ELSEVIER Artificial Intelligence in Medicine 23 (2001) 149–169

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

Knowledge discovery approach to automated cardiac SPECT diagnosis

Lukasz A. Kurgan[a,b,*], Krzysztof J. Cios[a,b,c,d,f], Ryszard Tadeusiewicz[e], Marek Ogiela[e], Lucy S. Goodenday[f]

[a]*University of Colorado at Denver, P.O. Box 173364, Denver, CO 80217-3354, USA*
[b]*University of Colorado at Boulder, Boulder, CO, USA*
[c]*University of Colorado Health Science Center, Denver, CO, USA*
[d]*4cData, LLC, Golden, CO, USA*
[e]*University of Mining and Metallurgy, Krakow, Poland*
[f]*Medical College of Ohio, Toledo, OH, USA*

Received 30 April 2000; received in revised form 2 January 2001; accepted 2 May 2001

Abstract

The paper describes a computerized process of myocardial perfusion diagnosis from cardiac single proton emission computed tomography (SPECT) images using data mining and knowledge discovery approach. We use a six-step knowledge discovery process. A database consisting of 267 cleaned patient SPECT images (about 3000 2D images), accompanied by clinical information and physician interpretation was created first. Then, a new user-friendly algorithm for computerizing the diagnostic process was designed and implemented. SPECT images were processed to extract a set of features, and then explicit rules were generated, using inductive machine learning and heuristic approaches to mimic cardiologist's diagnosis. The system is able to provide a set of computer diagnoses for cardiac SPECT studies, and can be used as a diagnostic tool by a cardiologist. The achieved results are encouraging because of the high correctness of diagnoses.

Keywords: Knowledge discovery and data mining; SPECT myocardial perfusion imaging; CLIP3 machine learning algorithm

1. Introduction

Modern medicine generates huge amounts of image data that can be analyzed and processed only with the use of specialized computer software. Since imaging techniques

* Corresponding author. Tel.: +1-303-556-6314; fax: +1-303-556-8369.
E-mail address: kcios@carbon.cudenver.edu (K.J. Cios).

0933-3657/01/$ – see front matter
PII: S0933-3657(01)00082-3

like single proton emission computed tomography (SPECT), PET, and MRI can generate gigabytes of data per day. There are many advantages of computerized analysis of data over human analysis: lower price, shorter time, automatic recording of analysis results, consistency, relatively inexpensive re-use of previous solutions.

Our goal is to create a computer system that is able to semi-automate the cardiac SPECT myocardial perfusion diagnostic process, using knowledge discovery approach, to reveal some new and useful information from the data. The data are cardiac SPECT images, clinical information, and physician's interpretation.

Our first sub-goal is to automatically obtain a set of partial diagnoses for regions of the left ventricle (LV) muscle, and the second sub-goal is the overall diagnosis, which describes perfusion of the entire LV cardiac muscle. To achieve these goals we first calculate cardiologist-defined set of features from cardiac SPECT images using image analysis algorithms. Then, using these features we generated diagnostic rules.

Two sets of diagnostic rules will are generated. One for partial diagnoses using features extracted directly from images, and the second for overall diagnosis (using partial diagnoses as an input). A heuristics approach is used to generate the set of rules for partial diagnoses, and inductive machine learning algorithm CLIP3 [2,4] is used to generate overall diagnostic rules.

In this work, we follow the six-step knowledge discovery process of Cios et al. [5,6,13], which is summarized as follows.

1. Understanding the problem domain — learning the terminology and current solutions, determination of medical and data mining goals.
2. Understanding the data — understanding mechanisms of data collection, initial data exploration and verification.
3. Preparation of the data — deciding algorithms inputs, cleaning and reformatting the database, creation of a new database (for latter use). This new database will consist of cardiac SPECT images, patient's clinical information, and other attributes derived at this step. Data preparation is the most time consuming step but its quality largely determines the success of the entire project.
4. Data mining — design of an image analysis algorithms, features extraction, deciding on training and testing procedures, generation of diagnostic rules.
5. Evaluation of the discovered knowledge — description and discussion of the results, description of possible improvements into the diagnostic algorithm.
6. Using the discovered knowledge — before the discovered knowledge can be used it needs to undergo clinical trials, which is not a part of this investigation.

2. Understanding the medical problem domain

First we need to understand the medical problem domain. Our system is primarily designed for cardiologists and cardiology fellows as the second-opinion computer-based diagnostic tool.

2.1. Cardiac SPECT diagnosing

SPECT imaging is used as a diagnostic tool for myocardial perfusion. The patient is injected with radioactive tracer (in our case Tl-201). Then two studies are performed, one 10–15 min after injection during maximal stress — called stress study (stress image), and one 2–5 h after injection — called rest study (rest image). The studies are collected as two sets of three-dimensional images. All the images represent LV muscle perfusion that is proportional to distribution of radioactive counts within the myocardium [10]. Cardiologists compare stress and rest studies in order to detect abnormalities in the LV perfusion.

Since contemporary devices cannot directly display 3D images several transformations are performed to display them. As a result we obtain two-dimensional images or three-dimensional surface rendering [15]. The 2D images preserve intensity information while 3D relations are hard to reconstruct. In the rendering case all 3D information is explicit, but density information is represented only indirectly by color and shape of the 3D surface [11,12].

Normally the SPECT images are presented to a cardiologist as three sets of two-dimensional images, which contain series of intensity slices (about 15–30 slices each). Only these images are used in this paper. Slices that show left ventricle perpendicular to its long axis are called ''short axis'', parallel to its long axis are called ''vertical long axis'' and ''horizontal long axis'', see Fig. 1. The three-dimensional relations can only be mentally reconstructed from these images. The cardiologists diagnose such studies by visually comparing corresponding slices with their mental model of the normal LV.

There are also other visualization methods used for cardiac SPECT images. One is bull's-eye method that is based on projection of 3D image of LV into 2D plane by radial projection into spherical coordinates [16], or into combination of spherical and cylindrical coordinates [20]. Another family of methods is connected with 3D surface rendering of the LV; they use gated blood-pool SPECT images in order to visualize motion of the heart muscle [7]. There are also studies that concern dynamic cardiac scenes interpretation [18]. Cardiac motion analysis in general enables to identify pathologies related to myocardial anomalies or coronary arteries circulation deficiencies. Similarity to the technique described in this paper they use 2D LV contour images to perform quantitative and qualitative evaluation of the heart function.

Many visualization methods have been developed to help in interpretation of the SPECT images, but it has been shown that visual analysis of images is often inconsistent and error prone [9]. Thus, there is a need for tools to assist in diagnosis that is based on computer-aided image display and quantification. Quantitation decreases the variability of image interpretation [14].

