagnoses, including codes for perfusion in the ROIs and the overall left ventricular perfusion classification.

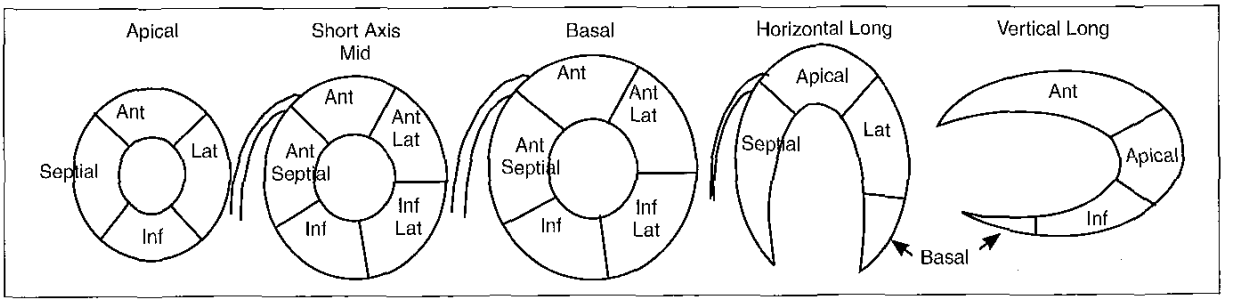
The initial effort was data warehousing. Data contained in the spreadsheet were converted to a relational database. The proprietary SPECT image file format had been reverse-engineered to the level that allowed the most critical information to be extracted—the actual 3-D images and patient identification information (hospital number and SPECT date) stored in the header. Software for automatic indexing of images, using patient identification information, was also created. Image indexes were stored in the database table. Images were stored outside of the database, within a predetermined directory structure. The database design objective was simplicity of maintenance and ability to easily add new patient records and images as they became available. Software for browsing patient records with simultaneous display of available images in several modes was written. The database also stored data generated by various data mining activities, such as information about generated models of a normal ventricle, and features extracted from the SPECT images. This way, an SQL query can be directly used to generate variants of data sets needed for the diagnosis.

Verification of Data Quality

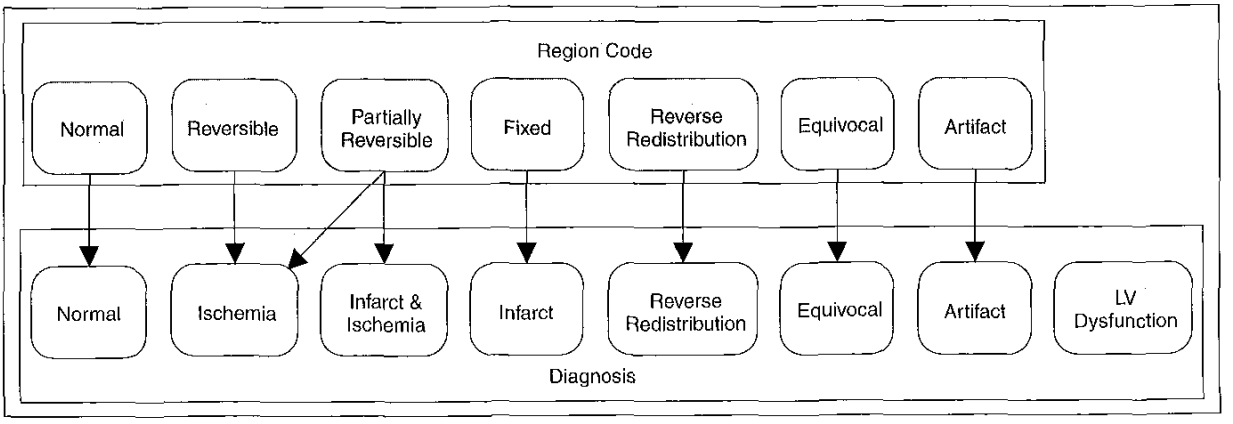
The original data have been semi-manually inspected to eliminate errors such as typos. We intended not to modify the data unless the correction was straightforward. Rather, where possible, we constructed SQL queries to filter undesired records.

The numbers of records available were counted to estimate the statistical validity of expected results; for instance, whether a sufficient number of examples existed for each learning class. Table 1 shows the number of cases in the database for each of the LV perfusion classifications. The top of the table corresponds to records with a single classification code, the bottom to records that contain more than one classification code.

By matching image sets to database records, we found some data errors; most were resolved as typographical, but some of the images remained without matching patient records and were thus discarded. Next, image sets were checked for completeness (e.g., stress study missing) and for quality of the individual images, mostly related to sufficient contrast (photon count).



2. Slices and regions of interest. The first three images correspond to a short-axis view, the last two to a horizontal long-axis view and a vertical long-axis view, respectively.



3. Diagnosis of LV condition based on local classifications.

Preparation of the SPECT Data

The largest task in the data preparation was extraction of features from the SPECT images. It was a highly iterative process, as a number of feature-extraction approaches were tested. Each time, the quality of the extracted features was ultimately tested by an attempt to build a classifier. This also led to modification and clarification of data mining objectives. For instance, initially, we implicitly assumed that the database was noise-free. However, we realized later that noise in the database might be responsible for biasing the assessment of feature extraction. Consequently, more emphasis was put on assessing quality (noise) in the patient database.

