Bayesian learning for cardiac SPECT image interpretation

Jarosław P. Sacha[a,*], Lucy S. Goodenday[b],
Krzysztof J. Cios[c,d,e,f]

[a]*Health Care Research Center, Corporate R&D, Procter & Gamble, 8700 Mason-Montgomery Road, Mason, OH 45040-9462, USA*
[b]*Medical College of Ohio, 1192 Hospital Building, 3000 Arlington Avenue, Toledo, OH 43614, USA*
[c]*University of Colorado at Denver, Denver, CO, USA*
[d]*University of Colorado at Boulder, Boulder, CO, USA*
[e]*University of Colorado Health Sciences Center, Boulder, CO, USA*
[f]*4cData, LLC, Golden, CO, USA*

Abstract

In this paper, we describe a system for automating the diagnosis of myocardial perfusion from single-photon emission computerized tomography (SPECT) images of male and female hearts. Initially we had several thousand of SPECT images, other clinical data and physician–interpreter's descriptions of the images. The images were divided into segments based on the Yale system. Each segment was described by the physician as showing one of the following conditions: normal perfusion, reversible perfusion defect, partially reversible perfusion defect, fixed perfusion defect, defect showing reverse redistribution, equivocal defect or artifact. The physician's diagnosis of overall left ventricular (LV) perfusion, based on the above descriptions, 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. Because of the complexity of the task, we decided to use the knowledge discovery approach, consisting of these steps: problem under-standing, data understanding, data preparation, data mining, evaluating the discovered knowledge and its implementation. After going through the data preparation step, in which we constructed normal gender-specific models of the LV and image registration, we ended up with 728 patients for whom we had both SPECT images and corresponding diagnoses. Another major contribution of the paper is the data mining step, in which we used several new Bayesian learning classification methods. The approach we have taken, namely the six-step knowledge discovery process has proven to be very

---

* Corresponding author. Tel.: +1-513-622-1556; fax: +1-425-944-4469.
E-mail addresses: jarek@ieee.org (J.P. Sacha), lgoodenday@mco.edu (L.S. Goodenday), krys.cios@cudenver.edu (K.J. Cios).

successful in this complex data mining task and as such the process can be extended to other medical data mining projects.

1. Introduction

In this paper, we follow a knowledge discovery process in which we use a new Bayesian learning method as a data mining step for interpretation of cardiac single-photon emission computerized tomography (SPECT) images. The paper starts with an overview of cardiac SPECT imaging and overview of the cardiologist's way of interpreting them using a 22-feature Yale system. We describe data collection, warehousing and verification of data quality. Section 2 is devoted to the problem of feature extraction from the SPECT images of the left ventricle (LV). As part of this effort a static three-dimensional model of the normal LV is constructed for male and female patients, respectively. A major section is devoted to image registration and extraction of features from the registered images. We also describe Bayesian network classifiers (BNCs) and introduce a new family of Bayesian classifiers. Results of using the classifiers for partial classification of the LV perfusion defects for each of the 22 regions, and for the overall classification of the LV perfusion, are discussed.

2. Understanding the problem domain

2.1. SPECT imaging

SPECT of myocardial perfusion is accomplished by acquiring multiple planar (two-dimensional) gamma camera views of the heart after the intravenous injection of a radionuclide that traces myocardial blood flow. Tomographs are constructed from the multiple planar projections by filtered back-projection. These tomographs are then re-aligned to three cardiac planes: horizontal short axis, vertical long axis and horizontal long axis.

There are many potential artifacts that plague this technique, which may arise from the raw data itself (such as body habits, soft tissue attenuation, motion and ''feed through'' of radiation from adjacent structures), from the acquisition technique (for example, timing, collimation, center of rotation and crystal uniformity) or from the reconstruction process itself (pre- and post-filtering, alignment). In this paper, we ignore these pitfalls and assume that the tomographs are of the best possible technical quality. It must be emphasized, however, that any automated process of interpretation relies on the artifact-free validity of the tomographs.

Cardiac SPECT is a set of three-dimensional images that visualizes distribution of radioactive counts within the myocardium and the surrounding structures [10]. Images represent radioactive count densities within the heart muscle that are proportional to
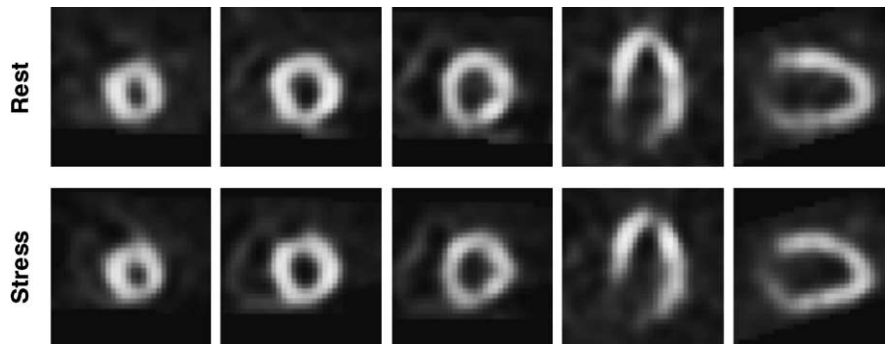
Fig. 1. Paired stress and rest images in each of the cardiac planes.

muscle perfusion, in particular of the LV, which is normally thicker than other cardiac structures. Two studies are performed after a patient is injected with a tracer: at rest (rest image) and at stress (after injection during maximal stress; stress image). The stress images are acquired after intervention that increases myocardial blood flow above that of the rest images; accomplished either by exercise or by administration of pharmacological agents. The studies are represented by two three-dimensional density images. Each tomographic study is displayed as paired stress–rest images in each of the cardiac planes (Fig. 1). Clinicians compare the two images to detect abnormalities in distribution of blood flow within the LV myocardium.

Visualization of the SPECT images is complicated by the fact that three-dimensional density images cannot be directly presented using contemporary display devices that produce two-dimensional pictures; a transformation must be performed that introduces reduction of information. There are two approaches: two-dimensional density images or three-dimensional surface rendering [20]. The first preserves most of the intensity information but the three-dimensional relations are only implicit. The second makes three-dimensional information explicit but density is represented indirectly through the shape of the three-dimensional surface and/or its color [15,16].

Typically, the LV is visualized as a series of two-dimensional intensity slices. When sliced perpendicular to the long axis of the LV, the view is called ''short axis''. Slices parallel to the long axis of the LV are called ''vertical long axis'' and ''horizontal long axis'' views (Fig. 1). Three-dimensional relations are only implicit in the views; it is up to the physician to mentally reconstruct them as a three-dimensional object.

Another family of two-dimensional visualization methods is based on projections in non-Cartesian coordinate systems. The three-dimensional LV is unwrapped on a two-dimensional plane by radial projection into spherical coordinates [21], or a combination of spherical and cylindrical coordinates [43]. They are referred to as bull's eye or polar map methods since they produce pairs of rest–stress round images (Fig. 8).

A number of three-dimensional surface rendering methods exist. They are occasionally used in association with gated blood-pool SPECT [7] to produce motion sequences of LV function. In this paper, we focus on static SPECT images of myocardial perfusion rather than on chamber geometry.

Several techniques exist to aid in the classification of the images; most of them are concerned with visualization. It has been shown, however, that interpretation of images by strictly visual techniques is fraught with error and inconsistency [9]. For this reason assistance in diagnosis through the use of computer-derived image display and quantitation is sought to decrease the variability in image interpretation [17]. One of few examples of automatic interpretation of SPECT images is the PERFEX expert system [14], which infers the extent and severity of coronary artery disease from perfusion distribution.

2.2. Cardiologist's interpretation process

Each patient study contains two three-dimensional SPECT image sets of the LV (one at rest and one at stress). Comparing the two allows diagnoses, like ischemia or infarct. Evaluation of the images is a highly subjective process with potential for substantial variability [9]. To analyze the images, we followed a procedure originally described in [5]. The raw image data taken from multiple planar views are processed by filtered back-projection to create a three-dimensional image. These three-dimensional images are displayed as three sets of two-dimensional images corresponding to the short axis view, horizontal long axis view and vertical long axis view.

From these two-dimensional views, the physician selects five slices for the final report. Three slices are selected from the short axis view; one slice is taken near the heart's apex, one at the middle of the ventricle and one near the heart's base. For each of the horizontal and vertical axis views, a single slice is selected corresponding to the center of the LV cavity. Each of these five images is subdivided into a number of regions of interest (ROIs), from four to five, along the walls of the LV, for a total of 22 regions (Fig. 2). The appearance of the LV and maximum count in each of the regions is evaluated. Corresponding ROI locations on the stress and rest images are then 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 analysis result, is concluded from the results of analysis in each of the ROIs. From the overall analysis the physician categorizes a
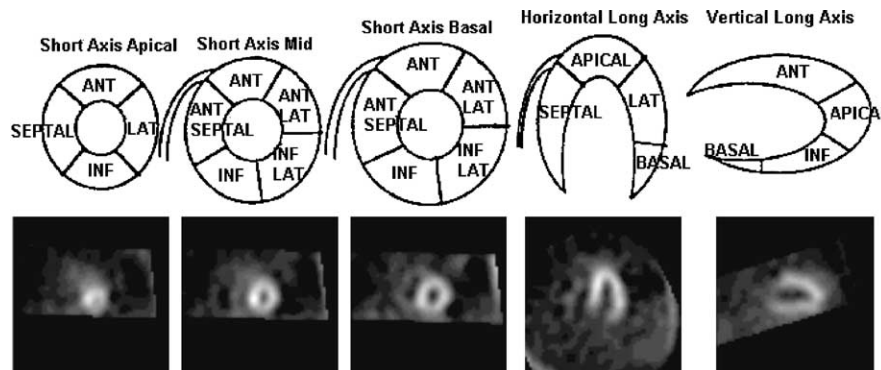

Fig. 2. The 22 ROIs within a LV (22 ROIs). The first three images correspond to short axis view slices, the last two to horizontal and vertical long axis views, respectively.
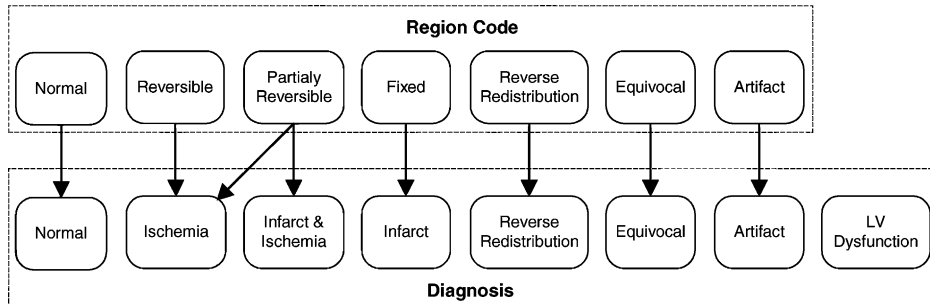
Fig. 3. Diagnosis of LV condition based on local classifications.

study as showing one or more of eight possible conditions: normal, ischemia, infarct and ischemia, infarct, reverse redistribution, equivocal, artifact or LV dysfunction (Fig. 3). Note that the classification LV dysfunction is included as a possible condition, 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 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, which is complicated by two factors that create a challenge for any algorithmic implementation.

- *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 the heart's anatomy and experience with cardiac SPECT imaging. However, this is a major challenge for computer algorithms.
- *Artifacts*: A common artifact in myocardial perfusion imaging is that of a decrease in radioactive 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 artifact may lead to false diagnosis since the decrease in counts may be erroneously taken for a real perfusion defect.

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.

3. Understanding the data

A prerequisite for a knowledge discovery process is data warehousing. Data warehousing refers to a number of activities involved in collection and cleaning of data to make

them available for online analysis and decision support. The objective is to organize the available data into a coherent database system and provide well-defined methods of efficient data access.

3.1. Original data

The SPECT data used here was collected at the Medical College of Ohio (MCO). The recorded data was entered manually into an Excel spreadsheet. Each row corresponds to a single patient visit—a SPECT procedure. One hundred and eighty-four parameters were recorded, including age, sex, height; information about the procedure and the nuclear cardiologist's interpretation of the SPECT images (by ROIs) and the overall perfusion classification, i.e. diagnosis. We had 4828 such records for review.

SPECT images were stored in a proprietary format, without database organization. We had 8142 SPECT image files. Typically there were six three-dimensional images and a number of auxiliary files per case. After cleaning it corresponded to 728 cases. Each set was stored in a separate directory. The following information was contained in each directory.