2.2. Determination of medical goals

From the medical point of view our goal is to semi-automate cardiac SPECT diagnostic process in order to assist a cardiologist in diagnosing cardiac SPECT images, to make this procedure easier, more consistent, and efficient.

To achieve these goals we will follow a knowledge discovery process described above.

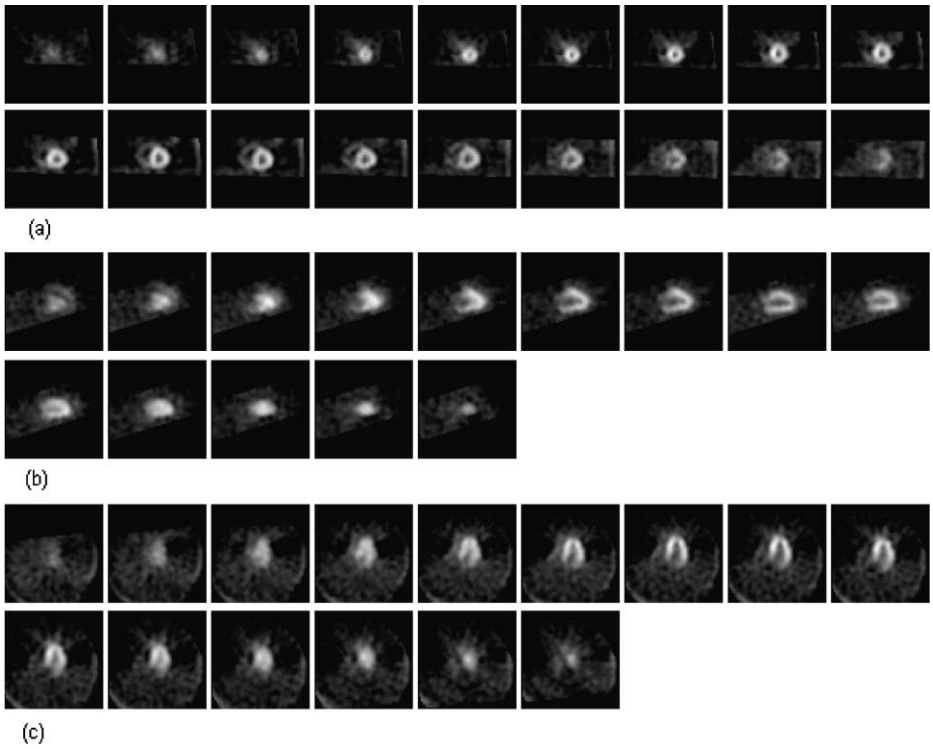

Fig. 1. Series of 2D-intensity slices: (a) short axis; (b) vertical long axis; (c) horizontal long axis view.

2.3. Determination of data mining goals

The main data mining goal is to identify key features from cardiac SPECT images. This will be done using image analysis and processing algorithm. Next, two types of diagnostic rules will be generated.

- *For partial diagnoses*: These rules will use features extracted from SPECT images; they will be generated using heuristics approach.
- *For the overall diagnosis*: These rules will use partial diagnoses as an input. They will be generated using the CLIP3 machine learning algorithm [2,4].

The set of input variables consist of: images from rest and stress SPECT studies, and one clinical information: patient's sex, because sex is the only information used by a physician during diagnosing SPECT images. Extraction of features from SPECT images is a very complex task due to anatomical differences between patients, artifacts in the images, diverse location of objects representing heart perfusion on the image, etc. Thus, we build a model of the normal left ventricle for reconstruction of the poorly perfused parts that may not be visible in the images. We also have to consider anatomical differences between male and female hearts.

In addition, image analysis functions like scaling, rotation, labeling, indexing, and feature extraction etc. need to be implemented. Parameters of the image analysis algorithms are optimized during development stage and after completion of the project. Our experience in the field of cardiac SPECT imaging [19] helped us to determine and define these goals.

3. Understanding the cardiac SPECT data

The data was collected at the Medical College of Ohio (MCO). Patient clinical records were manually entered into MS-Excel spreadsheet. There are about 180 parameters recorded for each patient. Each row in the spreadsheet corresponds to a single patient visit. Images were recorded selectively to reduce storage capacity requirements, thus from over 4000 recorded visits we will analyze only about 600, only those which are accompanied by the corresponding SPECT image. Only about half of the records and SPECT image sets have complete physician interpretation that consists of partial diagnoses for the regions of interest along the LV muscle, and the overall LV perfusion classification.

3.1. SPECT imaging

During cardiac SPECT study patient is injected with radioactive agent, (Tl-201), that during decay emits single photon of 150 (keV) energy. The detector collects the emitted photons. The collimator, part of the detector, collects photons only from specified direction — in this way information about location of the source of emission can be reconstructed. During the study the detectors are located around the patient body and rotated. Using high-level reconstruction algorithms one 3D image, from a set of 2D planar views at different angles, is created.

Cardiac SPECT imaging technique is characterized by low sensitivity, high signal to noise ratio, and application of very complex image reconstruction algorithms. However, it is successfully used in clinical trials because of the relatively low cost.

Cardiac SPECT images represent LV myocardial muscle perfusion that is proportional to distribution of radioactive counts within the myocardium. Typical 2D-image resolution is 64×64, all the images are black and white, 8 bits per pixel with 256 shades of gray. Brighter areas on the image correspond to well perfused areas of myocardium. When part of myocardium is not visible an Ischemia is suspected, see Fig. 2.

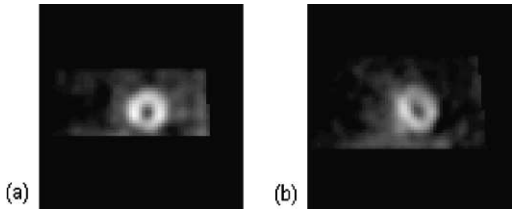

Fig. 2. Perfusion on cardiac SPECT images: (a) Normal perfusion; (b) Abnormal perfusion.

3.2. Description of the SPECT database

The SPECT database consists of images and clinical patient records. Data contained in the spreadsheet is converted to a relational database. Then it is analyzed and the significant attributes are extracted: encrypted patient ID, sex, weight, height, encrypted date of the study, 22 partial diagnoses, and the overall diagnosis. All are recorded in a text file.

Image database is also analyzed. Images are stored in a predetermined directory structure, defined according to the encrypted study date, and encrypted patient hospital number. For each patient there are two 3D images, one for each study and six 2D images (three for each study).

The database design goal is to simplify maintenance and ability to add new patients records and images when they become available.

3.3. Verification of the quality of the data

The original database was semi-manually analyzed to eliminate errors like typos, etc. All records and image sets that were incomplete were eliminated. Available records were counted to avoid situation when there is no sufficient number of examples for each diagnostic case. Most common encountered errors included typographical errors, missing values or images, or a very poor quality of a recorded image (in terms of contrast). In Table 1, a breakdown of the records and images available, after cleaning, is shown for each overall perfusion diagnosis.