Each diagnostic patient study contains two 3-D SPECT cardiac image sets of the LV (one each of rest and stress). Comparing the two image sets allows the interpreting physician to decide on a diagnosis, such as ischemia, infarct, or artifact. Evaluation of the images is a highly subjective process, with potential for substantial variability [11]. To analyze the images, we followed a procedure originally described by us in [16]. The raw image data taken from multiple planar views are processed by filtered backprojection to create a 3-D image. These 3-D images are displayed as three sets of 2-D images, corresponding to the *short-axis view*, *horizontal long-axis view*, and *vertical long-axis view*.

From these 2-D views, the interpreting physician selects five slices to represent the final report. From the short axis view, one slice is taken near the heart's apex, one at the mid-line of the ventricle, and one near the base. With this technique, a single slice is selected for each of the horizontal and vertical axis views, corresponding to the center of the LV cavity. Each of these five images is subdivided into four to five ROIs along the walls of the LV, for a total of 22 regions (see Fig. 2). The appearance of the LV and maximum count in each of the regions is then evaluated. Corresponding ROI locations on the stress and rest images are compared. Perfusion in each of the regions is classified into one of seven defect categories: *normal, reversible, partially reversible, fixed, defect showing reverse redistribution, equivocal,* or *artifact*.

The physician's impression of overall LV perfusion, or the final SPECT image analysis result, is concluded from the results of analysis in each of the ROIs. From the analysis, the interpreting physician categorizes a study as showing one or more of eight possible conditions: *normal, ischemia, infarct and ischemia, infarct, reverse redistribution, equivocal, artifact,* or *LV dysfunction* (see Fig. 3). Note that classification *LV dysfunction* is included as a possible condition diagnosed by the interpreting physician; however, it does not describe perfusion and does not depend on region codes. Some of the perfusion categories may coexist; for example, normal and artifact, reverse redistribution and infarct, etc.

The most fundamental operation performed by the interpreter during analysis of SPECT images is comparison of the case at hand to a mental image of a normal LV. The first task is to establish the location of the ROIs within the SPECT image. This process is complicated by two factors that create a challenge for any algorithmic implementation, as defined below:

- **Actual LV detects.** Changes in perfusion of the LV are manifested as changes in the brightness of the SPECT image. When perfusion is reduced, the radioactive counts are low, and, in effect, parts of the LV may not even be apparent in the image, due to extremely poor perfusion. The interpreting physician deals with this loss of counts by mentally "reconstructing" the missing contour of the image, based on knowledge of heart anatomy and experience with cardiac SPECT imaging. However, this is a major challenge for computer algorithms.
- **Artifacts.** The most common artifact for Tl^{201} imaging is decreased counts, usually from attenuation by breast tissue in females, or by the diaphragm in males. Artifacts may complicate localization of the 22 ROIs. Also, even after the analysis regions are determined correctly, presence of artifacts may lead to false diagnosis, since the decreased count may be erroneously taken for real perfusion defects.

Once the predefined ROIs are established, differences between rest and stress images in each location are analyzed, and counts within each region are compared to that of a normal model. The overall impression of myocardial perfusion is directly concluded from the results of this analysis from each of the regions.

In this work, we used a combination of computer vision and machine learning to mimic the diagnostic process performed by an interpreting physician. Before a machine learning algorithm can be used, a set of features needs to be extracted from the 3-D images. The most natural approach is to extract a single feature corresponding to each ROI in both rest and stress images (see Fig. 2), as was originally done by us in [16]. Each feature can be represented by a single number; e.g., maximum count, mean count, median count, etc. Thus, we

Table 2. Estimation of Classification Error Level Using 10-Fold Cross Validation. Mean of Percent Error and Standard Deviation Are Shown.

Data Set	# of Variables	# of Cases	Constant Classifier	C4.5 Trees		C4.5 Rules	Discrete Naïve Bayes
				No Pruning	Pruning		
Code22	22	1576	73.73 ± 0.89	24.50 ± 0.88	20.94 ± 0.82	23.67 ± 0.78	21.26 ± 0.92
Count	7	1576	73.73 ± 0.89	**18.15 ± 0.91**	**17.58 ± 0.97**	**17.96 ± 0.92**	**18.28 ± 0.77**
Extras	13	1433	73.83 ± 1.00	68.94 ± 1.25	67.62 ± 1.33	64.40 ± 1.16	63.85 ± 0.95
Code22 & Extras	35	1433	73.83 ± 1.00	26.87 ± 1.30	23.87 ± 1.46	23.52 ± 1.32	22.68 ± 0.71
Count & Extras	20	1433	73.83 ± 1.00	23.38 ± 0.99	18.98 ± 1.20	20.97 ± 1.14	19.48 ± 1.45
Code22 & Count	29	1576	73.73 ± 0.89	23.23 ± 0.85	18.72 ± 1.04	19.93 ± 1.13	20.12 ± 1.00
Code22 & Count & Extras	42	1433	73.83 ± 1.00	24.64 ± 1.47	20.94 ± 1.75	20.66 ± 1.45	20.80 ± 1.22

have a set of 44 features for each patient, which can be used to classify LV perfusion. Another approach is to perform local classification first, in each of the 22 regions using information from rest and stress images, and then use these 22 intermediate regional classifications to classify overall LV perfusion.