- Patient identification number, name and visit date (these were not given to us).
- Raw two-dimensional SPECT data for rest and stress studies (planar projections before three-dimensional reconstruction).
- Three-dimensional SPECT images corresponding to short, vertical long and horizontal long axis views of the heart, for rest and stress studies.

Not all directories contained complete datasets. All software for manipulation and visualization of SPECT images was written for this project [37].

3.2. Data warehousing

Data from the spreadsheet were converted to a relational database. The proprietary SPECT image file format had been reverse-engineered to the level that allowed for extraction of the critical information—the actual three-dimensional images. Software for automatic indexing of images was created and stored in the database table. The images themselves were stored outside of the database in a predetermined directory structure. The database design objective was simplicity of maintenance and ability to add new patient records and images easily as they become available. The database also stores data generated by various data mining activities, such as information about generated models of a normal LV and features extracted from the three-dimensional SPECT images. SQL query can be directly used to generate variants of datasets needed for the automation of diagnosis, e.g. learning BNCs.

3.3. Data selection—verification of the data quality

We semi-manually inspected the original data to eliminate typographical errors. The intention was not to modify the data unless the correction was straightforward. Rather, where possible, we constructed SQL queries to filter undesired records.

Table 1
Cases with complete sets of SPECT images and complete diagnosis data

Perfusion code	Female			Male			Total		
	No ART	ART	**Total**	No ART	ART	**Total**	No ART	ART	**Total**
NL	36	36	**72**	23	19	**42**	59	55	**114**
IS	17	4	**21**	21	6	**27**	38	10	**48**
INF	18	4	**22**	41	4	**45**	59	8	**67**
IS–IN	9	2	**11**	52	2	**54**	61	4	**65**
EQ	1	0	**1**	3	0	**3**	4	0	**4**
REV	1	1	**2**	0	1	**1**	1	2	**3**
LVD	2	0	**2**	0	0	**0**	2	0	**2**
	84	47	**131**	140	32	**172**	224	79	**303**
IS, IS–IN	0	0	**0**	1	0	**1**	1	0	**1**
IS, REV	1	0	**1**	2	1	**3**	3	1	**4**
IS, LVD	0	1	**1**	1	0	**1**	1	1	**2**
IS–IN, LVD	0	0	**0**	7	0	**7**	7	0	**7**
INF, IS	1	0	**1**	11	0	**11**	12	0	**12**
INF, IS–IN	0	0	**0**	2	0	**2**	2	0	**2**
INF, REV	0	0	**0**	3	1	**4**	3	1	**4**
INF, LVD	2	0	**2**	10	0	**10**	12	0	**12**
INF, LVD, REV	0	0	**0**	1	0	**1**	1	0	**1**
LVD, REV	0	0	**0**	1	0	**1**	1	0	**1**
	4	1	**5**	39	2	**41**	43	3	**46**
Total	88	48	**136**	179	34	**213**	267	82	**349**

Records were counted to estimate statistical validity of the expected results, i.e. if there were sufficient number of examples for each learning class. Table 1 shows the number of cases for each of the LV perfusion classifications (NL, normal; IS, ischemia; INF, infarct; IS–IN, ischemia and infarct; EQ, equivocal; REV, reversible redistribution; LVD, LV dysfunction; ART, artifact). At the top, we see records with single classification code and at the bottom the records that contain more than one classification code.

The images without the matching patient records were eliminated from the analysis. Next, image sets were checked for completeness, e.g. both studies present, and for quality of individual images, mostly related to sufficient contrast (photon count).

4. Extraction of features from SPECT images

Computer vision and machine learning techniques were used to mimic the diagnostic process of a physician. First, a set of attributes (features) must be extracted from the three-dimensional images for each case study. The most natural approach is to extract a single attribute feature corresponding to each ROI in both rest and stress images (see Fig. 2) as it was originally done in [5]. In the currently described work, we used more than a single attribute for each ROI, namely maximum photon count, mean count and median count. Then, to diagnose a patient, we used two approaches. One was to directly perform the

overall classification of LV perfusion using the extracted attributes. The second was to perform partial classification first, in each of the 22 regions using information from rest and stress images and then use these 22 intermediate classifications to come up with the overall classification.

It is difficult to automatically determine the ROIs directly from the three-dimensional images due to artifacts, actual LV defects and anatomical differences between patients. We approached the task of ROI detection by creating a model of a normal LV [37]; location of the regions is 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. Since the image may be matched with a number of models, the one with the highest correlation ratio was selected and used to locate slices and ROIs in the image. Regional perfusion was determined based on the count/intensity of the LV walls within the region. Even when the image and a model were correctly registered, the walls of the model and case under investigation would not always overlap exactly, thus compromising quality of the feature extraction process. For correct determination of the myocardial wall location, we used a method similar to the radial search used in SPECT bull's eye methods [21,43]. The model is used to determine the center of the LV 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 is based on the premise that counts within the LV wall are higher than in the surrounding areas.

Another critical issue was normalization of the image intensity range. Not only do counts vary significantly between patients, they are also different between the rest and stress images for the same patient. We normalized the numerical values as a percentage of the maximum count within the heart in a given three-dimensional image.

4.1. Model of the normal left ventricle

Physics-based deformable models are becoming popular for describing the LV [11] because they are supposed to be well suited for dealing with natural anatomical differences. Their drawbacks, however, are their complexity and difficulty of using them when there are large perfusion defects. Nonetheless these models are useful for tracking motion of the LV in gated SPECT imaging. We decided to build a rigid model of the LV by "averaging" a set of images corresponding to cases diagnosed by the physician as normal. We also decided to use, for models only, the images that were evaluated by the most experienced physician. Images were additionally screened for presence of excess noise and artifacts. Before averaging, the selected case images were translated, rotated and scaled to obtain the best match between 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 the averaged model. Due to anatomical differences between patients, models for females and males were created separately. The format of a model is the same as a three-dimensional SPECT image, so a cardiologist can easily evaluate its quality. Each model was manually inspected and the locations of slices and ROIs for that particular model were recorded. An example of three-dimensional rendering of a male rest model is shown in Fig. 4. The rendering was created using the visualization toolkit library [40].
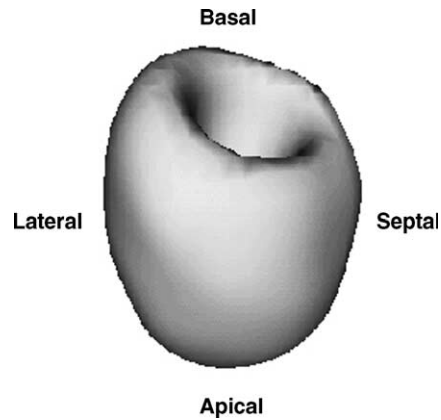
Fig. 4. Three-dimensional rendering of a normal LV model.

4.2. Three-dimensional image registration

Ideally, the object of interest on the analyzed images must have the same spatial orientation and scale to ensure repeatability and robustness of the feature extraction process. Such an operation is called registration or image matching. An example of a two-dimensional image registration is shown in Fig. 5 and a three-dimensional image registration in Fig. 6. Image (a) represents a desired orientation and scale. Image (b) has to be translated, rotated and scaled to match the object model in image (a).

Let I_M denote the three-dimensional reference image—the image containing the model object. Let I_C denote the three-dimensional image that needs to be registered to I_M. Let $I_R = T(I_C)$ denote the image I_C after registration. The problem of image registration is to find a transformation T, which applied to image I_C matches the considered object (in our case the LV) with the model object in image I_M.

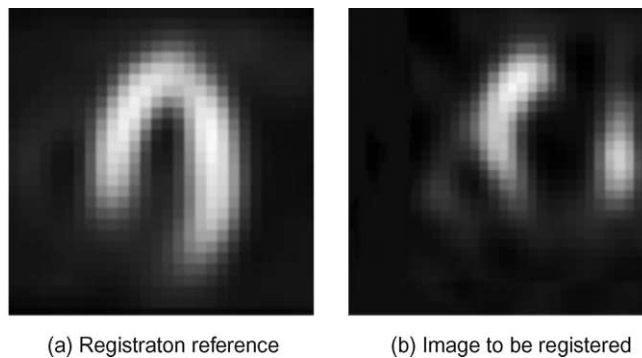(a) Registraton reference   (b) Image to be registered

Fig. 5. Sample of a two-dimensional image registration problem: (a) registration reference; (b) image to be registered. Images represent horizontal long axis view of a LV. Image (a) is from a model of a normal LV; image (b) from a patient with infarct.

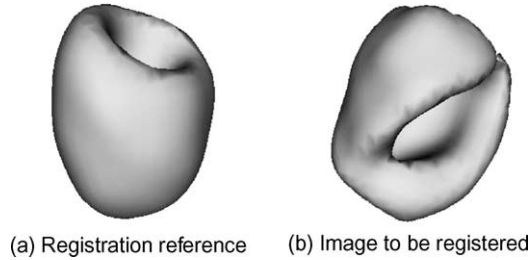(a) Registration reference   (b) Image to be registered

Fig. 6. Sample of a three-dimensional image registration problem: (a) registration reference; (b) image to be registered. Images represent isosurface rendering of three-dimensional cardiac SPECT images. Image (a) is from a model of a normal LV; image (b) from a patient with an infarct.

We searched for the optimal registration transformation T by maximizing the cross-correlation coefficient [37]:

$$r(T) = \frac{\int_W I_M(\boldsymbol{x}) T[I_C(\boldsymbol{x})] \, \mathrm{d}\boldsymbol{x}}{\sqrt{\int_W I_M^2(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \int_W T^2[I_C(\boldsymbol{x})] \, \mathrm{d}\boldsymbol{x}}}$$

where $\boldsymbol{x} = [x_1, x_2, x_3]^T$ is a three-dimensional pixel coordinate, $I(\boldsymbol{x})$ the gray level value of pixel $\boldsymbol{x}$ in image I, W the integration domain (typically it is the whole volume of the reference image I_M). The cross-correlation coefficient is zero for totally dissimilar images and reaches maximum of one for identical images. Alternative approaches to LV registration, including those based on physics-based deformable models, can be found in [11,34].

4.2.1. Three-dimensional image transformation

For registration of three-dimensional SPECT images, we used a transformation T that is a superposition of three component transformations: (1) translation T_t described by three parameters $\boldsymbol{t} = [t_1, t_2, t_3,]^T$ where t_i is a translation along axis X_i; (2) rotation T_α described by three parameters $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^T$ where α_i is an angle of rotation in the plain perpendicular to axis X_i and a center of rotation $\boldsymbol{x}^{(0)}$; (3) isotropic scaling T_s described by a scalar parameter s and a center of rotation $\boldsymbol{x}^{(0)}$ (that is the same as center of rotation); details can be found in [37].

4.2.2. Computation of an image registration transform

Computation of the cross-correlation coefficient for three-dimensional images is computationally intensive. The three-dimensional registration presented above has seven degrees of freedom: translation in three dimensions: t_1, t_2, t_3; three rotation angles: α_1, α_2, α_3; a scaling factor: s. To make the registration of cardiac SPECT images based on cross-correlation coefficient practical, we first estimated the registration transformation and then refined the transformation parameters by performing the best-first heuristic search with decay (to dump oscillations) in the seven-dimensional parameter space.

4.2.3. Estimation of image registration transform

We estimated two components of the registration transform: translation T_t and scaling T_s based on detecting a three-dimensional blob representing the LV walls in the model and

in the target image. A SPECT image was first thresholded at 55% of the maximum intensity of that image to create a binary image. Next, a three-dimensional connected-component labeling algorithm was applied to the binary image to label all blobs in the image. The largest blob near the center of the image was considered to be the LV. The thresholding level had been selected experimentally to ensure that the largest blob near the center actually corresponds to LV.

Let B denote a set of pixels constituting a blob. We will assume that $\boldsymbol{x} \in B$ means that a pixel with coordinates $\boldsymbol{x}$ is a member of the blob B. We can also write $B = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\}$, where n is the number of pixels in the blob.

A center of a blob B, denoted by $\bar{\boldsymbol{x}}$, is defined as an average of blob points' coordinates:

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}^{(i)}$$

This definition gave us better results in estimating translation transformation T_t than other approaches, including one defining the blob's center as a center of gravity.