Table 1
Breakdown of the SPECT database after cleaning of the data

Overall diagnosis	Number of correct records	Overall diagnosis	Number of correct records
NL (Normal)	49	INF, LVD	8
IS (Ischemia)	27	REV, LVD	1
INF (Infarct)	47	IS, ART	8
ART (Artifact)	31	IS-IN, LVD	3
IS-IN (Ischemia and Infarct)	48	ART, IS-IN	4
EQ (Equivocal)	3	IS, IS-IN	1
REV (Reversible Redistribution)	0	REV, ART	1
LVD (LV Dysfunction)	2	IS-IN, INF	2
IS, REV, ART	1	REV, INF	2
NL, ART	6	IS-IN, LVD	2
INF, ART	6	IS, ART, LVD	1
ART, REV	1	IS, REV	3
IS, INF	9	INF, REV	1
Sub-total	230	Sub-total	37
Total number of correct records	267		

4. Preparation of the cardiac SPECT data

4.1. Description of SPECT data

As said before each patient study contains two three-dimensional cardiac SPECT image sets of the LV. A cardiologist diagnoses, say, Ischemia, Infarct or Artifact, by comparing these two images. Evaluation of the images is a highly subjective process, with a great potential for substantial variability [9]. We use procedure similar to one described by us in [3] to analyze the images. The raw image data taken from multiple planar views are processed by filtered back-projection to create a three-dimensional image. Each of these 3D images is displayed as three sets of two-dimensional images. These 2D images correspond to the following sections of the LV myocardium: short axis view, horizontal long axis view and vertical long axis view. From these 2D sets of images a cardiologist selects five slices for each study that constitute the final report of the Yale system. Definitions of these five slices and regions of interest are shown in Fig. 3.

The five slices are chosen according to the following.

- Three slices for short axis view-one slice near heart's apex, one in middle of the LV and one near the heart base.
- One slice corresponds to the center of the LV cavity for horizontal long axis view.
- One slice corresponds to the center of the LV cavity for vertical long axis view.

Each of these five images is divided into four or five regions of interest (ROI), along the LV myocardium. As a result for each study there are 22 regions of interest. The cardiologist evaluates appearance and count in each of these regions. Comparison between corresponding ROIs in stress and rest study is performed. Partial diagnoses are made for each ROI by the cardiologist; they are classified into seven categories: Normal, Reversible, Partially Reversible, Artifact, Fixed, Equivocal and Reverse Redistribution. Cardiologist makes the overall diagnosis based on the partial diagnoses. The overall diagnosis is classified into

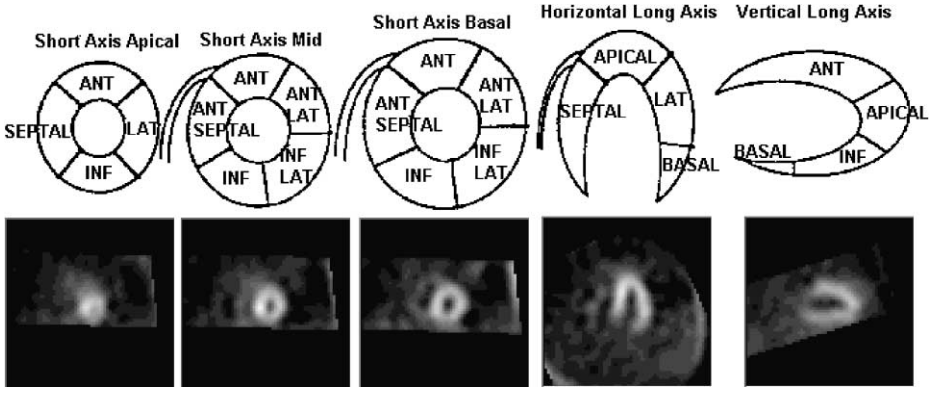

Fig. 3. Slices and regions of interest. From the left first three images are in short axis view, the fourth in horizontal long axis view, the last in vertical long axis view. At the bottom five slices chosen according to the definition.

eight categories: Normal, Ischemia, Infarct, Infarct and Ischemia, Artifact, Equivocal, Reverse Redistribution, and the LV Dysfunction. Some of these categories may coexist, e.g. Ischemia and Artifact, Normal and Artifact etc. (see Table 1).

The cardiologist during diagnosing process compares cardiac SPECT images to his or her mental image of the normal LV. In case of automatic image analysis one of the most difficult tasks is to properly establish the location of the ROI within the SPECT image. This task is complicated by two factors.

- *LV defects*: Changes in perfusion of the LV myocardium affect the changes in brightness of the cardiac SPECT image. When perfusion is low the corresponding brightness (count) is also low; as a result part of the LV myocardium can be absent from the image. The cardiologist deals with this fact by mentally reconstructing missing parts of the myocardium.
- *Artifacts*: The most common artifact for Thallium 201 tracer is decreased brightness of the LV myocardium; usually it is caused by breast tissue in females, or by diaphragm in males. Artifacts not only complicate localization of ROI but also presence of artifact in a region can lead to incorrect computer diagnosis because of the decreased brightness in the region.

In addition, SPECT quantitative information can be incorrect because of physical phenomena like radiation attenuation, scatter of radiation, 'partial volume effect', etc. [1]. All those factors can change the shape and visibility of the LV in the cardiac SPECT image; they are obviously cause of major problems for image analysis. After establishing the localization of ROI we compare corresponding regions in stress and rest studies, and counts in each region, against the same regions in the predefined normal LV model (described in Section 5.1).

4.2. Cleaning of the SPECT images

We manually looked for incomplete patient records, paying attention to missing partial and overall diagnoses, and information about patient sex. Each record with missing information was discarded.

The second part of data cleaning is cleaning the SPECT image database. We looked for incomplete image sets and for very poor quality images (without sufficient contrast). Each incomplete image set was discarded. Another problem was that pixel count values recorded in image files were in different ranges for different patient images. We found the solution for this problem by calibration of pixel values to the range from 0 to a maximal value of pixels in all images that are recorded for each patient. We also had to calibrate pixel values of every patient image to values in the range from 0 to 255. After cleaning we obtained a new database consisting of 267 patients covering all the diagnostic cases. The number of patients for each diagnostic case is shown in Table 1.