It is difficult to automatically perform correct and repeatable determination of the ROIs directly from the 3-D images, because of artifacts, actual LV defects, and anatomical differences between patients. In order to do that, we used a model of a normal LV. Location of the regions was a part of the model description. The model plays a role analogous to the interpreter's mental image of a normal LV. The first step in the feature-extraction process is registration—matching the image at hand to the model using translation, rotation, and scaling operations. The image may be matched with a number of models. The model with highest correlation ratio is selected and used to locate slices and ROIs in the image. Regional perfusion is determined, based on the count/intensity of LV walls within the region. Even when the image and a model are correctly registered, the walls of the model and case under investigation may not completely overlap, thus compromising quality of the feature-extraction process.

Correct determination of myocardium location is difficult. The approach we used, besides direct reference to the normal LV model, is similar to the radial search as used in SPECT bull's-eye methods [2, 8, 9]. The model is used to determine the center of the left ventricular cavity. A search is performed in a desired direction, starting from the center of the cavity; the maximum intensity value along the search direction is recorded. This method is based on the premise that counts within the LV wall are higher than in surrounding areas.

Another critical issue is normalization of the image intensity range. Not only do counts vary significantly among patients, they are also different between the rest and stress images for the same patient. There is no easy way to correct this. Typically, numerical values are normalized as a percentage of the maximum count in a given 3-D image.

Data Mining

Relevance Analysis and Classification Error Reference Level

Figure 3 shows a simplified perfusion classification process that is deterministic and entirely based on classification of perfusion in each of the ROIs. It is reasonable to stipulate that in practice this classification is more complex, nondeterministic, and involves a number of other factors. So our data mining subgoal is to verify that the classification of myocardial perfusion is indeed nondeterministic, to establish classification noise (classification error rate), and to determine if use of additional features from the database reduces classification error. In other words, we want to determine how consistent the diagnosis is, and whether it can be based on other features than those present in the SPECT images alone. Note that if classification of myocardial perfusion were completely deterministic, given the classification within ROIs, then the reference classification error would be zero. Since it is not, we want to estimate it so that more realistic performance requirements can be set for classifiers that use only the features extracted from SPECT images.

The upper boundary of the classification noise present in the database can be estimated by constructing a particular classifier and measuring its error rate using cross validation. To make this assessment goal, we used a C4.5 classifier [17]. In addition, to make sure that the results produced were correct, we compared them to naïve-Bayes [18] and *constant* classifiers. A constant classifier predicts a constant class—the majority class in the training set. It is used to establish a base performance level.

We used only records that had no missing, incorrect, or multiple values for features considered. Three sets of features were used to estimate classification error level:

- Diagnosis by a cardiologist for each of the 22 ROIs (Fig. 2). We refer to this 22-feature set as `Code22`. Each of the features can take on seven values that describe defect type, coded as NL - normal, F - fixed, R - reversible, P - partially reversible, E - equivocal, X - reverse redistribution, ART - artifact. This set has 1576 cases.
- A new data set was obtained from the `Code22` set by counting how many times each of the defect codes were present within the 22 regions for a given case. These data became a set of seven features, taking on integer values from 0 to 22. Similar to the `Code22` set, this set has 1576 cases. We refer to this data set as `Count`.
- Still another set was created using criteria such as relevance, completeness, and correctness of feature values. This resulted in 16 features, with a reasonably high number of corresponding entries in the `Code22` data set. The following variables were selected: sex, age, height, weight, body surface area coefficient (BSA), diabetes mellitus, family history, HTN, smoker/nonsmoker, chest pain, high cholesterol, prior MI, and left ventricular size. This set has 1433 cases. We refer to it as `Extras`.

Table 3. Example of Rules Generated for the Count Data Set

Training Set Error	Test Set Error	Used Train/Test	Wrong Train/Test	IF	THEN
0.4%	1%	241/103	1/1	NL>21	NL
11.3%	14.6%	106/48	12 / 7	ART>0 & R<=3 & P=0 & F=0 & X<=1	IS
15.7%	18.8%	204 / 80	32 / 15	ART=0 & R=0 & P <=2 & E=0 & X>1 & X<=11	ART
7.8%	15.9%	115 / 80	9 / 15	ART=0 & R<=2 & P>6	INF
22.0%	17.8%	82 / 45	18 / 8	ART=0 & R<=1 & P>3	INF
16.9%	15.4%	154 / 65	26 / 10	R>1 & X<=1	IS-IN
25%	0%	4 / 1	1 / 0	NL>15 & ART>0 & R=0 & P>0 & P<=2 & X<=1	IS-IN

The first two sets contain only information from the interpretation of SPECT images. The last one contains other features from patient records that might influence the diagnosis of myocardial perfusion. Table 2 contains classification error rates for each of the data sets as well as their combinations. Ten-fold cross validation was used and mean of percent error and standard deviation are shown.