Let $\bar{\boldsymbol{x}}_M$ be a center of the blob $\boldsymbol{B}_M$ representing LV in the reference image I_M and let $\bar{\boldsymbol{x}}_C$ be a center of the blob $\boldsymbol{B}_C$ representing LV in the target image I_C. We estimated the translation transformation T_t as a distance between centers of blobs $\boldsymbol{B}_M$ and $\boldsymbol{B}_C$:

$$\boldsymbol{t} = \bar{\boldsymbol{x}}_M - \bar{\boldsymbol{x}}_C$$

In order to estimate the scaling transform T_s, we introduced the notion of a scale factor. A scale factor of a blob $\boldsymbol{B}$, denoted by ξ is a median distance of pixels belonging to the blob from the blob's center $\bar{\boldsymbol{x}}$. Let ξ_M be a scale factor of the blob $\boldsymbol{B}_M$ and ξ_C be a scale factor of the blob $\boldsymbol{B}_C$. Then, the scaling transform T_s can be estimated as follows:

$$s = \frac{\xi_M}{\xi_C}$$

4.2.4. Registration refinement

The initial estimate of the registration transformation T was refined by performing an optimization/search that maximizes the correlation coefficient $r(T)$. The search was performed as a series of local neighborhood searches. We started with a search step $\varDelta$:

$$\varDelta = [\varDelta_{t_1}, \varDelta_{t_2}, \varDelta_{t_3}, \varDelta_{\alpha_1}, \varDelta_{\alpha_2}, \varDelta_{\alpha_3}, \varDelta_s]^{\mathrm{T}}$$

A search neighborhood is defined by a search center $T^{(0)}$ and the maximum number of steps γ that can be taken from the center. The search neighborhood consists of $(2\gamma + 1)^7$ points. For $\gamma = 1$, it is $3^7 = 2187$ points, for $\gamma = 2$ it consists of $5^7 = 78{,}125$ points and so on. Typically $\gamma = 1$ is a good enough trade-off between accuracy and computational complexity. The transformation T' in the search neighborhood that has the highest correlation coefficient $r(T')$ becomes the center of the next search neighborhood. The local search is repeated as long as the change in the correlation coefficient $\left|r(T') - r(T^{(0)})\right|$ is greater than some limit value or until the maximum number of iterations has been reached. To eliminate a tendency to get into a sustained oscillation cycle two or more iterations long, we introduced a decay factor that decreases the search step in each iteration.

4.2.5. Image registration database

The calculation of the registration transformations for each of the images for our 349 cases (with corresponding interpretation) in the SPECT database was time consuming. However, this calculation needed to be performed only once. We have added to the SPECT database a table containing the registration transformation parameters and the corresponding value of the cross-correlation coefficient for each image and a model of the normal LV.

4.3. Extraction of features from the registered images

Feature extraction was done in two phases. First, objects of interest in the image were detected. Next, for each of the objects or their parts, some features of interest were calculated. For SPECT images, the objects of interest are the walls of the LV. Following the approach described Section 2.2, we were interested in parts of the LV walls corresponding to the 22 ROIs shown in Fig. 2. We interpreted these 22 regions as three-dimensional parts of the LV walls. Once we were able to locate a region within a three-dimensional SPECT image, a set of features was calculated for each of them separately. Then a feature could be defined, for instance, as median pixel intensity within a region.

4.3.1. Detection of objects in the SPECT images

We considered several approaches for detection of the LV walls. One of them was described in Section 4.2.3—an image is thresholded and objects are detected by connected-components labeling. This approach is computationally efficient but not robust enough for detection of the LV walls. An approach based on spherical radial gradient search has been proposed by Declerck et al. [11]. The difficulty with this approach is that the SPECT images typically are of low contrast and transitions between pixels in infarcted or ischemic LV wall are small, making gradient detection methods impractical. This is especially true for Tl^{201} SPECT images used here. A method specifically designed for detection of free-shape objects on low contrast images has been developed by Sacha et al. [39]. It performs object detection using a region growing approach. We initially used it for detection of the LV walls but encountered problems related to low resolution of SPECT images (small pixel size of objects). The method that proved most robust is based on the spherical radial search of Goris et al. [21]. We eventually used a variant of this method that performs a search in cylindrical coordinates since it is closer to the cardiologist's interpretation process.

4.3.1.1. Spherical radial search. Goris et al. [21] modified the bull's eye image creation method by using only spherical search. A typical bull's eye method uses a combination of cylindrical and spherical radial search; spherical search is performed in the apex area, while the cylindrical search is used in the mid and basal section of the LV. The method uses only spherical radial search with a center positioned near the base of LV [21]. In what follows, we will refer to this method as GBB. The initial step in the GBB method is background subtraction done at a fixed level of 33% of the maximum pixel value in the three-dimensional image.

In the GBB method, the LV image is manually reoriented to normalize its position. The operator selects a center of radial search manually before the search is performed. In our
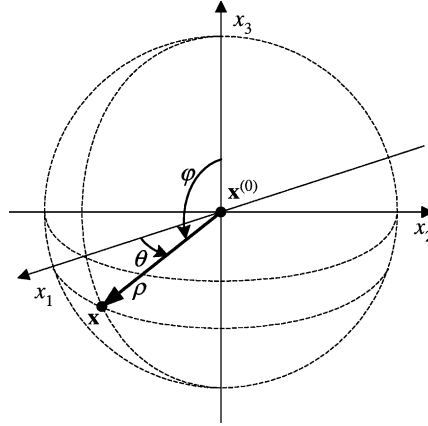
Fig. 7. Spherical coordinates system.

approach, a model of a normal LV is utilized for reorientation (registration) and automatic selection of the radial search center.

The spherical system of coordinates is shown in Fig. 7. $\boldsymbol{x}^{(0)}$ is the center of the spherical coordinates (the same as the center of the radial search). Coordinates of a point $\boldsymbol{x}$ are represented by a triple (θ, φ, ρ), where θ is an angle between vector $\vec{\boldsymbol{r}} = \boldsymbol{x} - \boldsymbol{x}^{(0)}$ and axis X_1, φ an angle between vector $\vec{\boldsymbol{r}}$ and axis X_3, and ρ the length of vector $\vec{\boldsymbol{r}}$. The search is performed along the vector $\vec{\boldsymbol{r}}$ for ρ ranging from 0 to some maximum value. Angle θ is changed in the full range form 0 to 360° and angle φ from −135 to +135°.

Two mappings are created during the search: $I_{\text{MAX}}(\varphi, \theta)$ and $I_{\text{TOT}}(\varphi, \theta)$. The MAX mapping contains the maximum value found along vector $\vec{\boldsymbol{r}}$ for fixed φ and θ. The TOT mapping contains the integral of values found along vector $\vec{\boldsymbol{r}}$ for fixed φ and θ.

To make the visualization of the radial search results more intuitive, mappings $I_{\text{MAX}}(\varphi, \theta)$ and $I_{\text{TOT}}(\varphi, \theta)$ are transformed to (x_1, x_2) coordinates forming two round images:

$$x_1 = \varphi \sin \theta + x_1^{(0)}$$
$$x_2 = \varphi \cos \theta + x_2^{(0)}$$

In the $I_{\text{MAX}}(x_1, x_2)$ and $I_{\text{TOT}}(x_1, x_2)$, the center of an image represents the apex of the LV, edges of an image are near the base of the LV. An example of $I_{\text{MAX}}(x_1, x_2)$ and $I_{\text{TOT}}(x_1, x_2)$ for rest and stress SPECT images created during our experiments is presented in Fig. 8.

4.3.1.2. Cylindrical radial search. Cylindrical radial search, or cylindrical unwrapping, is similar to the spherical radial search except it is performed in the cylindrical coordinate system (see Fig. 9). The $I_{\text{MAX}}(z, \theta)$ and $I_{\text{TOT}}(\varphi, \theta)$ mappings are calculated in a fashion similar to the spherical search. We did not convert these mappings to (x_1, x_2) coordinates, as is done in bull's eye methods [43]. We used them directly for feature extraction since conversion would have introduced additional interpolation errors.
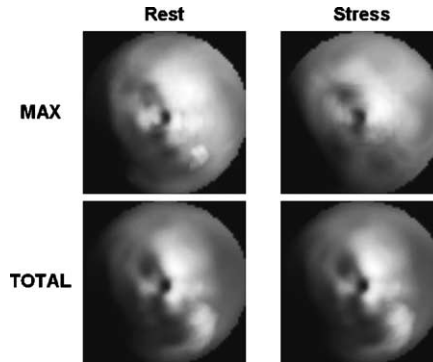
Fig. 8. Bull's eye images created by spherical unwrapping of the LV.

4.3.2. Radial search and location of the 22 three-dimensional ROIs

Radial search transforms three-dimensional SPECT images to create two-dimensional maps. The objective is to remove irrelevant information from the three-dimensional images and present the relevant information in a simpler two-dimensional form. Perfusion codes recorded in our database come from physicians inspecting slices of three-dimensional SPECT images in short and long axes views. To achieve a closer match between the physician interpretation and the automated interpretation process, we found it more appropriate to extract features in cylindrical coordinates rather then in the spherical coordinates.

4.3.2.1. Short axis views. The cylindrical search is well suited for detection of the LV walls in the short axis views (Fig. 2). The walls in these views are roughly cylindrical in shape. We can interpret each of the short axis ROIs, in three-dimensional images, as a wedge. These wedges can be one or more slices in thickness. Fig. 10a shows our selection of angles that define wedges in three-dimensional images. The map created by the spherical radial
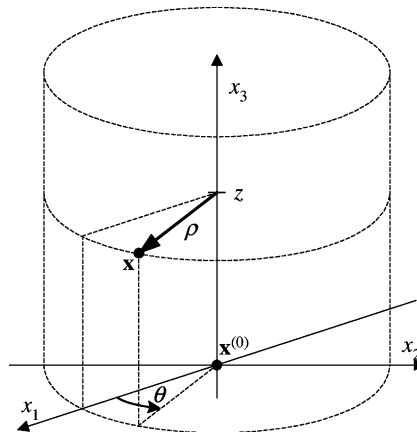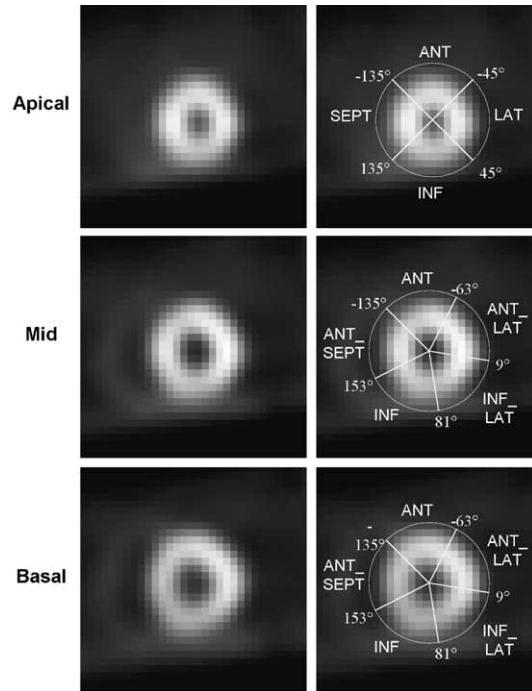

Fig. 9. Cylindrical coordinates system.

(a) Definations of ROIs superimposed on male models of normal left ventricle.