4.3. Construction of the new database

New database is created based on the initial SPECT database (after cleaning). This database contains the following modules:

1. Module with patient images consisting of 267 patients, 10 2D images per patients (five images for rest and five for stress study). Images are recorded in a predefined structure according to the encrypted date of the SPECT study and the encrypted patient hospital ID.
2. Module with clinical patient records, including encrypted patient hospital ID, sex, age, weight, height, encrypted date of the SPECT study and complete diagnosis (22 partial diagnoses and 1 overall diagnosis). It is used for evaluation of the results.
3. Module containing additional images: pattern-masks images for model of the normal LV, and images chosen for evaluation of the normal LV model. The total number of image files in this module is 225.

Images included in module 1 are chosen from cardiac SPECT images according to the procedure that is performed by a cardiologist during the diagnostic process as described in Section 4.1. Process of choosing the 10 images is done using our software that makes it possible to view the entire set of the 2D images using sliders. The user can freeze and save the chosen set of 10 images into 10 separate files within the predefined directory structure.

The entire record (including images) for one patient in the new database occupies about 51 kB of memory. The entire new database contains 2896 files of size of about 15 MB.

New database is created for the purpose of calculating the accuracy of the system. The system requires only currently diagnosed patient data to calculate diagnoses.

5. Data mining

We use image analysis in combination with machine learning tools to mimic a diagnostic process performed by a cardiologist. Automatic extraction of information from images is difficult because we need to model the human way of analyzing the images [8]; humans do it at the sub-conscious level.

To achieve it we first have to correctly establish each of the predefined ROI in the SPECT images. Next, we extract a set of features from the images to estimate brightness in each of the ROIs. We extract a single feature for each ROI for both stress and rest studies. Each feature is represented by a single number describing perfusion inside one of the ROI's. This approach is similar to the one described in [3]. As a result 44 features are used for classification of the LV perfusion. Based on them and using a set of production rules, we obtain partial diagnosis, one for each ROI. For generation of the rules for partial diagnoses we use two separate approaches: inductive machine learning CLIP3 algorithm [2,4], and the heuristic approach. After comparison of the results obtained by these two approaches we decided to keep the rules generated heuristically. The heuristic diagnostic algorithm very closely imitates cardiologist's diagnostic process. Then, using partial diagnoses as input we apply CLIP3 algorithm to generate the overall diagnostic rules. The diagram of the entire diagnostic process is shown in Fig. 4.

5.1. Image analysis algorithm

The goal of the image analysis algorithm is to extract 44 features from cardiac SPECT studies. Each feature is a single number that measures radioactive counts that represents

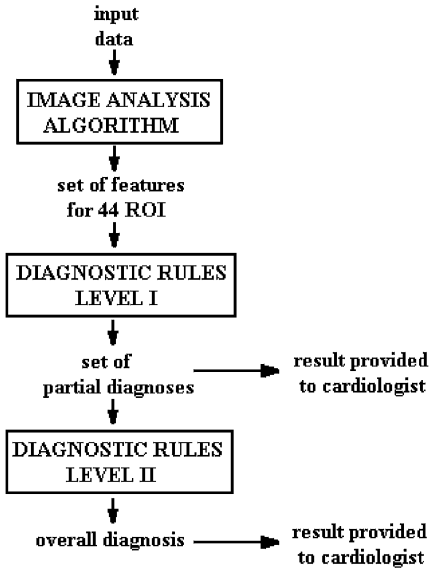

Fig. 4. Schematic diagram of the proposed diagnostic algorithm.

perfusion in the LV muscle in a specified ROI. This task is complicated by changes in appearance of the LV myocardium due to artifacts, actual LV defects and anatomical differences between patients. To deal with these problems two approaches are used.

1. We calculate one set of 44 features by using template-masks created directly from the diagnosed images. We use SPECT image of the currently analyzed patient to create template-mask, which defines position of all ROI in this image. This set is called the *basic-set*.
2. The template-masks of the normal LV model are used to calculate the second set of 44 features. The template-masks are stored in the database; they are part of the created model of the normal LV. This second set of features is called *emergency-set* — if some of the features cannot be correctly calculated using the first approach, then we use features calculated in the second approach.

It is important to remember that features from the *basic-set* are more accurate since they are derived directly from input images. The *emergency-set* has a full set of correctly extracted features, which we can use when the first approach fails. As a result there are two sets of 44 features to operate on. From these two sets a final set of 44 features is created: simply the features from the *basic-set* that were calculated incorrectly by the algorithm are replaced with features taken from the *emergency-set*.

The inputs for the image analysis algorithm are: 10 cardiac SPECT images chosen using the rules described in Section 4.1, and information about the patient's sex. Due to the anatomical differences of LV myocardium between males and females the diagnostic process is automated for the two sexes separately. We also built a separate model of the normal LV for males and females.

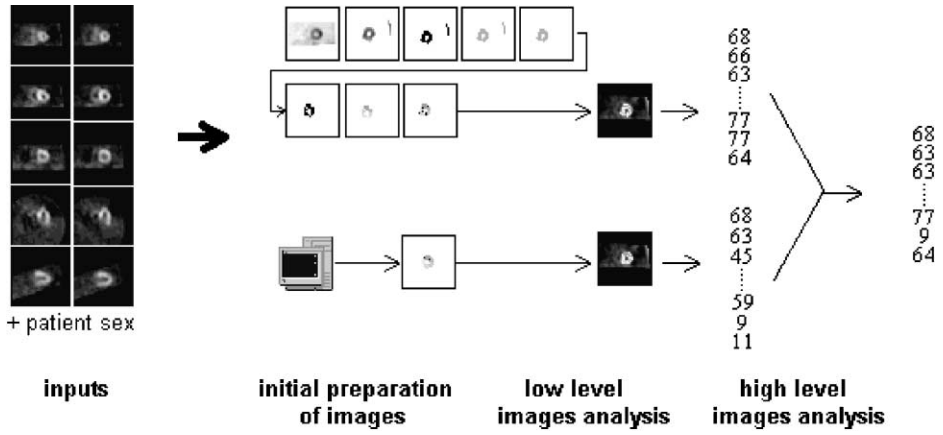

Fig. 5. Block diagram of image analysis algorithm.

Image analysis algorithm is divided into three parts: initial preparation of images, low-level image analysis, and high-level image analysis. Fig. 5 shows diagram of the algorithm.

Initial preparation of images includes operations like color inversion, background standardization, and binarization. Two threshold values are used and we consider different binarization methods in the initial preparation step. These parameters are subject to system optimization, see Section 5.3. Goal of the initial preprocessing is to reduce the amount of information by leaving only few objects of interest in the image.

Low-level image analysis includes operations like labeling, indexing, contour extraction, division of object into regions, calculation of object parameters, etc., see Fig. 6. The goal of this step is to process objects on the image (objects represent the LV myocardium) to obtain image with just one object divided into a number of ROIs.