The following conclusions can be drawn from Table 2:

- The data seem to have relatively high classification noise. The lowest classification error is above 17%.
- Representation of the regional classification codes, the `Count` data set, seems to be *easier* for learning.
- Additional (`Extras`) features contain some useful information for the classification. The error rate for the `Extras` data set is high; however, it is almost 10% lower (discrete naïve Bayes classifier) than that of the constant classifier, suggesting that some of the features in the `Extras` data set *are* correlated to classification category.
- Classification using the extended data sets (i.e., regional information with added features) suffers from increased classification error.

Table 4. Examples of Rules Generated for Count & Extras Data Set

Training Set Error	Test Set Error	Used Train/Test	Wrong Train/Test	IF	THEN
0.9%	0%	215/97	2 / 0	NL>21	NL
0%	0%	32 / 17	0 / 0	Chest_Pain=Y & NL<=21 & P=0 & F=0 & E=0 & X<=1 & ART=0	IS
4.3%	5.9%	46 / 17	2 / 1	R>2 & P=0 & F=0 & ART=0	IS
1.5%	6.9%	66 / 29	1 / 2	BSA>1.43 & NL<=5 P>2 & P<=11 & F>0	IS-IN
0%	0%	12 / 8	0 / 0	Sex=M & Smoker=N & P>5 & F=0 & ART<=1	IS-IN
22%	0%	9 / 9	2 / 0	Smoker=Y & Prior_MI=N & NL<=18 & P>0 & ART<=1	IS-IN
8.3%	0%	12 / 3	1 / 0	Height>67 & BSA<2.29 & Prior_MI=Y & P>0 & F=0 & ART<=1	IS-IN
0%	0%	5 / 1	0 / 0	F>6 & ART>1	INF
15.6%	7%	96 / 43	15 / 3	Family_History=Y & NL<=21 & R=0 & P=0 & F<=1 & X=0	ART

Table 5. Ranking of Features in Count & Extras Data Set

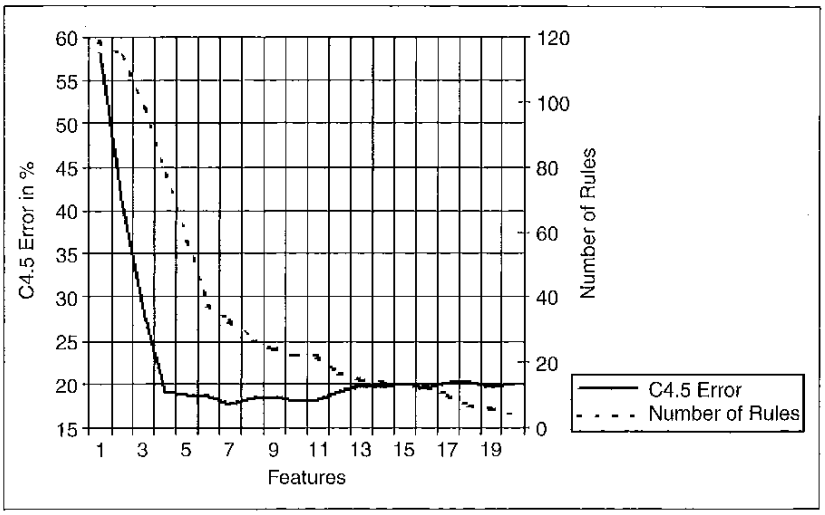
Feature	Rank			Number of Features Used	C4.5 Error
	Train	Test	Total		
Count P	46	73	119	1	58.20
Count F	42	72	114	2	41.13
Count AART	35	63	98	3	27.64
Count R	26	52	78	4	19.05
Count NL	24	33	57	5	18.71
Count X	12	26	38	6	18.63
Chest pain	1114	19	33	7	17.80
Sex	9	18	27	8	18.50
Count E	12	12	24	9	18.64
High cholesterol	9	13	22	10	18.01
Family history	9	13	22	11	18.29
Height	8	9	17	12	19.06
BSA	8	7	15	13	19.75
Prior MI	6	8	14	14	19.56
Age	7	6	13	15	19.96
LV Size	7	6	13	16	19.47
Smoking	6	4	10	17	20.17
Weight	3	4	7	18	20.14
HTN	4	2	6	19	19.62
DM	1	3	4	20	19.97

To obtain better understanding of classification, we investigated rules generated by the C4.5 algorithm. Some examples of rules and their error rate for the Count data set are shown in Table 3. Rules shown in the table had a classification error lower than average for that particular data set.

Some rules are similar to those shown in Fig. 3. For instance, the first rule from Table 3 states that if all 22 regions are classified as normal, then the left ventricular perfusion is normal. Some rules, like the third one, are at first puzzling. The first clause of the rule states that there are no artifacts, ART = 0. However, it classifies the case as "artifact." It seems that studies showing between two and 11 ROIs with reverse redistribution (X), and fewer than three ROIs with partial redistribution (P), were diagnosed as artifacts by the interning physician. This rule has a relatively high accuracy and correctly covers a large number of cases.