(b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

Fig. 10. ROIs for cylindrical search—short axis views: (a) definitions of ROIs superimposed on male models of normal LV; (b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

search is a rectangular image, I_{MAX} or I_{TOT}. Each of the wedges corresponds to a rectangular area in that image, as shown in Fig. 10b. A single image contains data for the apical, mid and basal views. The center of search and range for z and ρ depends on the model of the normal LV used. The center of search is located at the center of the LV cavity. The range of z typically spans nine slices, three slices for each of the views.
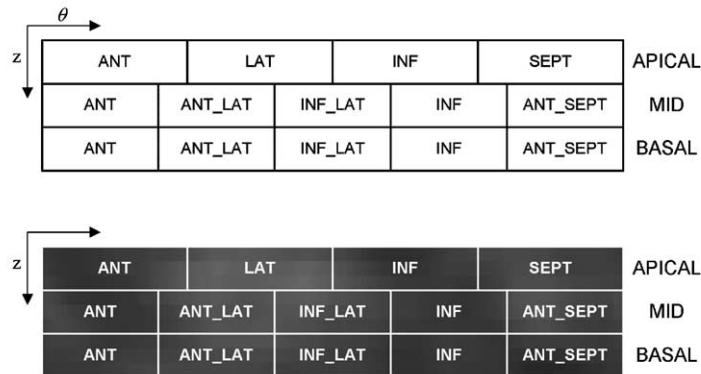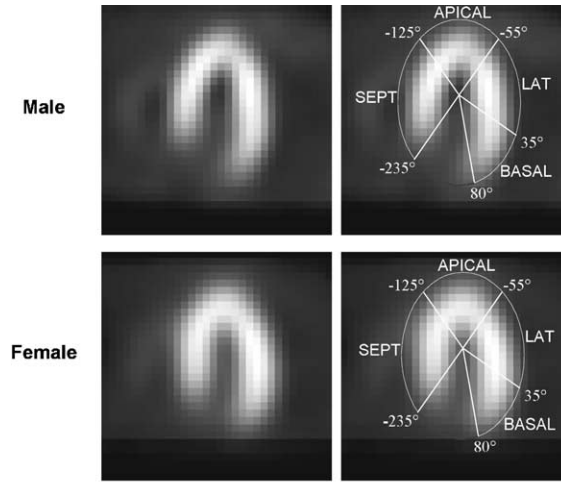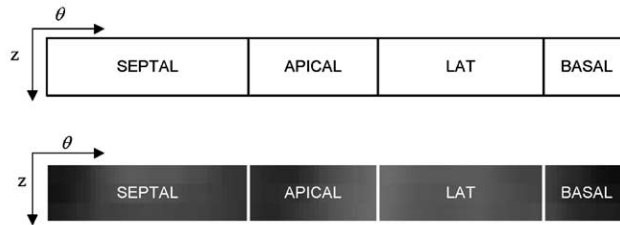
(a) Definations of ROIs superimposed on male models of normal left ventricle.

(b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

Fig. 11. ROIs for cylindrical radial search—horizontal long axis view: (a) definitions of ROIs superimposed on male models of normal LV; (b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

4.3.2.2. Long axis views. We also used a radial cylindrical search for the detection of LV walls in the long axis views. The LV walls are roughly cylindrical in shape in these views, as shown in Figs. 11 and 12. Cylindrical search maps were created separately for the horizontal long and short axes views.

4.3.3. Feature extraction for radial search maps

We used the radial search maps I_{MAX} and I_{TOT} created for three-dimensional rest and stress study for feature extraction. Each of the maps was partitioned into regions corresponding to 22 ROIs, as described in Section 4.3.2. For each partition, we calculated the maximum, mean, median and standard deviation of pixel values in that partition. This way we had 16 features extracted in each ROI. Features were stored in the SPECT database. Feature records also contained information about normal LV models used for image registration, registration transformation and parameters used for creation of I_{MAX} and I_{TOT} images. We used this information to prepare data for perfusion classification experiments.

(a) Definations of ROIs superimposed on male models of normal left ventricle.

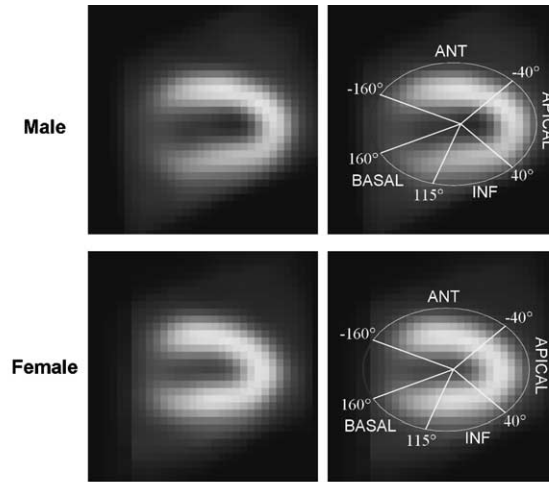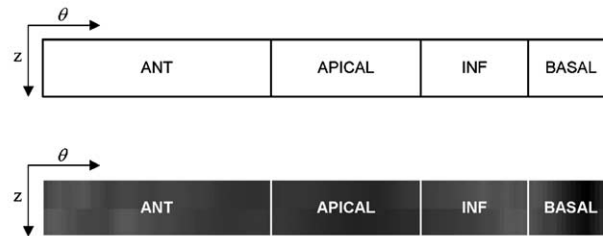(b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

Fig. 12. ROIs for cylindrical radial search—vertical long axis view: (a) definitions of ROIs superimposed on male models of normal LV; (b) ROIs in the cylindrically unwrapped image are the same for I_{MAX} and I_{TOT}.

5. Data mining

After organizing the database and extracting the relevant features, we were then ready to perform the step of actual data mining. Of the multitude of possible data mining tools that could have been used, we decided to concentrate in this work on using Bayesian classifiers.

5.1. Reasoning under uncertainty

Fig. 3 shows a simplified perfusion classification process that is deterministic and entirely based on classification of perfusion in each of the ROIs. In practice, however, this classification is more complex, non-deterministic and involves a number of other factors that may not be recorded in the database. A cardiologist's interpretation process is, to a large extent, qualitative. The data in the database are by their nature quantitative; some numerical and some categorical. The need to encode the qualitative information into a numerical or categorical form leads to a level of uncertainty associated with each recorded value. Intentionally there were only a limited number of codes that could be used to record

perfusion. The intention had been to standardize the recording process by limiting the number of defect codes, with each code having a clear description. Several cardiologists contributed to our database. Each may have had a different bias how to code the cases that do not belong to just a single category. There is a significant variability in diagnosing of the same images between cardiologists [9] and there is always the possibility of data errors in the database, e.g. typos.

Until recently, using a strictly mathematical approach to reasoning under uncertainty was considered impractical. This was due to the problem of computing the joint probability distribution of a large number of random variables involved in reasoning. The last decade, however, has seen increased interest in graphical models representing dependency relationships within a set of random variables. Nodes in a graph represent random variables. An arc in the graph intuitively corresponds to a dependency relationship between two variables. The lack of an arc can be intuitively interpreted as a lack of dependency between two variables (Fig. 13). This intuitive graphical interpretation of dependencies between variables is one of the reasons for the popularity of graphical models. Other reasons are significant progress in theory and algorithms for inference, as well as advances in learning structure and parameters of graphical models [22,26].

5.1.1. Bayesian network classifiers

One of the most popular types of graphical models are Bayesian networks. What differentiates them from other graphical models is that arcs in a Bayesian network are directed to represent conditional dependence among variables (Fig. 14).

A Bayesian network for a set of random variables $X = \{X_1, \ldots, X_n\}$ is a pair $B = \langle S, P \rangle$, where S is a directed acyclic graph (DAG) whose nodes are in one-to-one correspondence with random variables in X. P is a set of local probability distributions associated with each variable. X_i denotes both the variable and its corresponding node in S. We use Pa_i to denote parents and pa_i to denote configuration of parents of node X_i in S as well as variables corresponding to these parents. The joint probability represented by the structure S is given by:
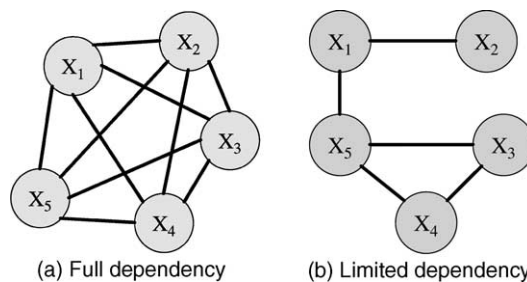
$$p(x) = \prod_{i=1}^{n} p(x_i | \mathrm{pa}_i)$$

Fig. 13. Examples of graphs representing graphical models: (a) full dependency; (b) limited dependency.
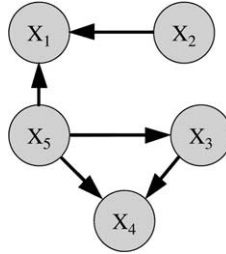
Fig. 14. Bayesian network.

The local probability distributions P are the distributions corresponding to the terms in the above equation. For example, the probability distribution represented by the network shown in Fig. 14 is:

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1|x_2, x_5) \cdot p(x_2) \cdot p(x_3|x_5) \cdot p(x_4|x_3, x_5) \cdot p(x_5)$$

A Bayesian network can be used for classification in a quite straightforward way. One of the variables is selected as a class variable (C) and the remaining variables as attribute variables ($A_1, \ldots, A_{n-1}$) (see Fig. 15). A classification is performed by selecting value of the class variable $c^{(k)}$ that has the highest conditional probability given values of attributes, $p(c^{(k)}|a_1, \ldots, a_{n-1})$. The problem of learning BNCs is different from the general problem of learning Bayesian networks in that the former approximate marginal distribution of a class variable while the latter approximate joint distribution of all variables.

The simplest BNC is the naïve Bayes [12]. Its efficiency and low classification error rate made it one of the most commonly used classifiers. The naïve Bayes has a fixed network structure; only parameters are adjustable. The class node is a parent to all attribute nodes and there are no edges between the attribute nodes (Fig. 16). In other words, the naïve Bayes classifier assumes that all attribute variables are conditionally independent given the class variable. Despite the fact that these assumptions are in most cases violated, the naïve Bayes classifier performs as well as many of the state of the art classifiers. Its properties have been intensively studied [25,27,28,30–32]. It is believed that the secret of naïve Bayes performance is its small number of parameters, as compared to other classifiers; they can be estimated with low variance (high accuracy) even from a limited number of training
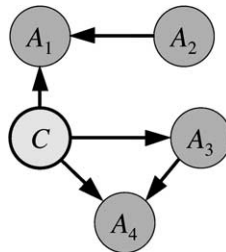

Fig. 15. Bayesian network classifier.
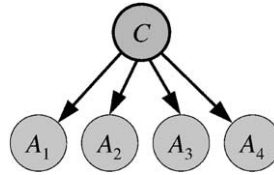
Fig. 16. Naïve Bayes classifier.
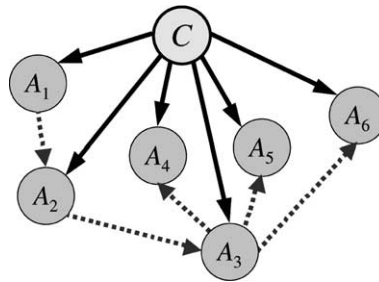
Fig. 17. Bayesian network representing tree-augmented naïve Bayesian classifier.

samples [18]. In other words, a naïve Bayes network with few parameters compensates for computational precision of a more parameter-rich system.

The main drawback of the naïve Bayes is the assumption of conditional independence of attributes (represented by fixed network structure). Friedman et al. [19] proposed a method that introduced dependencies among attributes using the dependency discovery approach of Chow and Liu [4]. This method assumes that dependencies among attribute nodes can be represented by a tree structure (Fig. 17). They called it a tree-augmented naïve Bayes (TAN) and demonstrated that it performs as well or better than naïve Bayes and also the C4.5 classifier [35]. The important property of this approach is that the augmenting tree-like network can be efficiently calculated in a closed form.

In Bayesian network-augmented naïve Bayes (BAN), the naïve Bayes is augmented with an unrestricted Bayesian network of dependencies among the attributes. Unlike TAN, the BAN cannot be constructed in a closed form since the problem of constructing unrestricted Bayesian networks is NP-hard.[1] Variants of BAN approach have been proposed [3,19,41]. The BAN of [19] uses search and scoring for creation of the augmenting network based on minimal description length principle [29,42]. Another BAN approach [41] creates augmenting network using Bayesian network learning algorithm of Cooper and Herskovits [6].

---

[1] NP-hard means non-polynomial-hard. Generally, algorithms are considered tractable when the time needed to execute the algorithm is a polynomial of the number of parameters, e.g. number of nodes in the network [13]. In particular, this function can be linear in the number of parameters. When there are no algorithms to solve a given problem so that the time needed to execute that algorithm grows no faster then according to some polynomial function of parameters, the problem is called NP-hard. For instance, a problem for which the most efficient algorithm has the execution time exponential with the number of parameters is NP-hard.

That method searches through domain of all possible network structures and scores them using a Bayesian scoring metric. The BAN of [3] uses dependency discovery approach to Bayesian network construction [2]. However, the BANs computational complexity is typically high, thus making them unsuitable for large applications.

5.1.2. New family of Bayesian network classifiers

The best BNCs are the naïve Bayes and TAN classifiers [19]. We used them as a starting point to come up with a new family of BNCs [37]. The approach combines the best of the search and scoring approach with dependency discovery approaches to achieve low computational complexity and high classification accuracy. The classification networks are constructed under two assumptions: (1) an attribute node can have at most two parent nodes; (2) class node has no parents. The first assumption is to control numerical complexity of search algorithms, while the second is necessary to eliminate parameter redundancy in a classifier network to lower the variance of parameter estimation. A software package implementing the new family of BNCs and some of the other approaches is available at: http://jbnc.sourceforge.net/.