High-level image analysis includes operations like systematized labeling, calculation of the features, registration (registrations is a process of scaling, shifting and rotating the object to reach the desired position), calculation of output features from *basic* and *emergency* features sets, see Fig. 7. The goal of the high-level image analysis is to calculate the set of features, which represent the counts in each of the ROIs.

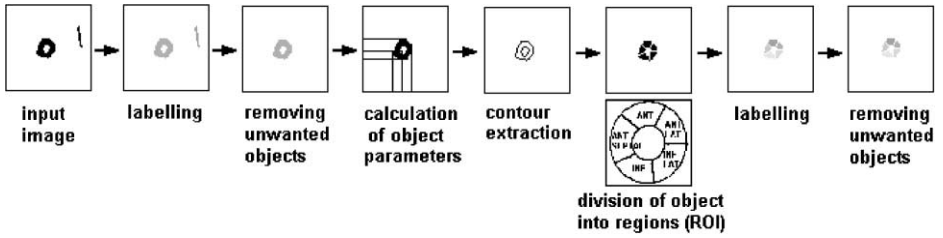

Fig. 6. Important operations performed during low-level image analysis.

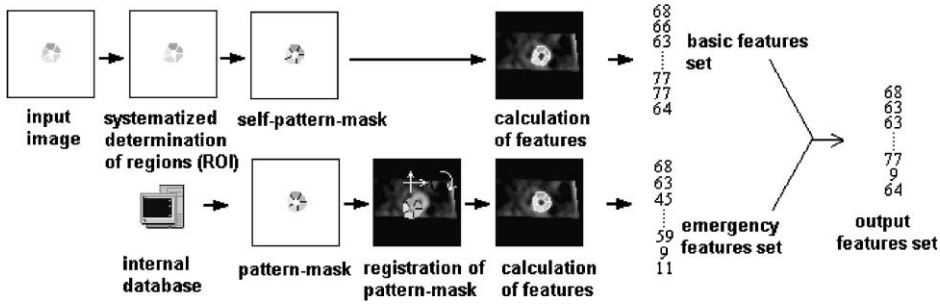

Fig. 7. Important operations performed during high-level image analysis.

Each feature is calculated according to the following equation:

$$\text{feature}_i = \frac{\sum_R C(m,n)}{N(R)} \times \frac{100}{m_j}\ (\%)$$

where R is current region (one of the ROI), $N(R)$ the number of pixels belong to R, $C(m, n)$ the value of pixel belong to R with coordinates (m, n) (brightness of pixel), m_j the maximal value of pixel taken from two corresponding images of LV, from rest and stress study, j [1–5] the index of current image, feature_i the value of feature, i the index of feature in features set.

Features are calculated as an average count of pixels located in the current ROI, which is defined by the template-mask. This value is calibrated as a percentage value of the highest brightness of two corresponding images, from rest and stress studies. We considered different methods for calculation of the features — this was the subject of the system optimization, see Section 5.3.

To create the model of the normal LV we took from the database 20 cases diagnosed earlier by cardiologist as Normal (10 females and 10 males). The model is build by averaging 10 image sets of male patients and 10 image sets of female patients, respectively. Model of the normal LV contains information about localization of all ROIs, it consist of 10 images corresponding to 10 slices chosen by the cardiologist during the diagnostic process. Fig. 8 shows template-model of the normal LV for females.

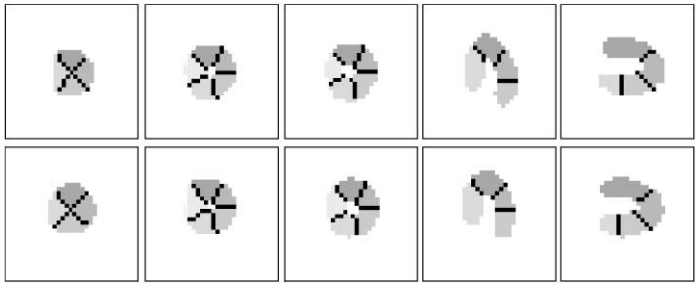

Fig. 8. Template-model of normal LV for female patients.

The output of the image analysis algorithm is a set of 44 features representing perfusion in 22 ROIs, for stress and rest studies, respectively. Based on these features we generate rules for 22 partial diagnoses; each partial diagnosis describing perfusion in one of the ROIs.

One of the advantages of this algorithm is that it produces correct set of features, even for very noisy images, or images showing strong abnormality. The algorithm is able to work with images in which parts of the myocardium are not apparent, or with images that have a very low contrast. Another advantage of the algorithm is that it includes mechanisms for checking the values of the features. We check if each of the ROIs is correctly established and also the correctness of the sequence of the established ROIs.

Original images are used without enhancing image quality. The assumption was that any change to an original image can deform either the shape or count information inherent in cardiac SPECT images, and thus should be avoided.

5.2. Diagnostic rules for partial diagnoses

To generate rules for partial diagnoses we use the CLIP3 algorithm, and a heuristic approach mimicking the cardiologist's diagnostic process. After both sets of rules are generated we compare the results. As a reference point cardiologist's partial diagnoses were used. We decided to keep heuristic rules because they gave us slightly better results. These rules have very specific structure and are based on comparison of values between features in the corresponding ROIs. An example rule is

$$\text{IF feature}_i > (\text{feature}_j + \text{threshold}) \text{ THEN} \ldots$$

Most, if any, machine learning algorithms, are not able to generate this type of rules. Below the process of generating diagnostic rules using the heuristic approach is described.

We look only for two categories of partial diagnoses: *Normal* (NL) — which corresponds to Normal diagnosis performed by a cardiologist and *Abnormal* (ABN) — which corresponds to six diagnoses: Reversible, Partially Reversible, Artifact, Fixed, Equivocal and Reverse Redistribution.

We also know that cardiologist's diagnosis of cardiac SPECT images has the following properties.

- Cardiologists compare perfusion in the corresponding regions of myocardium in stress and rest studies, to evaluate if there is a defect in this particular region.
- They also compare perfusion across all regions in the diagnosed image, to find if there is a difference in perfusion between the regions in the entire cross-section of the heart muscle.