Table 4 presents some of the rules generated by C4.5 for the Count & Extras data set. The rules in the table are examples of rules with smaller than average error rate for the test set. The first important observation is that inclusion of additional features leads to creation of highly accurate rules. However, Table 2 suggests that adding too many additional variables deteriorates classification quality [19]. We investigated further by quantitative analysis of the generated rules. We selected the Count & Extras data set for qualitative evaluation of relevance of individual features for correct classification. Importance of the variables was ranked by counting the number of times they appear in the rules that have better than average error rate for a particular data set. Ten-fold cross validation was used; the number of appearances in each of the 10 sets were added together. Table 5 shows the results. The table has two parts: ranking of the features and classification error rate achieved by using the ranking for features selection.

Rank columns represent how many times a particular feature is present in the rules that have an error rate lower than average for a given cross-validation data set. Features are listed from the most to least frequent (also see Fig. 4). As expected, most of the regional location code counts are at the top of the list. The last column shows the error rate of C4.5-rules classification using the top-ranked features. For example, when three features were used

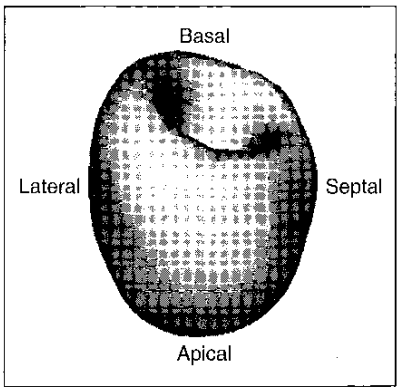


4. Ranking of features in Count & Extras data set and the classification error versus number of features used.

(Count P, Count F, and Count ART), the classification error rate was 27.64%. Initially, addition of features, starting with only one, significantly reduces the error rate. The lowest level is reached for seven features; for more, the error rate increases..

Model of the Normal Left Ventricle

There are a number of possible approaches to create a model of the LV. Recently, physics-based deformable models have been gaining popularity [20]. The premiss is that they are well suited to deal with natural anatomical differences. But their drawbacks are relative complexity, and they are difficult to use in cases when there are large perfusion defects. Thus, we decided to build a rigid model of the LV by "averaging" a set of images corresponding to cases diagnosed by the interpreting physician as normal. We also decided to use only the images that were evaluated by the most experienced physician. Images were additionally screened for excess presence of noise and artifacts. Case images are translated, rotated, and scaled to obtain the best match among them. A variant of best-first heuristic was used to make correlation search computationally feasible. Once matched to each other, case images were added, constituting an "averaged model." Due to anatomical differences, models for females and males were created separately. Models were built separately for stress and rest images. The format of a model is the same as the 3-D SPECT image, so a cardiologist can easily evaluate its quality. Each model was manually inspected,



5. Three-dimensional rendering of a normal left ventricle model.

and the locations of slices and ROIs for this particular model were recorded. An example of 3-D rendering of a male rest model is shown in Fig. 5. The rendering was created using the Visualization Toolkit library [21].

Perfusion Classification

Direct Classification Test

In our initial attempt, a direct classification method was used, with a model of a normal LV used as a reference. SPECT images were transformed to align with the model, using scaling, shifting, and rotation in 3-D. Locations of ROIs (see Fig. 2) were marked on the model. Once an image was registered with the model, the maximum, mean, and median count within each ROI were recorded, for both rest and stress images. A total of 44 variables was used, 22 from rest images and 22 from stress images, to permit comparisons. Three data sets, using max, mean, or

Table 6. Error Rate For Direct Perfusion Classification Tests								
Sex	**Feature Type**	**# of Variables**	**# of Cases**	**Constant Classifier**	**C4.5 Trees**	**C4.5 Rules**	**Naïve Bayes**	**Discrete Naïve Bayes**
Male	Max	44	121	16.43 ± 2.90	27.02 ± 3.22	28.69 ± 3.14	28.93 ± 4.16	24.76 ± 3.06
	Mean	44	121	16.43 ± 2.90	18.10 ± 2.65	18.10 ± 2.62	28.10 ± 3.64	17.50 ± 3.72
	Median	44	121	16.43 ± 2.90	16.31 ± 3.25	**14.64 ± 3.44**	26.43 ± 3.27	17.50 ± 3.91
Female	Max	44	70	38.53 ± 6.58	28.33 ± 6.12	28.33 ± 6.12	29.58 ± 6.06	31.25 ± 5.27
	Mean	44	70	38.53 ± 6.58	30.42 ± 5.71	28.75 ± 5.91	34.17 ± 5.13	28.33 ± 4.46
	Median	44	70	38.53 ± 6.58	**22.50 ± 6.17**	23.75 ± 6.57	31.25 ± 5.20	27.08 ± 4.94

median as values representing each region, were created separately for female and male patients. Only records that had a single perfusion diagnosis, with no artifacts, were selected (see Table 1). All defect classes were combined into one class, called "abnormal." Twenty-fold cross validation was used. The classification error rate is shown in Table 6. Notice that the best results presented in Table 6 are not much larger than the reference classification error given in Table 2.

