Automated Heart Wall Motion Abnormality Detection From Ultrasound Images using Segmental Knowledge

Maleeha Qazi
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, USA
maleeha.qazi@siemens.com

Glenn Fung
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, USA
glenn.fung@siemens.com

Sriram Krishnan
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, USA
sriram.krishnan@siemens.com

ABSTRACT

Coronary Heart Disease can be diagnosed by measuring and scoring regional motion of the heart wall in ultrasound images of the left ventricle (LV) of the heart. Studies have shown that the quality of diagnosis has great inter- and intra-observer variation, even among world-class expert cardiologists. This variability in diagnosis quality is particularly critical for CHD, because early diagnosis is a key factor in improved prognosis. We describe a completely automated and robust technique that detects diseased hearts based on automatic detection and tracking of the endocardium and epicardium of the LV. The local wall regions and the entire heart are then classified as normal or abnormal based on the regional and global LV wall motion. In order to leverage structural information about the heart we applied Bayesian Networks to this problem, and learnt multiple potential structures off of the data using different algorithms. We checked the validity of these structures using anatomical knowledge of the heart and medical rules as described by doctors. We then used a novel feature selection technique based on mathematical programming to add imaging features to the basic learnt structures. The resultant classifiers thus depend only on a small subset of numerical features extracted from dual-contours tracked through time. We verify the robustness of our systems on echocardiograms collected in routine clinical practice at one hospital, both with the standard cross-validation analysis, and then on a held-out set of completely unseen echocardiography images.

Keywords

Coronary Heart Disease, echocardiogram, heart wall motion, feature selection, classification, Bayesian Networks, Recursive Minimal Entropy Partitioning, Kolmogorov-Smirnov.

1. INTRODUCTION

Early detection (along with prevention) is an excellent way of controlling *Coronary Heart Disease* (CHD). CHD (along with Congestive Heart Failure) can be detected by measuring and scoring the regional and global motion of the left ventricle (LV) of the heart; CHD typically results in *wall-motion abnormalities*, i.e., local segments of the LV wall move abnormally (move weakly, not at all, or out of sync with the rest of the heart), and sometimes motion in multiple regions, or the entire heart, is compromised. The LV can be imaged in a number of ways. The most common method is the echocardiogram – an *ultrasound video of different 2-D cross-sections of the LV*. Echocardiograms are unfortunately notoriously difficult to interpret, and even the best of physicians can misdiagnose heart disease. Hence there is a tremendous need for an automated "second-reader" system that can provide objective diagnostic assistance, particularly to the less-experienced cardiologist. Inter-observer studies have shown that even *world-class experts agree on their diagnosis only 80% of the time* [19], and intra-observer studies have shown a similar variation when the expert reads the same case twice at widely different points in time. Furthermore, the agreement between less-experienced cardiologists and the experts is often below 50%.

In this paper, we address the task of building a computer-aided diagnosis system that can automatically detect wall-motion abnormalities from echocardiograms. Our goal is to develop a system to assist physicians to interpret wall motion scores, and thereby reduce variability and improve diagnostic accuracy of wall motion analysis. Section 2 provides some medical background on cardiac ultrasound and the standard methodology used by cardiologists to score wall-motion abnormalities. In Section 3 we describe our real-life dataset, which consists of echocardiograms used by cardiologists at Erasmus Medical Centre to diagnose wall-motion abnormalities. Sections 4 and 5 provide an overview of our proposed system which we built on top of an algorithm that detects and tracks the inner and outer cardiac walls [17, 24, 13, 14]. It consists of a classifier that classifies the local region of the heart wall (and the entire heart) as normal or abnormal based on the wall motion. We describe our methodology for feature selection and classification, followed in Section 6 by our experimental results. We conclude with some thoughts about our plans for future research in Section 7.