Using this knowledge we generate rules that follow this way of reasoning. All threshold values are initially estimated heuristically and then changed during the optimization process, see Section 5.3. Below we show the rules.

$$\text{Rule 1}: \quad \text{IF } [\text{feature\_stress}_i - \text{feature\_rest}_i] < \text{threshold THEN Region}_i \text{ ABN},$$
$$\text{where } i \text{ is the number of ROI}$$

This rule compares features representing perfusion in corresponding ROIs from stress and rest studies, and by using the threshold decides if the difference is big enough to

classify the region as Abnormal. For normal patients perfusion in stress should be bigger than perfusion in rest for the same region of the heart muscle. Threshold value is calculated using values of all features from the same cross-section of the myocardium for both rest and stress studies. The value of the threshold is adjusted for each case and each heart cross-section due to differences in contrast.

$$\text{Rule 2}: \quad \text{IF } [\text{feature\_rest}_i < 50 \text{ AND feature\_strees}_i < 50] \text{ THEN Region}_i \text{ ABN}$$

This rule checks if perfusion in stress and rest studies is high enough to classify the region as Normal. If perfusion is low then the value of feature is low and when perfusion is below certain threshold we consider the corresponding region as Abnormal. Threshold value is established as 50 — it depends on the formula used to calculate features. Regions with features values below the threshold are considered as Abnormal.

$$\text{Rule 3}: \quad \text{IF } [\text{feature}_i < \text{mean}_j - \text{threshold}] \text{ THEN Region}_i \text{ ABN},$$
where i is the number of features, j the number of heart cross-section,
mean_j the mean value of features along entire cross-section (entire image)

This rule compares appearance of perfusion in the entire cross-section of the heart muscle with perfusion inside one of the ROIs. If the difference between perfusion in analyzed ROI is below the mean value of the entire cross-section minus the value of the threshold, then this region is considered as Abnormal. Threshold values are different for features representing regions from stress (threshold = 14) and rest (threshold = 11) studies. The difference results from our knowledge that perfusion of normal patients in stress should be higher than in rest. The reason for this comparison is in a situation when low perfusion will occur only in one, or few regions of the heart muscle, and it is lower than perfusion in the surrounding regions, but the same when comparing perfusion between stress and rest studies. If the difference is greater then a certain threshold the region is classified as Abnormal; this is in case where both Rules 1 and 2 were not able to recognize it.

$$\text{Rule 4}: \quad \text{IF } [\text{feature\_stress}_i - \text{feature\_rest}_i] < \text{threshold THEN Region}_i \text{ ABN},$$
where i is the number of ROIs from the specific part of the LV

This rule was found useful since it improved accuracy of the algorithm during the optimization process. The structure of this rule is identical to the Rule 1, but it is used only after Rules 1, 2 and 3 have been used (we have initial classification for all ROIs into NL and ABN categories). The value of the threshold is much lower then the value used in Rule 1. The reason for Rule 4's re-diagnosing is that cardiologists have tendency to diagnose the entire parts of the heart muscle (e.g. anterior or inferior part) with the same category label. For example, if anterior part of the heart near the apex is diagnosed with low perfusion then most likely low perfusion will occur also in the entire anterior part of the heart muscle (mid and base). Thus, we define five different parts of the heart muscle (A, B, C, D, E) as being connected. They are shown in Table 2. In order to include this anatomical dependency all the ROIs belonging to that part of the heart muscle, in which we found at least one Abnormal region using the first three rules, are re-diagnosed. To achieve that Rule 4 with low threshold value (threshold = 4) is used.

Table 2
Definition of anatomically connected parts of the LV used to re-diagnose partial diagnoses with Rule 4

Cross-section	A	B	C	D	E
Short Apical	ANT	LAT	LAT	INF	SEPTAL
Short Mid	ANT	ANT-LAT	INF-LAT	INF	ANT-SEPTAL
Short Basal	ANT	ANT-LAT	INF-LAT	INF	ANT-SEPTAL

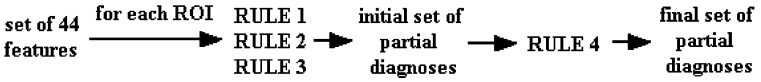

Fig. 9. Process of obtaining partial diagnoses from 44 features extracted from SPECT images.

The entire process of converting features extracted from cardiac SPECT images, using the four diagnostic rules into partial diagnoses is shown in Fig. 9.

5.3. Optimization of the system and partial diagnoses results

After rules for partial diagnoses are generated we optimize important parameters to achieve the best accuracy of diagnosing. To perform this task we choose group of 24 patients (training data set) from the new database that includes all diagnostic cases: three patients for each of eight groups for overall diagnoses, as described in Section 4.1. Using this set, the following parameters of the system were adjusted.

- Binarization method and binarization threshold value in the image analysis algorithm.
- Threshold values used to produce the final set of features from *basic-set* and *emergency-set* of features.
- Formula to calculate a feature value for each ROI (four different formulas were considered).
- Threshold values used in diagnostic rules for partial diagnoses.

The parameters were adjusted one at a time. We choose the value corresponding to the best performance of the system, and then switch to optimize another parameter. In each step when change in parameter value gives improvement in performance then this new value is kept, otherwise the old value is kept. We also assumed certain ranges for parameter adjusting.

Three accuracy criteria are used: ALL (performance of the rules versus 22 partial diagnoses for the entire training data set as done by cardiologists), NL (performance of the rules for partial diagnoses of patients who were diagnosed by cardiologists as Normal) and ABN (performance of the rules, but only for partial diagnoses of those who were diagnosed by cardiologists as Abnormal). The values in tables represent percentage of correct partial diagnoses. Table 3 shows the results before and after optimization process.

After the optimization process the rules are validated on all 267 diagnostic cases from the new database.

Table 3
Comparison of system accuracy for partial diagnoses calculated before and after optimization

Criterion	ALL (%)		NL		ABN	
	%	Correct/ total	%	Correct/ total	%	Correct/ total
Before optimization (24 patients)	67.99	359/528	65.66	262/399	73.69	95/129
After optimization (24 patients)	84.09	444/528	87.22	348/399	74.42	96/129

In addition, we analyzed the results for groups of patients defined as having different overall diagnoses. Within these groups we analyzed accuracy of rules for partial diagnoses, as well as for the entire validation data set. The results are shown in Table 4. The accuracy of the system for partial diagnoses is about 81%. For partial diagnoses evaluated by cardiologists as Normal, the system correctness is about 85%, and for those evaluated by cardiologists as Abnormal, the system correctness is about 72%. The results for individual diagnostic cases have the same level of correctness except for ABN patients diagnosed as NL; this is a result of artifacts. These results were obtained using a database of 267 cases, each case includes 22 partial diagnoses, which gives 5874 partial diagnoses in the validation data set. Because of the size of the database the results are quite reliable.