One of the main factors contributing to a steep increase in the number of parameters of a Bayesian network is the number of parents for each variable in the network. The number of parameters associated with each node in the network is exponential with the number of this node's parents. Thus, the first principle of our synthesis was to limit the number of parents for each node. We assumed that each node has no more than two parents.

Our second assumption, constraining the structure of the network, was that any edge between a class node and an attribute node is always directed away from the class variable; or in the other words, the class variable has no parents. This assumption comes from the way information is passed between nodes in a Bayesian network during inference. When the value at a node is known, the node is instantiated and the node blocks any information passing from its parents to its children. On the other hand, the information can flow from children to parents. This property of Bayesian networks is called dependency separation [24,33]. Typically, when inference about the class node is performed, all attribute nodes are instantiated. If an attribute node is a parent of the class node, then since it is instantiated, it will block the information path between its parents and the class variable; it will make the class node conditionally independent from its parents. We were only interested in the inference in the class node and did not want to have superfluous edges and parameters in the network that would not contribute to the classification task, thus we assumed that the class node has no parents.

When Friedman et al. [19] introduced a tree-augmented naïve Bayes classifier, they first assumed that every attribute depends on the class variable, as in the naïve Bayes. Then they extended the naïve Bayes network structure by adding tree-like dependency among attribute variables. In their approach, there is always an edge from the class node to every attribute node and there is always an undirected path between any attributes that do not pass through the class variable. This may force dependencies between random variables in the model that do not really exist and thus deteriorate classification performance.

In this work, we relaxed both of the constraints [19] as well as restrictions of the naïve Bayesian classifier. In our approach not all attributes need to be dependent on the class

variable and there may be no undirected path between two attribute nodes. We introduce a family of five network search algorithms.

- *Selective tree-augmented naïve Bayes* (*STAN*): A search algorithm decides which attribute nodes depend on the class node. Augmenting tree spans all of the attribute nodes.
- *Selective tree-augmented naïve Bayes with discarding* (*STAND*): A search algorithm decides which attribute nodes depend on the class node, other attributes are discarded and nodes corresponding to them removed from the network. Augmenting tree spans only attributes that depend on the class variable.
- *Forest-augmented naïve Bayes* (*FAN*): All attribute nodes depend on the class node. Dependency among attribute nodes has a form of a forest (a number of trees). The augmenting forest is created by a search algorithm based on the Kruscal's maximum spanning tree algorithm [8]. In particular the augmenting forest may be empty and the classifier has a naïve Bayes structure or it can have only one tree spanning all of the nodes and the classifier has a TAN structure.
- *Selective forest-augmented naïve Bayes* (*SFAN*): A search algorithm decides which attribute nodes depend on the class node. Augmenting forest may span all of the attribute nodes.
- *Selective forest-augmented naïve Bayes with discarding* (*SFAND*): A search algorithm decides which attribute nodes depend on the class node, other attributes are discarded and nodes corresponding to them removed from the network. Augmenting forest may span only attributes that depend on the class variable.

Each of the algorithms differs in terms of the trade-off it makes between computational complexity and richness of the network structures it creates. Fig. 18 illustrates richness of
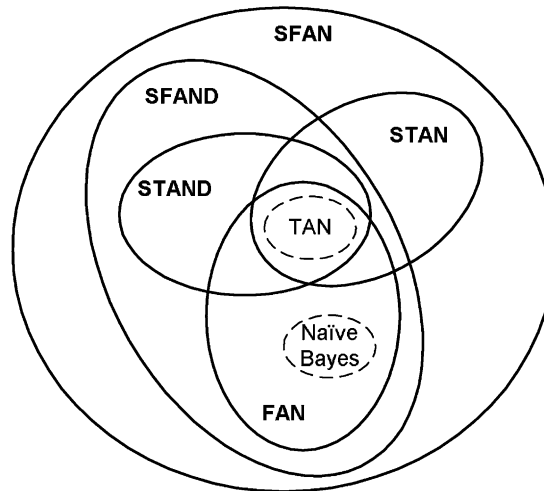

Fig. 18. New family of Bayesian network classifiers with their relation to naïve Bayes and TAN classifiers. Solid lines represent new algorithms and dashed lines the old algorithms.

structures that can be produced by each of our algorithms and their relationship to the naïve Bayes and TAN classifiers. The latter two are special cases of the proposed algorithms.

Various network structure quality measures can modify searches performed by each of the new algorithms. We used five quality measures: Bayesian dirichlet metric [23], standard Bayesian measure [1], leave-one-out cross-validation, 10-fold cross-validation and 10 times 10-fold cross-validation. A combination of the five search algorithms with five quality measures gave us 25 new BNC learning algorithms [37].

5.2. Perfusion classification error reference level

If classification of myocardial perfusion was completely deterministic, given the classification within ROIs, then the reference classification error would be zero. Since it is not, we wanted to estimate the perfusion classification error rate present in the gold standard. This helped us to set more realistic performance requirements for classifiers that used only features extracted from the SPECT images.

We used the perfusion classification recorded in the database by cardiologists as a gold standard while constructing a classifier based on features extracted from the SPECT images. The main objective of the experiment described in this section was to estimate goodness of the gold standard by estimating the perfusion classification error rate. In this experiment, the class variable represented the overall perfusion code as recorded in the database. The attributes represented a partial perfusion code recorded for each of the 22 ROIs and other information recorded in a study.

The second goal of this experiment was to see if some additional clinical information, for instance, patient's gender, weight, etc. might be useful in perfusion classification. We wanted to determine whether use of that additional information could improve the classification error rate.

The upper boundary of the classification noise that is present in the database can be estimated by constructing a number of classifiers and measuring their error rate using cross-validation. We used the constant classifier[2], C4.5 classifier [35], naïve Bayes classifier [12], tree-augmented naïve Bayes classifier [19] and 11 of the new BNCs [37]. We used only records that had no missing, incorrect or multiple values for considered attributes. Three sets of attributes were used to estimate classification error level.

- Diagnosis by a cardiologist for each of the 22 ROIs (Fig. 2). We refer to this 22-attribute set as Code22. Each of the attributes can take on seven values which describe defect type: NL, normal; F, fixed; R, reversible; P, partially reversible; E, equivocal; X, reverse redistribution; ART, artifact. This set has 1762 cases.
- A new dataset was obtained from Code22 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 attributes taking on integer values from 0 to 22. Similarly to the Code22 set, this set has 1762 cases. We refer to this dataset as Count.

---

[2] The constant classifier predicts a constant class—the majority class in the training set. This is the simplest classifier that can be built. It is used to establish a base performance level. If no classifier can reach performance better then the constant classifier it suggests that a dataset does not contain any information useful for classification; a dataset contains only noise.

Table 2
Summary of the estimation of reference classification error level using 10-fold cross-validation

Dataset	No. of variables	No. of cases	Constant classifier	C4.5	Naïve Bayes	TAN	Best new BNC
Code22	22	1762	73.38 ± 1.25	21.28 ± 1.04	20.94 ± 1.28	20.37 ± 1.48	**19.30 ± 1.15**
Count	7	1762	73.38 ± 1.25	**16.63 ± 0.72**	18.05 ± 0.98	17.82 ± 0.99	17.59 ± 0.99
Extras	16	1619	73.44 ± 0.74	66.89 ± 0.96	60.72 ± 0.83	60.22 ± 1.29	**59.85 ± 1.01**
Code22 & Extras	35	1619	73.44 ± 0.74	22.54 ± 1.58	21.74 ± 1.09	20.45 ± 0.92	**19.64 ± 0.77**
Count & Extras	20	1619	73.44 ± 0.74	18.78 ± 0.85	17.66 ± 1.04	18.47 ± 0.92	**17.60 ± 0.84**
Code22 & Count	29	1762	73.38 ± 1.25	**17.54 ± 0.90**	19.69 ± 1.23	18.22 ± 1.12	17.70 ± 1.07
Code22 & Count & Extras	42	1619	73.44 ± 0.74	18.53 ± 0.85	20.26 ± 0.98	17.91 ± 0.91	**17.66 ± 0.87**

The first number represents an error in percent; the second is the standard deviation of the sample mean.

- Still another set was created using criteria such as relevance, completeness and correctness of attribute values. This resulted in 16 attributes with a reasonably high number of corresponding entries in the Code22 data. The following variables were selected: sex, age, height, weight, body surface area (BSA) coefficient, diabetes mellitus, family history, HTN, smoker/non-smoker, chest pain, high cholesterol, prior MI and LV size. This set has 1619 cases. We refer to it as Extras.

The first two sets contain only information from the interpretation of the SPECT images. The last one contains other attributes from patient records that might influence the diagnosis of myocardial perfusion.

We used these sets of attributes and their combinations to construct seven datasets: Code22, Count, Extra, Code22 & Extra, Count & Extra, Code22 & Count and Code22 & Count & Extra. Ten-fold cross-validation was used. The summary of the classification results is shown in Table 2. Results for selected new BNCs are shown in Tables 3 and 4. The first number is the mean cross-validation classification error in percent; the second is the standard deviation of the mean.[3] The following conclusions can be drawn from Table 2.

- The data seem to have high classification noise; the lowest classification error is above 16%.
- Regional classification codes (the Count dataset) seem to be "easier" for learning.
- Additional attributes (Extras) contain some useful information for the classification. The error rate for the Extras dataset is high but it is over 13% lower than the constant classifier, suggesting that some of the attributes in the Extras dataset are correlated to the classification category.
- Classification using the extended datasets, i.e. regional information with added attributes suffers from increased classification error.

---

[3] Experimental standard deviation of the mean of a sample $x_1, \ldots, x_\phi$:

$$\sigma(\bar{x}) = \sqrt{\frac{1}{\phi(1-\phi)} \sum_{i=1}^{\phi} (x_i - \bar{x})^2}$$

where $\bar{x} = (1/\phi) \sum_{i=1}^{\phi} x_i$ is the experimental mean of the sample. The standard deviation of the mean is not the standard deviation of the sample. It shows how variable the mean is, which is smaller than the variability of the sample itself by a factor of sample's size [36].

Table 3
Estimation of reference classification error level using 10-fold cross-validation and FAN classifier

Dataset	FAN				
	HGC	SB	LC	LOO	CV1010
Code22	19.81 ± 1.54	19.64 ± 1.24	19.30 ± 1.15	19.58 ± 1.23	20.34 ± 1.13
Count	**17.59 ± 0.99**	17.93 ± 1.04	17.76 ± 1.04	17.65 ± 1.01	17.70 ± 1.01
Extras	60.53 ± 1.27	**60.22 ± 0.86**	60.47 ± 0.95	60.41 ± 0.94	60.41 ± 0.94
Code22 & Extras	20.57 ± 0.86	**20.14 ± 0.96**	20.38 ± 1.01	20.94 ± 0.81	21.19 ± 0.88
Count & Extras	18.10 ± 1.06	**17.60 ± 0.84**	17.60 ± 0.98	17.79 ± 0.97	17.66 ± 0.97
Code22 & Count	18.33 ± 1.27	17.93 ± 1.23	17.99 ± 1.31	**17.76 ± 1.27**	17.88 ± 1.30
Code22 & Count & Extras	19.02 ± 1.02	18.40 ± 0.80	18.28 ± 0.99	**17.85 ± 0.85**	18.03 ± 0.84

In our previous work [38], we studied the relation between the number of attributes selected for classification and the classification error. In particular, we investigated the Count & Extras dataset created from earlier version of database. That dataset contained 1433 cases. The attributes were first ordered according to their contribution to the classification goal based on analysis of classification rules generated by C4.5 classifier. As expected, attributes from the Count set ranked highest. We started with a dataset that contained only one attribute, with the highest rank and performed a 10-fold cross-validation test using the C4.5 classifier. Then, we added the second attribute to the dataset and repeated the cross-validation test and so on. The results are shown in Fig. 19. These results agree with the ones presented in Table 2: addition of extra attributes from the database that are not directly related to myocardial perfusion does not improve classification results and may actually increase the error. This is in spite of the fact that, as shown in Table 2, the Extras set on its own contains information useful for classification. This is not a contradiction. The larger number of attributes in the dataset means that a classifier needs to be described by a larger number of parameters. When we keep the number of cases in a training dataset constant and increase the number of parameters in a classifier, the variance of estimation of the classifier parameters using that dataset increases. Each of the parameters is estimated with a lower accuracy that may lead to a lower classifier performance. There is usually some optimal trade-off between the number of attributes