2. MEDICAL BACKGROUND KNOWLEDGE

2.1 What is Coronary Artery Disease?

The human heart is divided into four chambers: the left and right atrium, and the left and right ventricle. The left ventricle (LV) is the chamber responsible for pumping oxygenated blood to the entire body. As a result, it is the largest and strongest of the four chambers. Figure 2.1 shows the layout of the heart chambers in relation to one another; the LV is in the lower right part of the figure. The heart is fed by three major coronary arteries: the left anterior descending(LAD), right coronary artery (RCA), and the left circumflex coronary artery(LCX). All three of these vessels feed the

KDD 2006 Philadelphia, Pennsylvania USA

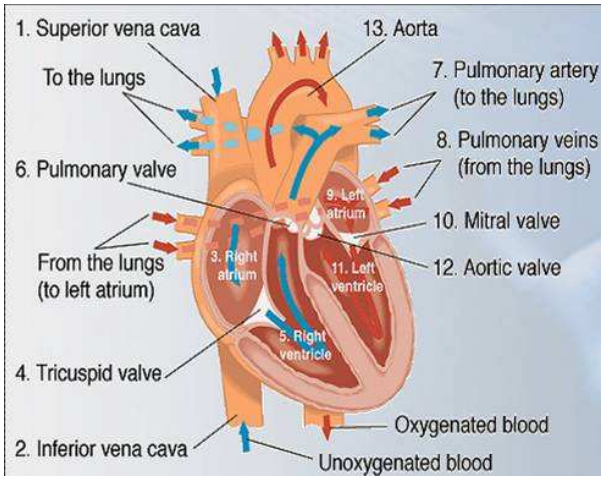

Figure 1: Major parts of heart labeled, including the four chambers of the human heart: the left and right atrium, and the left and right ventricle.

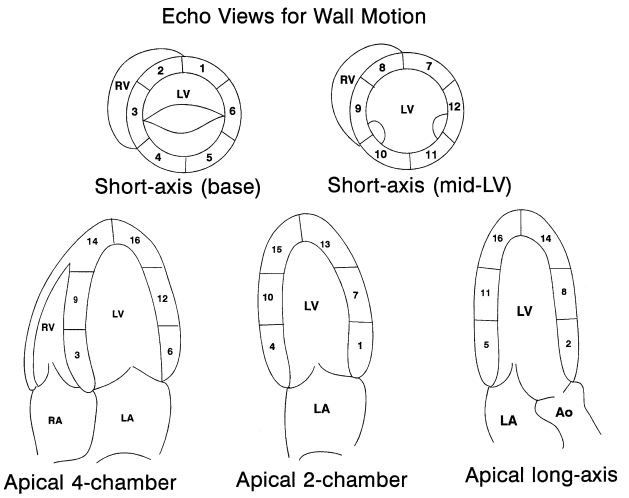

Figure 2: Echocardiographic views for wall motion evaluation. In the short-axis view, at the base and midventribular levels, the left ventricle is divided into anterior septum (2,8) and anterior free wall (1,7), lateral (6,12), posterior (5,11), inferior free wall (4,10), and posterior septal (3,9) segments. These same wall segments are seen in apical views as indicated, plus the anterior(13), septal (14), inferior (15), and lateral (16) apical segments are seen. Modified from reference "Textbook of Clinical Echocardiography" (segment numbers have been corrected to reflect standard naming convention being used).

muscle surrounding the LV. *Coronary artery disease* results from the development of plaque within the artery, which usually deposits along the walls. When the plaque restricts normal blood flow to an extreme extent the patient will experience chest pain, known as *angina*. When the blood flow to the heart muscle is reduced, the function of that piece of muscle fed by the blocked artery will begin to become impaired. This is known as *ischemia*. This functional impairment can be seen from ultrasound images of the heart, also called echocardiograms (echos). One of the first effects of coronary artery disease is that the motion of the heart wall during contraction will become impaired. Accurate regional wall motion analysis of the LV is an essential component of interpreting echos to detect this effect, particularly for early disease.