5.4. Generation of diagnostic rules using CLIP3 algorithm

The next task is to use partial diagnoses to come up with the overall diagnosis, which describes perfusion of the entire LV muscle. CLIP3 algorithm [4] is used to generate a set of diagnostic rules. We choose this algorithm because it was successfully used in another application with cardiac SPECT images [6] and performed as well as C4.5 and CN2 machine learning algorithms [5]. The CLIP3 algorithm is a descendent of the CLILP2 algorithm [2]. It is a hybrid of decision tree algorithms and rule-based algorithms [4]. CLIP3 uses IP model to select both key attributes and to generate the rules. There are three thresholds used by the CLIP3 algorithm [5]. The purpose of these thresholds is to control the complexity of the generated solutions, to decide when to stop forming rules, and to

Table 4
Results for partial diagnoses obtained using validation data set of 267 patients

Criterion	ALL		NL		ABN	
	%	Correct/ total	%	Correct/ total	%	Correct/ total
Overall diagnosis: NL (49 patients)	89.52	965/1078	91.76	946/1031	40.43	19/47
Overall diagnosis: IS (49 patients)	82.32	489/594	81.41	346/425	84.61	143/169
Overall diagnosis: INF (49 patients)	81.53	843/1034	86.43	618/715	70.53	225/319
Overall diagnosis: ART (49 patients)	81.52	556/682	88.17	462/524	59.49	94/158
Overall diagnosis: IS-IN (49 patients)	77.65	820/1056	78.06	459/588	77.14	361/468
Entire new DB (267 patients)	81.34	4778/5874	84.98	3581/4214	72.11	1197/1660

avoid creation of too many rules possibly describing noisy data. By using them we generate different sets of diagnostic rules.

Two databases are used to generate different sets of diagnostic rules using CLIP3 algorithm. They are later analyzed using multi-criterion evaluation and the best database is chosen for later use.

- Database of partial diagnoses (DPD) consist of 267 patient cases; it is created using previously described rules. Every record in this database consists of 22 partial diagnoses (each partial diagnosis is binary: Normal or Abnormal).
- Database of features (DF) extracted from cardiac SPECT images, as described before. Each feature is a number describing perfusion in one of the ROIs; it has an integer value in the range [0,100].

As said above our goal is to generate rules for the overall diagnosis, for two categories: Normal (NL) and Abnormal (ABN), which correspond to seven overall diagnoses: Ischemia, Infarct, Infarct and Ischemia, Equivocal, Reverse Distribution, Artifact and LV Dysfunction.

The new database contains complete diagnostic records, including 22 partial diagnoses and the overall diagnosis, as performed by cardiologists. Cases (examples) classified as Normal are called positive examples, and those classified as Abnormal are called negative examples. Table 5 shows details for both databases.

In the first step, the databases are divided into two parts: training and validation sets. Each set of rules generated from training data set is tested on validation data set.

Since both databases have only 55 positive examples (Normals) and 212 negative examples (Abnormals) two different divisions of data into training and validation data sets for each database are created. The reason for doing that is that better rules are generated when number of positive examples is approximately equal to the number of negative examples. Table 6 shows training and validation data sets. Because of the division we are

Table 5
Description of databases used to generate diagnostic rules for overall diagnoses

Database	Number of positive examples	Number of negative examples	Number of examples	Number of attributes	Values of attributes
DPD	55	212	267	22	Binary
DF	55	212	267	22	Continuous

Table 6
Training and validation data sets used to generate diagnostic rules

	Training data set		Validation data set		Total examples
	Number of positive examples	Number of negative examples	Number of positive examples	Number of negative examples	
DPD1	40	40	15	172	267
DPD2	40	162	15	50	267
DF1	40	40	15	172	267
DF2	40	162	15	50	267

Table 7
Possible outcomes from verification test and the verification test criteria

	Test result positive	Test result negative	Verification test criteria
Hypothesis positive	TP	FN	Sensitivity $= (TP/(TP + FN)) \times 100\%$, specificity $= (TN/(TN + FP)) \times 100\%$
Hypothesis negative	FP	TN	Predictive accuracy $= ((TP + TN)/(TP + TN + FP + FN)) \times 100\%$

able to directly compare the rules generated using different databases and divisions into training and validation data sets.

In order to evaluate the ''goodness'' of generated rules we check their ability to generalize on new, unseen validation data. The measures used to evaluate the rules are described below.

5.4.1. Verification test

In verification test there are three evaluation criteria [5]. Possible outcomes of a verification test, and the criterion definitions are shown in Table 7.

Sensitivity measures how many of positive examples from validation data set are correctly recognized.

Specificity measures how many of negative examples from validation data set are correctly recognized — they were excluded or not recognized as positive examples.

Predictive accuracy measures how many examples from entire validation data set are correctly recognized.

Using the verification test, the chosen best set of rules is the one that gives high values for all three criteria. The study performed by in [17] shows the importance of the trade-off between the sensitivity and specificity measures.

One additional criterion is also used: *number of rules* generated. The preferred situation is when there is a fewer number of rules generated.

Table 8 compares best sets of rules chosen for the four cases (chosen from all the rules generated using different values of CLIP3 parameters).

DPD1 is the set for which sensitivity and specificity values are close to each other and the number of corresponding rules is small. This set of rules has almost 80% predictive accuracy on the entire validation data set.

Table 8
Comparison of results obtained for the four considered databases using validation data set

Best sets of rules	Sensitivity		Specificity		Predictive accuracy		Number of rules
	%	Correct/total	%	Correct/total	%	Correct/total	
DF2	66.67	10/15	88.00	40/50	83.08	54/65	2
DF1	73.33	11/15	77.32	133/172	77.00	144/187	1
DPD2	66.67	10/15	74.00	37/50	72.31	47/65	6
DPD1	73.33	11/15	80.23	138/172	79.68	149/187	2

Table 9
Comparison of results obtained for validation data set using different modes for solving the IP model for the DPD1 database

Method for solving IP model	Sensitivity		Specificity		Predictive accuracy		Number of rules
	%	Correct/ total	%	Correct/ total	%	Correct/ total	
Standard	73.33	11/15	80.23	138/172	79.68	149/187	2
Semi-random	80.00	12/15	84.30	145/172	83.96	157/187	3

After choosing this set of rules further study using the DPD1 database was performed. The method of calculating solution of the IP model was modified. Original method is described in [4]. The new method searches through set of columns in a probabilistic way choosing one set of columns as a solution. We call it semi-random method for solving the IP model. This method was used to generate new set of diagnostic rules for our problem. Table 9 compares best sets of rules generated for the DPD1 database using standard and the new semi-random mode for solving the IP model. As can be seen the new method results in significant increase of the accuracy of the rules compared to the accuracy of the rules generated using the standard mode. The rules finally chosen for the overall diagnosis are shown below:

Rule 1 : IF[F3 = NL and F7 = NL and F11 = NL and F13 = NL and F15 = NL
and F16 = NL and F17 = NL and F18 = NL and F21 = NL]
THEN Overall Diagnosis = NL (covers 26 out of 40 positive examples)

Rule 2 : IF[F1 = NL and F4 = NL and F5 = NL and F6 = NL and F8 = NL
and F9 = NL and F10 = NL and F15 = NL and F16 = NL
and F17 = NL and F18 = NL and F19 = NL and F22 = NL]
THEN Overall Diagnosis = NL (covers 17 out of 40 positive examples)

Rule 3 : IF[F2 = NL and F3 = ABN and F5 = NL and F16 = NL and F22 = NL]
THEN Overall Diagnosis = NL (covers 4 out of 40 positive examples),

where features F1, . . . , F22 represent partial diagnoses corresponding to ROIs.