Apex Feature Extraction Test

In this test, we concentrated on classifying a defect within an ROI (see Fig. 3). Regions HORIZONTAL_LONG_APICAL and VERTICAL_LONG_APICAL intersect in three dimensions. We classified cases that had the same code in both of these regions. This time we tried to use more than a single feature extracted from a region. Four features (maximum, mean, minimum, and standard deviation of gray levels within the ROI) were recorded.

First, rest and stress 3-D short-axis SPECT images were mapped into two dimensions using the method presented in [8] (see Fig. 6). A 3-D image is converted into 2-D images using radial projection in spherical coordinates. Two types of projections are performed. In one, only the maximum value along the radius is recorded, called the MAX image. In the other, the values along the radius are integrated, called the TOTAL image. Before unwrapping, each 3-D image was reoriented and scaled to best match a LV model; only the cases for which good fit was achieved were used (correlation coefficient greater than 0.93 for females and 0.95 for males). The center of the 2-D region corresponds to the apex. We used two variants to delimit the region of the apex: a circle and a pair of intersecting strips (corresponding to vertical and horizontal long-axis views). Four parameters were calculated for each delimited apex region: minimum, maximum, mean and standard deviation. Thus, a feature set for each case consists of 16 values: four for the rest-MAX, rest-TOTAL, stress-MAX and stress-TOTAL, respectively.

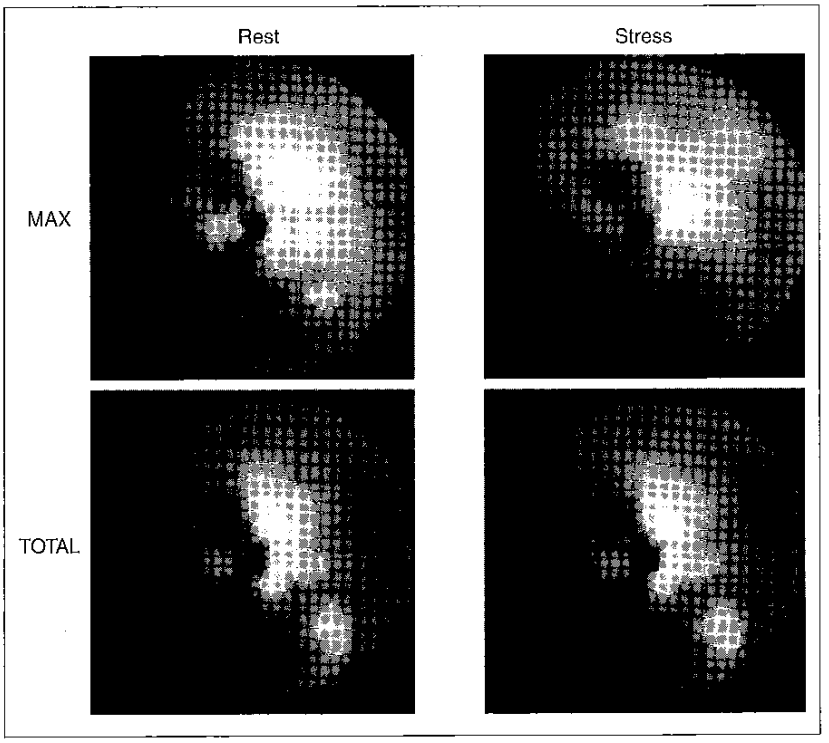
Due to the large disproportion between normal and abnormal cases, and low representation for particular abnormal types (see Table 7), all abnormal cases were combined into a single category, called abnormal. As a result, only two classes are used: normal and abnormal. Then, 10-fold cross validation was performed, and the results are shown in Table 8.

Evaluation of the Discovered Knowledge

Assessment of the Results Versus the Objectives

We investigated three data mining goals relevant to our objective of semi-automating the diagnosis of LV abnormalities based on SPECT imaging technique. Here we summarize our results:

Relevance analysis and reference classification error: The most important observation is that our SPECT database contains a significant level of noise, even after cleaning and record selection. This affects assessment of the remaining data



6. Example of unwrapping a 3-D SPECT image of the left ventricle onto a 2-D image using a method similar to [8]. The central portion of each of the images corresponds to the apex.

mining goals. Feature relevance analysis reveals that the overall perfusion diagnosis depends, to some extent, on factors other than the SPECT images. Addition of extra features only slightly reduces the classification error, but this is very quickly neutralized by increased error due to increase of dimensionality of the feature space. Transforming the features can alleviate the "curse of dimensionality": the `Count` data set (seven features) obtained from the original regional classification data set `Code22` (22 features) has a consistently lower classification error than the original data set.

Normal LV model: The model proved to be helpful in normalizing differences among patients. However, numerical aspects of using the model need to be optimized. Currently, matching of cases to a model is done off-line, and match results are stored in the database for convenient retrieval when testing variants of feature extraction.

Perfusion classification: Surprisingly, direct overall classification showed lower classification error than regional classification. On the other hand, we tested regional classification in the most difficult region for diagnosis—the apex. The direct classification error is only slightly higher than the reference classification error established by the first data mining goal.