Heart disease has no gender, geographic or socio-economic boundaries. Cardiovascular Disease (CVD) is a global epidemic that is the leading cause of death worldwide (17 mil. deaths per year) [22]. Since 1900, CVD has been the No. 1 killer in the USA every year except 1918. It claims more lives each year than the next 4 leading causes of death combined, which are cancer, chronic lower respiratory diseases, accidents, and diabetes mellitus [5]. *Coronary Heart Disease* (CHD) accounts for more than half the CVD deaths (roughly 7.2 mil. deaths worldwide every year, and 1 of every 5 deaths in the US), and is the *single* largest killer in the world. According to the CDC/NCHS, if all forms of major CVD were eliminated, life expectancy would rise by almost 7 years. By comparison, if all forms of cancer were eliminated, the gain would be 3 years [5]. Accurate early diagnosis of CHD - primarily with cardiac ultrasound - has been identified as a critical factor in improving patient outcomes for CHD.

2.2 Divisions of the Heart

There are many imaging modalities that have been used to measure myocardial perfusion, left ventricular function, and coronary anatomy for clinical management and research; for this project we chose to use echocardiography. The Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association has created a standardized recommendation for the orientation of the heart, angle selection and names for cardiac planes and number of myocardial segments [9]. This is the standardization used in this project.

Echo images are collected from four standard views: apical 4 chamber (A4C), apical 2 chamber (A2C), parasternal long axis (PLAX) or apical 3 chamber (A3C), and parasternal short axis (PSAX) – shown in Figure 2 The planes used to cut the heart to display these standard views are displayed in Figure 3 from reference [7]. The long-axis view extends from the LV apex through the aortic valve plane. The short-axis view is perpendicular to the long-axis view resulting in a circular view of the LV. The four-chamber view is perpendicular to both the long- and short-axis views and includes the left and right ventricle, and left and right atrium. If one rotates the 4-chamber view plane counterclockwise about 60 degrees, the two-chamber view is obtained which shows the LV and the left atrium.

The left ventricle (LV) is divided into 17 myocardial segments. The short-axis view which results in a circular view of the LV can be taken at 3 locations, near the apex (apical), at the middle (mid-cavity), or near the base (basal). The most desirable being the mid-cavity cut. If one lays these 3 resultant rings against one another, all segments of the heart are visible in relationship to one another, as shown in Figure 4 (modified from reference [9]). The left anterior descending(LAD) feeds segments 1, 2, 7, 8, 13, 14 and 17, the right coronary artery (RCA) feeds segments 3, 4, 9, 10 and 15, and the left circumflex coronary artery (LCX) feeds segments 5, 6, 11, 12 and 16.

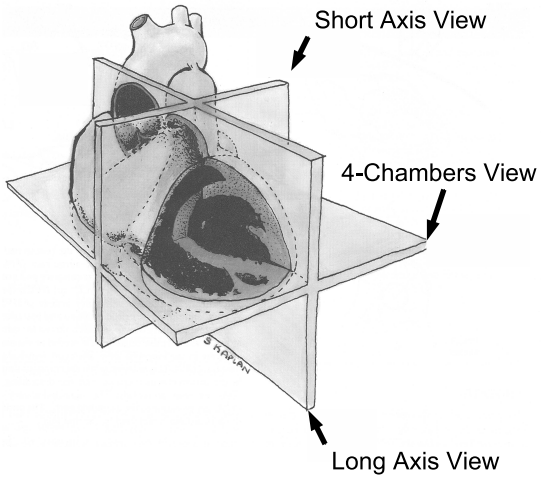

Figure 3: The three basic image planes used in transthoracic echocardiography. The ventricles have been cut away to show how these image planes intersect the left and right ventricles. *Dashed lines* indicate the image planes at the great vessel and atrial levels. From reference "Textbook of Clinical Echocardiography".