Table 4
Estimation of reference classification error level using 10-fold cross-validation, STAN and STAND classifiers

Dataset	STAN			STAND		
	HGC	SB	LC	HGC	SB	LC
Code22	22.02 ± 1.15	35.75 ± 1.64	22.64 ± 1.12	**19.97 ± 1.53**	43.02 ± 1.14	20.43 ± 1.18
Count	**17.82 ± 0.99**	**17.82 ± 0.99**	**17.82 ± 0.99**	17.76 ± 1.00	17.76 ± 1.00	17.76 ± 0.94
Extras	**59.85 ± 1.01**	60.96 ± 1.21	59.35 ± 0.90	61.70 ± 0.96	62.87 ± 1.06	59.97 ± 1.13
Code22 & Extras	22.05 ± 1.19	26.13 ± 1.42	23.59 ± 0.98	**19.64 ± 0.77**	24.77 ± 1.01	20.81 ± 0.95
Count & Extras	18.47 ± 0.90	19.09 ± 1.26	18.03 ± 0.95	17.91 ± 1.05	**17.60 ± 1.10**	18.16 ± 1.13
Code22 & Count	21.51 ± 1.02	25.42 ± 0.98	19.69 ± 0.99	18.05 ± 1.08	19.01 ± 0.90	**17.70 ± 1.07**
Code22 & Count & Extras	21.62 ± 1.33	26.19 ± 1.40	19.64 ± 0.91	**17.66 ± 0.87**	19.15 ± 1.09	18.44 ± 0.92
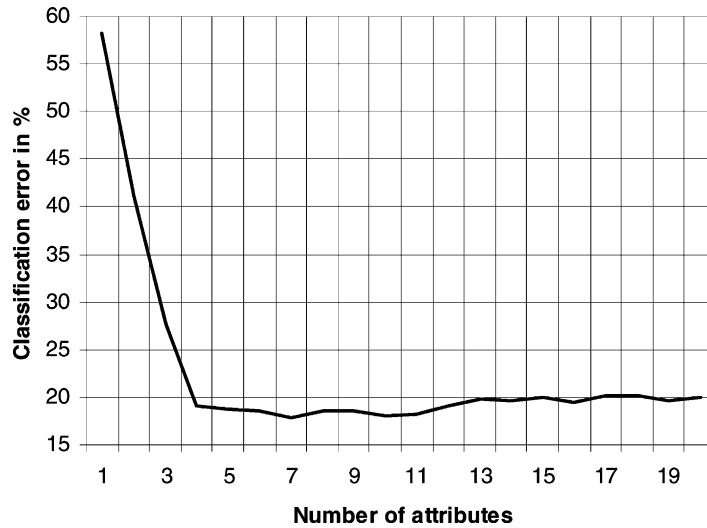
Fig. 19. Classification error as a function of number of attributes used for classification for dataset Count & Extras.

in the dataset and the classification error. This is illustrated in Fig. 19. Around seven attributes, for this particular dataset, seem to give the most optimal results. Based on these results, in order to minimize the parameter variance, we decided not to include attributes from the Extras.

5.3. Partial classification of left ventricular perfusion

This section presents the classification performed separately in each of the 22 three-dimensional ROIs (Fig. 2). We tested all new BNCs [37]. For reference, we also include classification results using the constant classifier, C4.5 classifier and a continuous version of the naïve Bayes classifier. We used 44 datasets having 16 attributes each (one dataset for each of the 22 three-dimensional ROIs, for females and males separately). The results are summarized in Tables 5–7. We show cross-validation result for each the datasets.

5.3.1. Feature extraction and creation of datasets

The features were extracted from three-dimensional SPECT images using a technique described in Section 4. We have used more than a single two-dimensional slice for each of the five views (see Fig. 2). This was to compensate for ambiguity of finding exactly the same slice that may have been used by an evaluating physician while recording perfusion codes. For each of the short axis views, we used three slices; the search radius was set to 9 pixels for females and 10 pixels for males. For the horizontal long axis view, we used three slices; the search radius was set to 11 pixels for females and 12 pixels for males. For the vertical long axis view, we used two slices; the search radius was set to 10 pixels for females and 12 pixels for males. The number of slices per view for females and for males was coincidentally the same, although it was not intentional. The number of slices and

Table 5
Case counts for partial LV perfusion classification tests

ROI name	Female						Total	Male						Total
	Defect codes							Defect codes						
	NL	F	P	R	X	A		NL	F	P	R	X	A	
SHORT_AXIS_AP_ANT	67	10	10	8	1	28	124	137	10	8	11	1	3	170
SHORT_AXIS_AP_LAT	110	4	6	2	0	2	124	147	8	7	7	1	0	170
SHORT_AXIS_AP_INF	106	12	2	2	0	2	124	64	53	22	16	4	9	168
SHORT_AXIS_AP_SEPT	114	2	1	5	0	2	124	154	6	4	5	0	1	170
SHORT_AXIS_MID_ANT	57	13	6	15	2	31	124	141	5	9	13	0	2	170
SHORT_AXIS_MID_ANT_LAT	99	7	3	8	0	7	124	157	2	5	5	1	0	170
SHORT_AXIS_MID_INF_LAT	106	9	2	6	0	1	124	104	23	20	12	2	8	169
SHORT_AXIS_MID_INF	109	9	2	3	0	1	124	67	53	26	10	2	11	169
SHORT_AXIS_MID_ANT_SEPT	106	4	2	10	0	2	124	145	6	8	9	1	0	169
SHORT_AXIS_BASAL_ANT	60	14	9	11	1	29	124	148	5	6	8	1	2	170
SHORT_AXIS_BASAL_ANT_LAT	107	6	2	4	0	4	123	164	1	1	2	1	1	170
SHORT_AXIS_BASAL_INF_LAT	104	10	3	7	0	0	124	101	28	21	9	0	10	169
SHORT_AXIS_BASAL_INF	103	11	3	3	1	2	123	62	50	27	13	2	15	169
SHORT_AXIS_BASAL_ANT_SEPT	102	3	4	9	1	5	124	146	3	9	11	0	1	170
HORIZ_LONG_SEPT	114	0	0	7	1	2	124	139	5	9	15	1	0	169
HORIZ_LONG_APICAL	100	8	5	7	1	2	123	111	31	7	19	1	1	170
HORIZ_LONG_LAT	103	6	4	7	1	3	124	143	11	10	4	2	0	170
HORIZ_LONG_BASAL	113	5	2	4	0	0	124	151	11	10	4	2	0	178
VERT_LONG_ANT	61	12	6	14	1	28	122	137	6	11	13	1	2	170
VERT_LONG_APICAL	103	8	5	5	1	2	124	95	32	11	26	2	3	169
VERT_LONG_INF	105	11	2	5	0	1	124	54	53	29	19	3	10	168
VERT_LONG_BASAL	109	10	1	3	1	0	124	141	17	4	3	0	5	170
Total	2158 (79)	174 (6)	80 (3)	145 (5)	12 (0)	154 (6)		2708 (72)	419 (11)	264 (7)	234 (6)	28 (1)	84 (2)	

The values given in parenthesis are in percentage.

Table 6
Ten-fold cross-validation error of the partial LV perfusion classification for reference classifiers and summary for new Bayesian network classifiers

ROI name	Constant classifier	C4.5	Naïve Bayes	TAN	Best new BNC
Female					
SHORT_AP_ANT	46.09 ± 3.86	37.69 ± 3.06	33.65 ± 4.05	31.99 ± 4.03	**27.18 ± 3.32**
SHORT_AP_LAT	11.45 ± 2.84	9.68 ± 2.57	14.36 ± 3.40	11.99 ± 3.56	**8.08 ± 2.39**
SHORT_AP_INF	14.42 ± 3.05	20.71 ± 3.43	14.29 ± 3.18	15.06 ± 3.91	**13.53 ± 3.31**
SHORT_AP_SEPT	8.01 ± 2.35	**7.18 ± 2.20**	9.68 ± 2.30	8.01 ± 2.35	8.01 ± 2.35
SHORT_MID_ANT	46.03 ± 4.03	36.35 ± 3.96	31.35 ± 2.88	32.24 ± 3.55	**27.44 ± 2.77**
SHORT_MID_ANT_LAT	20.13 ± 3.42	17.69 ± 5.07	20.19 ± 3.83	16.79 ± 4.48	**15.19 ± 4.89**
SHORT_MID_INF_LAT	14.55 ± 1.11	13.78 ± 2.42	15.96 ± 2.55	15.32 ± 2.73	**12.12 ± 1.78**
SHORT_MID_INF	11.79 ± 3.85	11.09 ± 3.48	8.72 ± 2.95	7.88 ± 2.30	**6.28 ± 1.93**
SHORT_MID_ANT_SEPT	14.42 ± 3.05	12.76 ± 3.11	14.62 ± 2.45	16.09 ± 2.96	**11.35 ± 1.85**
SHORT_BASAL_ANT	59.10 ± 4.04	25.83 ± 3.58	31.92 ± 5.97	28.59 ± 6.00	**22.44 ± 3.61**
SHORT_BASAL_ANT_LAT	13.01 ± 2.53	21.09 ± 3.69	17.05 ± 3.42	16.92 ± 4.45	**10.51 ± 2.11**
SHORT_BASAL_INF_LAT	**16.22 ± 1.80**	21.03 ± 3.40	21.15 ± 3.32	18.65 ± 2.52	**16.22 ± 1.80**
SHORT_BASAL_INF	**16.28 ± 2.11**	21.92 ± 3.82	22.05 ± 3.74	17.95 ± 3.23	17.12 ± 2.88
SHORT_BASAL_ANT_SEPT	**17.76 ± 2.36**	26.79 ± 5.54	**17.76 ± 2.36**	**17.76 ± 2.36**	**17.76 ± 2.36**
HORIZ_LONG_SEPT	7.95 ± 2.70	**5.58 ± 2.08**	7.12 ± 2.47	5.58 ± 1.74	**5.58 ± 1.74**
HORIZ_LONG_APICAL	18.46 ± 3.86	21.92 ± 3.23	25.13 ± 2.70	18.59 ± 3.76	**15.45 ± 2.57**
HORIZ_LONG_LAT	16.86 ± 3.24	16.73 ± 3.98	16.03 ± 2.00	16.03 ± 3.42	**12.76 ± 2.40**
HORIZ_LONG_BASAL	8.85 ± 2.59	9.68 ± 2.40	12.12 ± 2.75	10.45 ± 1.71	**7.95 ± 2.39**
VERT_LONG_ANT	58.14 ± 2.02	39.36 ± 4.44	31.86 ± 4.07	**30.19 ± 4.15**	**30.19 ± 4.15**
VERT_LONG_APICAL	**16.86 ± 2.81**	20.90 ± 3.19	21.73 ± 3.60	17.63 ± 2.87	**16.86 ± 2.81**
VERT_LONG_INF	**15.13 ± 3.08**	21.54 ± 3.82	29.81 ± 4.19	15.90 ± 3.18	**15.13 ± 3.08**
VERT_LONG_BASAL	12.05 ± 2.99	19.29 ± 4.58	25.71 ± 3.49	12.05 ± 2.99	**11.99 ± 2.43**
Male					
SHORT_AP_ANT	19.41 ± 1.53	26.47 ± 2.36	19.41 ± 2.33	17.65 ± 2.32	**16.47 ± 2.29**
SHORT_AP_LAT	13.53 ± 2.33	12.94 ± 2.45	12.94 ± 2.11	11.18 ± 2.23	**10.00 ± 1.97**
SHORT_AP_INF	38.09 ± 3.30	39.96 ± 3.51	32.28 ± 2.44	32.90 ± 3.08	**28.68 ± 3.32**
SHORT_AP_SEPT	**9.41 ± 2.35**	12.94 ± 1.92	13.53 ± 2.64	10.59 ± 1.47	**9.41 ± 2.18**
SHORT_MID_ANT	**17.06 ± 2.05**	21.18 ± 2.35	20.59 ± 2.67	20.59 ± 2.67	**17.06 ± 2.05**
SHORT_MID_ANT_LAT	**7.56 ± 1.76**	**7.65 ± 1.26**	8.24 ± 1.57	**7.65 ± 1.76**	7.65 ± 1.76
SHORT_MID_INF_LAT	38.38 ± 4.26	38.38 ± 2.58	33.09 ± 3.26	33.05 ± 3.25	**29.52 ± 2.99**
SHORT_MID_INF	39.74 ± 4.51	41.32 ± 4.32	27.17 ± 4.46	25.99 ± 4.38	**25.92 ± 5.04**
SHORT_MID_ANT_SEPT	**14.23 ± 2.53**	17.76 ± 3.04	21.88 ± 3.03	14.82 ± 2.54	**14.23 ± 2.53**
SHORT_BASAL_ANT	12.94 ± 2.75	14.12 ± 2.66	17.06 ± 2.97	12.94 ± 2.75	**12.35 ± 2.55**
SHORT_BASAL_ANT_LAT	**3.53 ± 0.96**	4.12 ± 1.26	**3.53 ± 0.96**	**3.53 ± 0.96**	**3.53 ± 0.96**
SHORT_BASAL_INF_LAT	40.18 ± 4.75	33.68 ± 4.17	33.09 ± 2.89	31.91 ± 3.02	**30.74 ± 3.56**
SHORT_BASAL_INF	36.65 ± 4.16	37.83 ± 4.00	31.36 ± 3.16	28.90 ± 4.39	**27.72 ± 3.00**
SHORT_BASAL_ANT_SEPT	**14.12 ± 2.51**	17.65 ± 2.63	18.82 ± 3.70	**14.12 ± 2.51**	**14.12 ± 2.51**
HORIZ_LONG_SEPT	17.79 ± 2.81	13.57 ± 3.03	17.13 ± 2.82	12.46 ± 2.26	**12.39 ± 2.83**
HORIZ_LONG_APICAL	34.71 ± 3.87	28.24 ± 3.49	25.88 ± 3.19	23.53 ± 3.82	**21.18 ± 2.66**
HORIZ_LONG_LAT	15.88 ± 3.17	18.82 ± 2.88	14.12 ± 3.53	14.12 ± 3.19	**11.76 ± 2.63**
HORIZ_LONG_BASAL	11.18 ± 2.05	13.53 ± 2.49	9.41 ± 2.51	8.82 ± 2.19	**8.24 ± 2.18**
VERT_LONG_ANT	**19.41 ± 2.33**	22.35 ± 2.29	20.00 ± 2.51	20.00 ± 2.51	**19.41 ± 2.33**
VERT_LONG_APICAL	43.75 ± 3.59	44.93 ± 1.89	39.60 ± 3.78	41.40 ± 2.27	**36.07 ± 3.52**
VERT_LONG_INF	**32.10 ± 3.26**	35.74 ± 3.19	**32.10 ± 3.26**	**32.10 ± 3.26**	**32.10 ± 3.26**
VERT_LONG_BASAL	17.06 ± 3.09	16.47 ± 3.49	15.88 ± 3.17	15.88 ± 3.17	**14.71 ± 3.07**
Average error	21.82	21.78	20.67	18.54	**17.55**
Average advantage	0.00	−7.85	−6.82	7.51	**11.98**