Reviewing the Entire Knowledge Discovery Process

Several iterations were performed during the knowledge discovery process described, especially during model construction and feature extraction for the perfusion classification. The results presented were reached at the end of the iterative process. Some of the iterations performed were very computationally intensive; for instance, a single iteration of matching (registering) normal LV models to all of the images in the database took almost 10 days using a PC with a Pentium II 300 MHz processor and 128 Mbytes of RAM. Registering is essential to the quality of the feature-extraction process. Whenever possible, we kept matching results in the database and modified only details of the feature extraction to improve computational efficiency.

Determining Actions to be Taken Based on the Achieved Results

In the course of the present work, we realized that the quality of the data contained in the database must be improved. This is critical for the building of perfusion classifiers. We need to address:

1. Noise in the database.
2. Adequate representation of all perfusion defects.

The second issue can only be addressed by collecting more data. The first can be addressed by improving ways in which data are collected. However, it is also possible to correct data already present in the database by rediagnosing archived SPECT images.

As mentioned above, the computational efficiency of using models of the LV needs to be addressed. On the other hand, we may be able to further improve our models by combining them with physics-based deformable model approaches. Such combinations may decrease computational efficiency; however, we expect that improved image matching capabilities are worth it.

During perfusion classification tests, we used only two types of classifiers: decision-tree-based C4.5 and simple statistical naïve-Bayes. Testing a larger number of classifier types may help to find one that better *fits* our problem. Encouraged by good performance of the simple naïve-Bayes classifier, which was close to and at times even better than C4.5, and recognizing intrinsic uncertainty in our data, we plan to try more sophisticated classifiers such as Bayesian network classifiers or other types of classifiers based on graphical models.

We did not reach the stage of using the discovered knowledge, as clinical trials are required before the system can be implemented. At this stage, we plan that the system will be tested by cardiology fellows at MCO to assess its success in distinguishing normal versus abnormal patients on new SPECT images.

Conclusions

Data mining and general knowledge discovery techniques appear to be useful for classification of SPECT cardiac images. Creation and mining of the database has illustrated the importance of precision in data input, which is not always present in the narrative, or even in graphics-coded descriptions of images provided by physicians. Using data mining techniques on the raw images themselves may actually improve diagnosis and help clean the database. At the moment, these are time-consuming tasks, but methods are available to improve time requirements. As in all diagnostic systems, addition of more representative cases in each classification should improve performance. Such techniques may also have

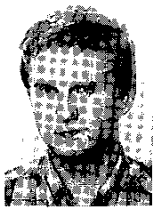
Table 7. Numbers of Available Examples of Defects in the Apex Area

Sex	Normal	Fixed	Partially Reversible	Reversible	Reverse Redistribution	Equivocal	Artifact	Total
Female	81	7	4	5	1	0	2	128
Male	81	26	6	13	1	0	0	100

Table 8. Cross-Validation Error Rate

		# of Variables	# of Cases	Constant Classifier	C4.5 Trees	C4.5 Rules	Naïve Bayes	Discrete Naïve Bayes
Male	Circle	16	128	36.31 ± 4.16	26.31 ± 5.00	26.31 ± 4.38	26.19 ± 4.80	**23.69 ± 3.53**
	Strips	16	256	36.70 ± 1.93	24.94 ± 2.78	22.60 ± 2.59	21.44 ± 2.61	**20.32 ± 2.67**
Female	Circle	16	100	19.00 ± 3.40	21.00 ± 3.40	21.00 ± 3.40	21.00 ± 2.28	**19.00 ± 3.40**
	Strips	16	200	19.00 ± 1.76	20.50 ± 3.20	20.50 ± 3.20	23.00 ± 3.85	**18.00 ± 3.29**

application in the quality control of diagnostic laboratories.



Jaroslaw P. Sacha received an M.S. degree in electronics and automatic control in 1991 from the University of Mining and Metallurgy, Krakow, Poland. In 1999, he received a Ph.D. degree in engineering science from the University of Toledo, Ohio. He has several years of professional experience working on industrial research projects in the area of computer vision, quality inspection, and knowledge discovery. Some of the companies he cooperated with include Computer Associates, Libbey-Owens-Ford (Pilkington), Owens-Corning, Owens-Illinois, NASA, and Philips Electronics. His current research interests are artificial intelligence, computer vision, and knowledge discovery. He is a member of the AAAI, IEEE, and SPIE.

Krzysztof J. Cios is a professor and interim chair of the Department of Bioengineering and Professor of Electrical Engineering and Computer Science at the University of Toledo. He is also an adjunct professor of medicine at the Medical College of Ohio. His research interests are in the areas of biologically inspired learning algorithms, in particular artificial neural networks and machine learning, and in data mining and knowledge discovery. NASA, NSF, American Heart Association, Ohio Aerospace Institute, and NATO have funded Dr. Cios's research. He has published two books, 35 journal articles, nine book chapters, and 48 peer-reviewed conference papers. He serves on the editorial boards of *Neurocomputing*; *IEEE Transactions on Systems, Man and Cybernetics*; *IEEE Engineering in Medicine and Biology Magazine*; and *Mathematical Neuroscience*. Dr. Cios has been the recipient of the Norbert Wiener Outstanding Paper Award, the Neurocomputing Best Paper Award, the University of Toledo Outstanding Faculty Research Award, and the Fulbright Senior Scholar Award. He received his Ph.D. from the University of Mining and Metallurgy, Krakow; his M.B.A. from the University of Toledo; and his D.Sc. from the Systems Research Institute, Polish Academy of Sciences. Dr. Cios is a member of the IEEE, AAAI, and Sigma Xi.