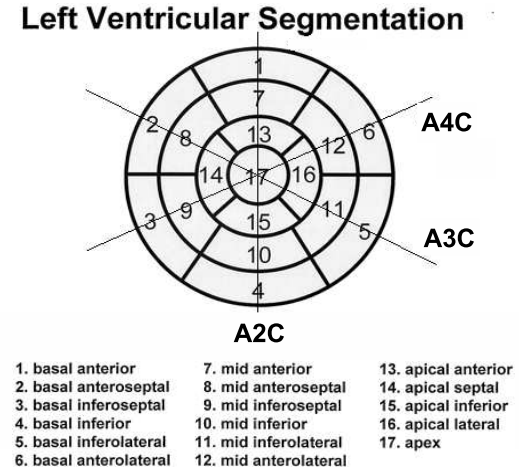

Figure 4: Display, on a circumferential polar plot, of the 17 myocardial segments and the recommended nomenclature for tomographic imaging of the heart. Modified from reference "Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart".

3. UNDERSTANDING THE DATA

The data is based on standard adult transthoracic B-mode ultrasound images collected from the four standard views described previously. Currently we utilize the three apical views - A2C, A3C, and A4C - which show all 16 segments of interest (we ignore the apex, segment 17, since it is near unto impossible to measure with ultrasound). These provide all the information needed to achieve our goal of classifying hearts.

Even though we have images at different levels of stress (resting, low-dose stress, peak-dose stress, recovery) this work is based on images taken when the patient was resting. The goal is to automatically provide an initial score, or classification, to determine whether a heart is normal or abnormal given the resting ultrasound.

The echo data was collected from Erasmus Medical Centre, Rotterdam, The Netherlands (abbrev: Erasmus). The Erasmus data consists of 345 cases for which we have associated images as well as ground truth and 2216 cases for which we only have ground truth (no images); all of which were generated using pharmacological stress, which allows the physician to control the amount of stress a patient experiences (in this case induced by dobutamine). All the cases have been labeled at the segment level by a group of trained cardiologists; this is what we refer to as "ground truth". Each of the 16 segments were labeled 1 - 5 (1 = normal, 2 = hypo-kinetic, 3 = akinetic, 4 = dyskinetic, 5 = aneurysm), for simplification purposes we converted this 5-class problem to a binary class problem (1 = normal, 2 - 5 = abnormal) for most of the tests we will describe (unless the 5-class problem is mentioned specifically, assume binary problem). The heart level classification labels can be obtained from the segment level labels by applying the following doctor provided definition: A heart is considered abnormal if two or more segments are abnormal.

4. PREPARATION OF THE DATA

Our application consists of two main parts: image processing, and classification. The echos are run through an algorithm which automatically detects and tracks both the interior (endocardial) and exterior (epicardial) borders of the LV [13, 14]. Motion interferences (e.g. probe motion, patient movement, respiration, etc.) are compensated for by using global motion estimation based on robust statistics outside the LV, this is done so that only the heart's motion is analyzed. Then numerical feature vectors, which are extracted from the dual-contours tracked through time, form the basis for the regional wall motion classification.

4.1 Image processing

The first step toward classification of the heart involves automatic contour generation of the LV [17]. Ultrasound is known to be noisier than other common medical imaging modalities such as MRI or CT, and echocardiograms are even worse due to the fast motion of the heart muscle and respiratory interferences. The framework used by our algorithm is ideal for tracking echo sequences since it exploits heteroscedastic (i.e. location-dependent and anisotropic) measurement uncertainties. The process can be divided into 2 steps:

1. Border detection: First the LV is localized on multiple frames of the image clip (shown in Figure 5 as a box drawn around the LV). Then the internal (endocardial) border of the LV is detected through contour shape inference (shape models learnt from a collection of training cases are used to help in this process). The shape of the border is represented by configurations of labeled control points or landmark points, assuming point correspondence from frame-to-frame. Landmark points are assigned based on anatomic features (e.g. the apex, the papillary muscles, etc.). Currently 17 control points are used for the apical views, and they are ordered in clockwise order, with the first point at the basal septal/inferior region and the mid-point assigned at the apex [24]. The external (epicardial) border is found by extending the internal (endocardial) border outwards [14].