Table 7
Partial classification: ranking of dataset quality

Rank	Dataset	Advantage ratio (%)	Rank	Dataset	Advantage ratio (%)
1	F_SHORT_AXIS_BASAL_ANT	62	23	F_SHORT_AXIS_MID_INF_LAT	17
2	F_VERT_LONG_ANT	48	24	F_HORIZ_LONG_APICAL	16
3	F_SHORT_AXIS_MID_INF	47	25	M_SHORT_AXIS_AP_ANT	15
4	F_SHORT_AXIS_AP_ANT	41	26	M_VERT_LONG_BASAL	14
5	F_SHORT_AXIS_MID_ANT	40	27	F_SHORT_AXIS_AP_SEPT	10
6	M_HORIZ_LONG_APICAL	39	27	F_HORIZ_LONG_BASAL	10
7	M_SHORT_AXIS_MID_INF	35	29	F_SHORT_AXIS_AP_INF	6
8	M_HORIZ_LONG_SEPT	30	30	M_SHORT_AXIS_BASAL_ANT	5
9	F_HORIZ_LONG_SEPT	30	31	F_VERT_LONG_BASAL	1
10	F_SHORT_AXIS_AP_LAT	29	32	F_SHORT_AXIS_BASAL_ANT_SEPT	0
11	M_HORIZ_LONG_BASAL	26	32	M_SHORT_AXIS_BASAL_ANT_LAT	0
12	M_SHORT_AXIS_AP_LAT	26	32	M_SHORT_AXIS_BASAL_ANT_SEPT	0
12	M_HORIZ_LONG_LAT	26	32	M_SHORT_AXIS_MID_ANT_SEPT	0
14	M_SHORT_AXIS_AP_INF	25	32	M_VERT_LONG_INF	0
14	F_SHORT_AXIS_MID_ANT_LAT	25	32	F_SHORT_AXIS_BASAL_INF_LAT	0
16	M_SHORT_AXIS_BASAL_INF	24	32	F_VERT_LONG_INF	0
16	F_HORIZ_LONG_LAT	24	32	M_SHORT_AXIS_MID_ANT	0
16	M_SHORT_AXIS_BASAL_INF_LAT	24	32	F_VERT_LONG_APICAL	0
19	M_SHORT_AXIS_MID_INF_LAT	23	32	M_VERT_LONG_ANT	0
20	F_SHORT_AXIS_MID_ANT_SEPT	21	32	M_SHORT_AXIS_AP_SEPT	0
21	F_SHORT_AXIS_BASAL_ANT_LAT	19	43	M_SHORT_AXIS_MID_ANT_LAT	−1
22	M_VERT_LONG_APICAL	18	44	F_SHORT_AXIS_BASAL_INF	−5

radius of search had been determined by inspecting each of the normal LV models independently.

For each view in three-dimensional rest and stress images, two-dimensional MAX and TOT images were created using a cylindrical radial search (Section 4.3.1). Each of the two-dimensional images was normalized by dividing pixel intensity values by the largest pixel intensity found in the corresponding three-dimensional image. For each of the ROI, we extracted four features from two-dimensional images: maximum intensity in the region, mean intensity, median intensity and standard deviation of the intensity in the region. This gave a total of 16 features for each ROI. A separate dataset was created for each ROI; separate for males and females.

Table 5 shows counts for females and males for each diagnostic type (NL, normal; F, fixed defect; P, partially reversible defect; R, reversible defect; X, defect showing reverse redistribution; A, artifact). We used cases that had a complete cardiologist's evaluation record for each of the 22 ROI. For males, we used cases that had a correlation match of 0.95 or better with the normal LV model (see Section 4.2); this resulted in 170 patient cases. For females, we used cases that had a correlation match of 0.93 or better with the normal LV model; this resulted in 124 cases.

The 'Total' column in Table 5 shows the total number of cases in each dataset. This does not always add up to 124 or 170 because of the coding errors in the databases that we were not able to resolve. For example, sometimes there was more than a single code recorded for a particular ROI. We did not drop these cases completely since they had correct diagnoses for other ROI and could be used for creation of other datasets.

The last two rows in Table 5 show the total number of times each defect was present in the 22 ROIs. Due to the low number of cases and low relative count for defect types other than normal, we combined all the defect codes into one abnormal. Thus, each of the datasets used for experiments had two classes (normal and abnormal) described by 16 attributes. There are 22 datasets for females and 22 datasets for males.

5.3.2. Experiments

Summary results are shown in Table 6. We used the constant classifier, C4.5, naïve Bayes, tree-augmented naïve Bayes and 25 new BNCs. Table 6 shows only the best performing new Bayesian classifiers; detailed results for each of the new classifiers can be found in [37]. The first number in a row is the 10-fold cross-validation error; the next is the standard deviation of the mean. Numbers in bold indicate the lowest error rate for a given dataset. The bottom of the table shows two indicators for each classifier: the average error rate and the average advantage ratio. The advantage ratio compares performance of a classifier to the constant classifier; it is described below.

5.3.3. Quality of the datasets with features extracted from SPECT images

By quality of a dataset, we understand the amount of information in the dataset that can be utilized to perform classification with a low error rate. In the extreme case (for supervised learning) a dataset may contain information only in the class variable, while all attribute variables contain noise. A reasonable decision in this case is to create a classifier that will always predict the class that is most frequent in the training data. This is the principle of the constant classifier—it always predicts the majority class regardless of the values of its attributes. Since constant classifier is the simplest possible classifier, we use its error rate as a reference for other classifiers.

If we can build a classifier that produces an error rate significantly lower than the constant classifier, we can say that this dataset is of good quality. For a perfect dataset, we should be able to build a classifier that has an error rate close to zero regardless of the error rate of the constant classifier. We rank quality of a dataset using the advantage ratio [37]:

$$\text{advantage ratio} = \frac{\varepsilon_{\text{const}} - \varepsilon_{\text{best}}}{\varepsilon_{\text{const}}} \times 100\%$$

where $\varepsilon_{\text{const}}$ is the error rate of the constant classifier for a given dataset and $\varepsilon_{\text{best}}$ the error rate of the best classifier built for that dataset. The advantage ratio defined above reaches 100% if an error rate equals to zero; it is 0% if the error rate of our best classifier is the same as of the constant classifier; it is negative if the best classifier produces an error rate higher than the constant classifier.

Table 7 shows ranking of datasets in decreasing order of the advantage ratio. We calculated the advantage ratio using 10-fold cross-validation results (Table 6). We say that datasets with advantage ratio close or less than zero are of low quality. A direct comparison of the classification error of the constant classifier versus a best classifier for a given dataset is shown in Fig. 20.

We see from Table 7 that there was a significant variability in the quality of the datasets. Dataset F_SHORT_AXIS_BASAL_ANT had the highest advantage ratio that indicates
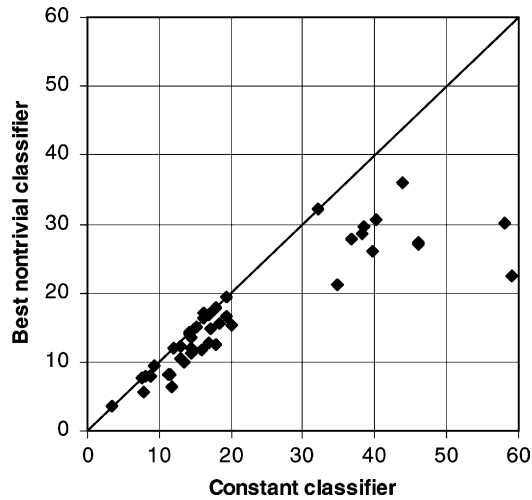
Fig. 20. Quality of the partial LV classification datasets: comparison of the constant classifier error rate to a best nontrivial classifier tested.

that it contained the most useful information for classification, but its best error rate of 22.44 % is quite high. The dataset M_SHORT_AXIS_BASAL_ANT_LAT had the lowest error rate of 3.53% but it had the advantage ratio equal to zero; indicating that it does not contain information useful for classification. Table 5 shows that this dataset contains no more than two examples for each of the defect types.

There are three factors that determine the quality of a dataset. First, the number of cases in a dataset; second, the noise present in coding values of the class variable; and third, the information coding scheme for the attributes. The advantage ratio is only able to judge these factors cumulatively; however, we can combine it with information shown in Tables 5 and 6. The overall conclusion is that the results of partial classification look very promising but a database containing a significantly larger number of cases would be needed in order to achieve more reliable results.

5.4. Overall classification of left ventricular perfusion

In Section 5.3, we described partial classification of LV perfusion to classify each of the 22 three-dimensional ROIs separately. The goal of the experiment described in this section was to determine performance of the overall classification of LV perfusion. We used the same feature extraction process as above, but we created different datasets. We selected cases that had a single overall perfusion code recorded in the database (Table 1). In addition, we used only defect types that had at least five cases. As before, we only used the cases that had sufficiently high LV model fit correlation ratio (0.95 for males and 0.93 for females). These cases are shown in Table 8.

The difficulty was with a low number of cases and a large number of attributes. Again, 16 features were extracted for each of the 22 three-dimensional ROI. This resulted in 352

Table 8
Case counts for overall LV perfusion classification tests

Sex	No. of attributes	No. of cases	Defect codes			
			NL	IS	INF	IS–IN
Female	44	74	36 (49)	16 (22)	14 (19)	8 (11)
Male	44	118	23 (19)	20 (17)	30 (25)	45 (38)

The values given in parenthesis are in percentage.

attributes available for the overall classification. Unfortunately, this was larger than the number of available cases, so we expected very high variance in estimation of the classifier parameters. Thus, we decided to use only one feature for each ROI in rest and stress images, resulting in 44 attributes. The feature was decided based on maximum, mean and median of pixel values in a three-dimensional region. We used either MAX or TOT radial search image. Thus, we had six feature types: MAX-Max, MAX-Mean, MAX-Median, TOT-Max, TOT-Mean and TOT-Median. One dataset was created for each of the six feature types, for females and males separately, resulting in 12 datasets. Each dataset had four classes: NL, normal; IS, ischemia; INF, infarct; IS–IN, ischemia–infarct.