To summarize, the rules for partial diagnoses are generated using heuristic approach and the rules for the overall diagnosis were generated using the CLIP3 algorithm. Rules for partial diagnoses use as an input cardiac SPECT images, and patient's sex information. The rules for the overall diagnosis use as an input only partial diagnoses.

6. Evaluation of the discovered knowledge

Several iterations were performed during the described knowledge discovery process. These iterations included reviewing and changing image analysis algorithms, optimizing the system to evaluate and adjust its parameters, and searching for different sets of the

overall diagnostic rules using different input variables. Our goal during this iterative process was to improve accuracy of the algorithm so that it is as close as possible to diagnoses performed by a cardiologist. This iterative reviewing process resulted in significant increase of accuracy: about 16% increase for the case of evaluation of partial diagnoses (see Table 2) and about 4% increase for the overall diagnosis (see Table 9).

6.1. Analysis of the results

Visual interpretation and diagnosing of clinical images is highly observer-dependent. Thus, there is a need to create computer-based tools, which can assist a cardiologist during the diagnostic process. Several observations need to be made about the following results.

- Partial diagnoses are 81% accurate and the overall diagnoses are 84% accurate.
- The results are supported by using database of 267 patients.
- A small number of diagnostic rules for both partial and overall diagnoses was generated.
- The generated diagnostic rules have anatomical and physiological interpretation; they can be evaluated by experts.
- The results are influenced by the fact that cardiologist interpretation of the data was inconsistent. Different cardiologists interpreted SPECT studies and depending on their experience they were more or less accurate, but we still used them as a gold standard to train the algorithm.

7. Using the discovered knowledge and discussion

The newly discovered rules need to be evaluated by medical professionals on much larger data sets before they can be used in a hospital setting. The knowledge discovery process used appears to be very practical. We have shown that iterative process used for the creation of the new knowledge, or diagnostic rules, can significantly increase accuracy of the entire system.

In this project, we processed cardiac SPECT images to extract important features. Using these features we generated diagnostic rules using heuristic and inductive machine learning approaches. Both of these approaches resulted in simple, easy to understand, and highly accurate diagnostic rules. The system can be used as an assistant tool by cardiologists to help them to make more consistent diagnosis of cardiac SPECT studies. In the future, we plan to use more patient information to improve system accuracy.

References

[1] Blokland JAK, Reiber JHC, Pauwels EKJ. Quantitative analysis in single photon emission tomography (SPET). Eur J Nuclear Med 1992;19:47–61.

[2] Cios KJ, Liu N. An algorithm which learns multiple covers via integer linear programming, Part I — the CLILP2 algorithm. Kybernetes 1995;24(2):29–50.

[3] Cios KJ, Goodenday LS, Shah KK, Serpen G. A novel algorithm for classification of SPECT images of a human heart. In: Proceedings of the CBMS '96, 1996 Jun 1–5; Ann Arbor.

[4] Cios KJ, Wedding DK, Liu N. CLIP3: cover learning using integer programming. Kybernetes 1997;26(4/5):513–36.
[5] Cios KJ, Pedrycz W, Swiniarski R. Data mining methods for knowledge discovery. Dordrecht: Kluwer Academic Publishers, 1998. http://www.wkap.nl/book.htm/0-7923-8252-8.
[6] Cios KJ, Teresinska A, Konieczna S, Potocka J, Sharma S. Diagnosing myocardial perfusion from SPECT bull's-eye maps — a knowledge discovery approach. IEEE Eng Med Biol Magazine, special issue on Med Data Mining Knowledge Discovery 2000;19(4):17–25.
[7] Corbett JR. Gated blood-pool SPECT. In: DePuey ED, Berman DS, Garcia EV, editors. Cardiac SPECT imaging. New York: Raven Press, 1995.
[8] Crevier D, Lepage R. Knowledge-based image understanding systems: a survey. Comput Vision Image Understanding 1997;67(2):161–85.
[9] Cuaron A, Acero AP, Cardenas M, et al. Interobserver variability in the interpretation of myocardial images with Tc-99m-labeled diphosphonate. J Nuclear Med 1980;21:1.
[10] Cullom SJ. Principles of cardiac SPECT. In: DePuey ED, Berman DS, Garcia EV, editors. Cardiac SPECT imaging. New York: Raven Press, 1995.
[11] Faber TL, Akers MS, Peshock RM, Corbett JR. Three-dimensional motion and perfusion quantification in gated single-photon emission computed tomograms. J Nuclear Med 1991;32:2311.
[12] Faber TL, Cooke CD, Peifer JW, et al. Three-dimensional displays of left ventricular epicardial surface from standard cardiac SPECT perfusion quantification techniques. J Nuclear Med 1995;36:697–703.
[13] Fayyad U, Piatetsky-Shapiro G, Smyth P, Uturusamy R. Advances in knowledge discovery and data mining. New York: MIT Press, 1996.
[14] Francisco DA, Collins SM, Go RT, et al. Tomographic thallium-201 myocardial perfusion scintigrams after maximal coronary vasodilation with intravenous dipyridamole: comparison of qualitative and quantitative approaches. Circulation 1982;66(2):370.
[15] Garcia EV, Ezquerra NF, DePuey EG, et al. An artificial intelligence approach to interpreting thallium-201 3 dimensional myocardial distributions (abstract). J Nuclear Med 1986;27:1005.
[16] Goris ML, Boudier S, Briandet PA. Two-dimensional mapping of three-dimensional SPECT data: a preliminary step to the quantification of thallium myocardial perfusion single photon emission tomography. Am J Physiol Imaging 1987;2:176–80.
[17] Kukar M, Kononenko I, Groselj C, Kralj K, Fettich J. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. Artif Intell Med 1999;16(1):25–50.
[18] Puentes J, Garreau M, Roux C, Coatrieux JL. Towards dynamic cardiac scenes interpretation based on spatial-temporal knowledge. Artif Intell Med 2000;19(2):155–83.
[19] P Sacha J, Cios KJ, Goodenday LS. Issues in automating cardiac SPECT diagnosis. IEEE Eng Med Biol Magazine, special issue on Med Data Mining Knowledge Discovery 2000;19(4):78–88.
[20] Van Train KF, Garcia EV, Cooke CD, Areeda J. Quantitative analysis of SPECT myocardial perfusion. In: DePuey ED, Berman DS, Garcia EV, editors. Cardiac SPECT imaging. New York: Raven Press, 1995.