Lucy S. Goodenday is an associate professor of medicine in the Division of Cardiology at the Medical College of Ohio in Toledo, Ohio, where she practices general and nuclear cardiology. She serves as the director of the Nuclear Cardiology and Exercise Laboratories at the Medical College Hospital. She graduated from Bryn Mawr College and received her medical degree from New York Medical College. After internship and residency, she was a USPHS-NIH trainee in cardiology at the University of California, San Francisco, and a cardiology research fellow at the University of London. She has practiced and performed research in nuclear cardiology since 1978.

Address for Correspondence: Krzysztof J. Cios, Department of Bioengineering, University of Toledo, Toledo, OH 43606-3390. E-mail: kcios@eng.utoledo.edu.

References

1. **Crevier D and Lepage R:** Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding* 67(2): 161-185, Aug. 1997.

2. **Cios KJ, Teresinska A, Konieczna S, Potocka J, and Sharma S:** A knowledge discovery approach to diagnosing myocardial perfusion. This issue, pp. 17-25.

3. **Cios KJ, Pedrycz W, and Swiniarski R:** *Data Mining Methods for Knowledge Discovery*. Norwell, MA: Kluwer, 1998.

4. **Cullom SJ:** Principles of cardiac SPECT. In: DePuey ED, Berman DS, Garcia EV (Eds): *Cardiac SPECT Imaging*. New York: Raven Press, 1995.

5. **Garcia EV, Ezquerra NF, DePuey EG, et al.:** An artificial intelligence approach to interpreting thallium-201 3-dimensional myocardial distributions (Abstract). *J Nucl Med* 27: 1005, 1986.

6. **Faber TL, Akers MS, Peshock RM, and Corbett JR:** Three-dimensional motion and perfusion quantification in gated single-photon emission computed tomograms. *J Nucl Med* 32: 2311, 1991.

7. **Faber TL, Cooke CD, Peifer JW, et al.:** Three-dimensional displays of left ventricular epicardial surface from standard cardiac SPECT perfusion quantification techniques. *J Nucl Med* 36: 697-703, 1995.

8. **Goris ML, Boudier S, and Briandet PA:** Two-dimensional mapping of three-dimensional SPECT data: A preliminary step to the quantification of thallium myocardial perfusion single photon emission tomography. *Amer J Physiologic Imaging*, 2: 176-180, 1987.

9. **Van Train KF, Garcia EV, Cooke CD, and Areeda J:** Quantitative analysis of SPECT myocardial perfusion. In: DePuey ED, Berman DS, Garcia EV (Eds): *Cardiac SPECT Imaging*. New York: Raven Press, 1995.

10. **Corbett JR:** Gated blood-pool SPECT. In: *In:* DePuey ED, Berman DS, Garcia EV (Eds): *Cardiac SPECT Imaging*. New York: Raven Press, 1995.

11. **Cuaron A, Acero AP, Cardenas M, et al.:** Interobserver variability in the interpretation of myocardial images with Tc-99m-labeled diphosphonate. *J Nucl Med* 21: 1, 1980.

12. **Francisco DA, Collins SM, and Go RT et al.:** Tomographic thallium-201 myocardial perfusion scintigrams after maximal coronary vasodilation with intravenous dipyridamole: comparison of qualitative and quantitative approaches. *Circulation* 66(2): 370, 1982

13. **Ezquerra N, Mullick N, Cooke N, Garcia E, and Krawczynska E:** PERFEX: An expert system for interpreting 3D myocardial perfusion. Graphics, Visualization & Usability Center, Georgia Institute of Technology, Atlanta, GA, Technical Report GIT-GVU-92-02, 1992.

14. **Chapman P, Clinton J, Khobaza T, Reinartz T, and Wirth R:** The CRISP-DM Process Model (discussion paper). On-line: http://www.crisp-dm.org. CRISP-DM Consortium, March 1999.

15. **Fayyad U, Piatetsky-Shapiro G, and Smyth P:** From Data Mining to Knowledge Discovery in Databases. *AI Mag* 17(3): 37-54, 1996.

16. **Cios KJ, Goodenday LS, Shah KK, and Serpen G:** A novel algorithm for classification of SPECT images of a human heart. In: *Proc CBMS'96*, Ann Arbor, MI, June 1-5, 1996.

17. **Quinlan JR:** *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

18. **Duda R and Hart P:** *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

19. **Friedman JH:** On bias, variance, 0/1 - loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery* 1(1): 55-77, 1997.

20. **Declerck J, Feldmar J, Goris ML, and Betting F:** Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images. *IEEE Trans Med Imag*, 16(6): 727-737, Nov. 1997.

21. **Schroeder W, Martin K, and Lorensen B:** *The Visualization Toolkit*, 2nd Ed. Englewood Cliffs, NJ: Prentice Hall, 1998.