Again, 10-fold cross-validation was performed using four reference classifiers (the constant classifier, C4.5, naïve Bayes and TAN) and 25 new BNCs. The summary of results is shown in Table 9. As before, the bold numbers indicate the lowest error for a given dataset. The best feature type, for both females and males, was the median of pixel intensity taken from the TOT images. We observed that the error rate and the advantage ratio was better for females than males even though the female sample was smaller. However, the male sample was also small compared to the number of attributes. Distribution of classes in the male sample was more uniform (see Table 8). In the

Table 9
Ten-fold cross-validation error rate of the overall LV perfusion classification using reference classifiers and the best results for the new Bayesian network classifiers

Sex	Image type	Feature type	Constant classifier	C4.5	Naïve Bayes	TAN	Best new BNC	Advantage ratio
Female	MAX	Max	**51.07 ± 4.39**	59.46 ± 2.78	**51.07 ± 4.39**	**51.07 ± 4.39**	**51.07 ± 4.39**	0.00
		Mean	**51.07 ± 4.39**	53.93 ± 5.07	57.86 ± 5.01	57.86 ± 5.01	57.86 ± 5.01	0.00
		Median	**51.07 ± 4.39**	61.96 ± 5.41	59.11 ± 5.29	59.11 ± 5.29	59.11 ± 5.29	0.00
	TOT	Max	51.07 ± 4.39	55.71 ± 4.51	53.75 ± 4.43	53.75 ± 3.24	**49.82 ± 3.63**	0.02
		Mean	51.07 ± 4.39	43.57 ± 3.71	43.04 ± 4.51	48.57 ± 5.44	**41.79 ± 4.46**	0.18
		Median	51.07 ± 4.39	48.21 ± 4.24	41.61 ± 4.73	40.36 ± 4.65	**38.75 ± 6.06**	0.24
Male	MAX	Max	61.82 ± 3.15	69.62 ± 4.84	62.80 ± 5.09	62.80 ± 5.38	59.47 ± 5.57	0.04
		Mean	61.82 ± 3.15	64.55 ± 3.14	61.06 ± 3.26	60.23 ± 2.68	55.15 ± 5.14	0.11
		Median	61.82 ± 3.15	66.21 ± 4.39	58.56 ± 3.95	56.74 ± 3.17	54.17 ± 3.63	0.12
	TOT	Max	**61.82 ± 3.15**	74.47 ± 3.69	66.29 ± 4.61	63.64 ± 3.21	62.73 ± 3.05	0.00
		Mean	61.82 ± 3.15	57.58 ± 3.13	61.06 ± 3.43	**55.30 ± 4.31**	55.98 ± 4.20	0.11
		Median	61.82 ± 3.15	71.89 ± 02.96	62.80 ± 3.46	61.97 ± 3.23	**57.73 ± 3.60**	0.07

The last column shows the value of the advantage ratio of the best classifier for a given dataset (higher is better).

female sample class, NL appeared in almost half of the cases, thus making it an easier concept to learn.

6. Discussion

New Bayesian network search algorithms were used as a data mining tool for inter-pretation of cardiac SPECT images. The Bayesian algorithms were based in part on naïve Bayes and TAN classifiers. The data used in this work was obtained from a clinical database and from the set of SPECT images. The SPECT image data were from two-dimensional tomographic slices of reconstructed three-dimensional radioactive count representations of myocardial perfusion. The clinical database included information, such as gender, body mass index, smoking, medications and exercise performance. Significant difficulty was caused by considerable noise present in the database, such as typographical errors, inconsistencies in recording and noise in the images. The noise is inherent in radionuclide techniques (such as SPECT) from many factors including radiation scatter, multiple levels of origin of the photons, variation in photon count and poor resolution. Another problem was missing values. We dealt with the problem by just removing all the records with missing values; this was possible because of the fairly large size of the initial dataset. For smaller datasets the approach could be to substitute missing values with most likely values or to replace the missing value with all possible values for a specific attribute. Still another problem was inter-patient variability in heart size, motion, position, rotation and artifacts from other soft tissues, as well as from the presence of perfusion defects. To address these problems we cleaned the image and clinical databases, restructured two-dimensional data into three-dimensional structures and created ''normal'' three-dimen-sional models for reference. The data from 22 regions of the two-dimensional slices were then projected onto the three-dimensional model.

We used the cardiologists' qualitative interpretations as the gold standard for compar-ison. Each of the 22 regions for each patient, along with its diagnostic classification, formed a separate database. The problem with this approach was a large number of variables with relatively few examples of specific diagnoses for each of the 22 regions. In addition, the gold standard itself had an error rate of 17% when the cardiologists' diagnoses were compared to the counts-based objective data. In spite of these problems the performance of Bayesian classifiers was relatively good for both partial and total classifications. We found that the error rate and the advantage ratio were better for females than males. This was probably due to more homogeneity in the data: women had many more regions classified as ''normal'' than men. A much larger dataset, with more examples for diagnoses in each of the regions would be needed for better classification results. However, the best achieved error rate (18%) was essentially the same as the error rate of the gold standard.

The major contributions of this work include: (a) the method for construction of normal gender-specific LV models; (b) image registration; (c) the new Bayesian learning classi-fication methods; (d) the use of the six-step knowledge discovery process that was very successful in this complex data mining task; as such, the process can be extended to other medical data mining projects. As an additional benefit to the readers we made the software

package, which includes all BNCs used in the experiments, available (including source code) free of charge at: http://jbnc.sourceforge.net/.

References

[1] Castillo E, Gutiérrez JM, Hadi AS. Expert systems and probabilistic network models. New York: Springer, 1997.

[2] Cheng J, Bell DA, Liu W. An algorithm for Bayesian belief network construction from data. In: Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics, AI&STAT'97, 1997. p. 83–90.

[3] Cheng J, Greiner R. Comparing Bayesian network classifiers. In: Proceedings of the 15th International Conference on Uncertainty in Artificial Intelligence. Los Altos, CA: Kaufmann, 1999.

[4] Chow CK, Liu CN. Approximating discrete probability distributions with dependence trees. IEEE Trans Inf Theory 1968;14:462–7.

[5] 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, Ann Arbor, 1–5 June 1996.

[6] Cooper GF, Herskovits E. Bayesian method for induction of probabilistic networks from data. Machine Learn 1992;9:309–47.

[7] Corbett JR. Gated blood-pool SPECT. In: DePuey ED, Berman DS, Garcia EV, editors. Cardiac SPECT imaging. New York: Raven, 1995.

[8] Cormen TH, Leiserson CE, Rivest RL. Introduction to algorithms. Cambridge: MIT Press, 1990.

[9] Cuaron A, Acero AP, Cardenas M, Huerta D, Rodriguez A, de Garay R. Inter-observer variability in the interpretation of myocardial images with Tc-99m-labeled diphosphonate. J Nucl Med 1980;21(1):1–9.

[10] Cullom SJ. Principles of cardiac SPECT. In: DePuey ED, Berman DS, Garcia EV, editors. Cardiac SPECT imaging. New York: Raven, 1995.

[11] Declerck J, Feldmar J, Goris ML, Betting F. Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images. IEEE Trans Med Imaging 1997;16(6):727–37.

[12] Duda RO, Hart PE. Pattern classification and scene analysis. New York: Wiley, 1973.

[13] Even S. Graph algorithms. Rockville: Computer Science Press, 1979.

[14] Ezquerra N, Mullick N, Cooke N, Garcia E, Krawczynska E. PERFEX: an expert system for interpreting three-dimensional myocardial perfusion. Technical Report GIT-GVU-92-02 Graphics, Visualization & Usability Center. Atlanta, GA: Georgia Institute of Technology, 1992.

[15] Faber TL, Akers MS, Peshock RM, Corbett JR. Three-dimensional motion and perfusion quantification in gated single-photon emission computed tomograms. J Nucl Med 1991;32:2311–7.

[16] Faber TL, Cooke CD, Peifer JW, Pettigrew RI, Vansant JP, Leyendecker JR, et al. Three-dimensional displays of left ventricular epicardial surface from standard cardiac SPECT perfusion quantification techniques. J Nucl Med 1995;36:697–703.

[17] Francisco DA, Collins SM, Go RT, Ehrhardt JC, Van Kirk OC, Marcus ML. Tomographic thallium-201 myocardial perfusion scintigrams after maximal coronary vasodilation with intravenous dipyridamole: comparison of qualitative and quantitative approaches. Circulation 1982;66(2):370.

[18] Friedman JH. On bias, variance, 0/1-loss and the curse-of-dimensionality. Data Mining Knowledge Discov 1997;1(1):55–77.

[19] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Machine Learn 1997;29(2/3):131–63.

[20] Garcia EV, Ezquerra NF, DePuey EG, Hise HL, Robbins W. An artificial intelligence approach to interpreting thallium-201 three-dimensional myocardial distributions. J Nucl Med 1986;27:1005[abstract].

[21] 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.

[22] Haddawy P. An overview of some recent developments in Bayesian problem-solving techniques. Artif Intell Mag 1999;20(2):11–9.

[23] Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learn 1995;20:197–243.

[24] Jensen FV. An introduction to Bayesian networks. Berlin: Springer, 1996.

[25] John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, San Mateo, CA. Los Altos, CA: Kaufmann, 1995.

[26] Jordan MI, editor. Learning in graphical models NATO science series. Dordrecht: Kluwer Academic Publishers, 1998.

[27] Kononenko I. Comparison of inductive and naïve Bayesian learning approaches to automatic knowledge acquisition. In: Welinga B, et al., editors. Current trends in knowledge acquisition. Amsterdam: IOS Press, 1990.

[28] Kononenko I. Semi-naïve Bayesian classifier. In: Proceedings of the Sixth European Working Session on Learning Porto, Portugal. London: Pittman, 1991. p. 206–19.

[29] Lam W, Bacchus F. Learning Bayesian belief networks: an approach based on the MDL principle. Comput Intell 1994;10:269–93.

[30] Langley P, Iba W, Thompson W. An analysis of Bayesian classifiers. In: Proceedings of the 10th National Conference on Artificial Intelligence. San Jose: AAAI Press, 1992.

[31] Langley P, Sage S. Induction of selective Bayesian classifiers. In: Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, Seattle, WA. Los Altos, CA: Kaufmann, 1994.

[32] Langley P, Sage S. Tractable average-case analysis of naïve Bayesian classifier. In: Proceedings of the 16th International Conference on Machine Learning, Bled, Slovenia. Los Altos, CA: Kaufmann, 1999.

[33] Pearl J. Probabilistic reasoning in intelligent systems: networks for plausible inference. Los Altos, CA: Kaufmann, 1988.

[34] Qian J, Mitsa T, Hoffman E. Contour/surface registration using physically deformable model. In: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, San Francisco, 1996.

[35] Quinlan JR. C4.5: programming for machine learning. Los Altos, CA: Kaufmann, 1993.

[36] Rice JA. Mathematical statistics and data analysis. Belmont, CA: Wadsworth, 1988.

[37] Sacha JP. New synthesis of Bayesian network classifiers and interpretation of cardiac SPECT images. Ph.D. Dissertation, University of Toledo, 1999.

[38] Sacha JP, Cios KJ, Goodenday LS. Issues in automating cardiac SPECT diagnosis. IEEE Eng Med Biol Mag 2000;19(4):78–88.

[39] Sacha JP, Shabestari BN, Cios KJ. Structured region growing and recognition algorithm for non-destructive evaluation. In: Proceedings of the SPIE's Annual Meeting on Applications of Digital Image Processing XIX, Denver, Colorado, 1996. p. 87–94.

[40] Schroeder W, Martin K, Lorensen B. The visualization toolkit: object-oriented approach to three-dimensional graphics. Upper Saddle River, NJ: Prentice-Hall, 1998. http://www.kitware.com.

[41] Singh M, Provan G. A comparison of induction algorithms for selective and non-selective Bayesian classifiers. In: Proceedings of the 12th International Conference on Machine Learning, 1995. p. 497–505.

[42] Suzuki J. A construction of Bayesian networks from databases based on an MDL scheme. In: Heckerman D, Mamdani A, editors. Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA. Los Altos, CA: Kaufmann, 1993. p. 266–73.

[43] 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, 1995.