2. Border tracking: Next the detected borders are tracked from one frame to the next through the entire video clip. The inner and outer contours are treated as a single shape for coupled

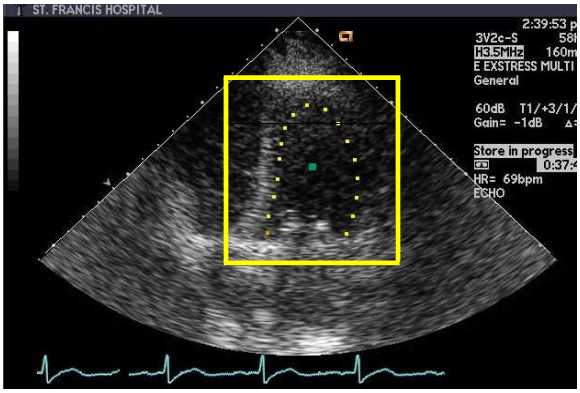

Figure 5: One frame from an A4C image clip with the (yellow) box showing the localized LV, and the (yellow) dots representing the control points along the detected contour.

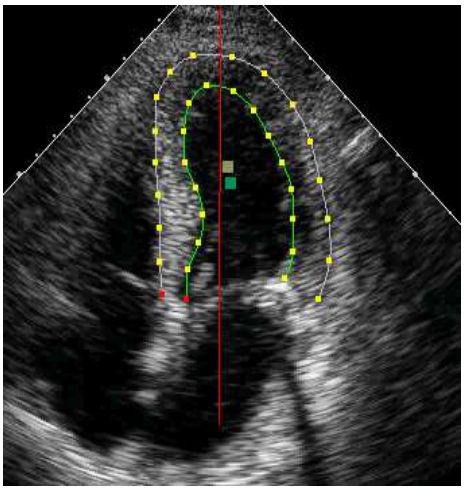

Figure 6: One frame from an A4C image clip with the outer and inner contour control points shown. The (red) vertical line shows use of global motion compensation, and the two squares denote the centers of the individual contours.

double-contour tracking. "Intuitively, a double-contour approach can propagate information from the endocardium to guide the localization of the epicardium (or vice versa), thus achieving more robust tracking of the two borders than tracking them independently" [24]. Motion interferences (e.g. probe motion, patient movement, respiration, etc.) are compensated for by using global motion estimation based on robust statistics outside the LV. This global motion estimation can be seen in Figure 6 as a vertical line near the center of the image. "The motion is always computed with reference to the neighborhood on the control point in the first frame of the sequence (i.e. the current frame is always compared to a model extracted from the first frame). Thus, error accumulation from frame to frame is avoided. Since the location of the model is updated at each frame, the motion estimation process always starts with a good initialization. The overall procedure is suitable for the tracking of periodic sequences such as the heart ultrasound data" [24].

After detection and tracking numerical features are computed from the dual-contours tracked through time. The features extracted are both global (involving the whole LV) and local (involving individual segments visible in the image), and are based on velocity, thickening, timing, volume changes, etc.

4.2 Extracted Features

A number of features have been developed to characterize cardiac motion in order to detect cardiac wall motion abnormalities, among them: velocity, radial and circumferential strain, local and global Simpson volume, global and local ejection fraction (EF) ratio, and segmental volume. Some of these features, including velocity, circumferential strain, and local EF ratio, are based on the inner (endocardial) contour.

Due to the patient examination protocol, only the systole (i.e. contraction phase of the heart) is recorded for some patients. In order for the features to be consistent, the systole is extracted from each patient based on the cavity area change. For each frame, the LV cavity area can be estimated accurately based on the inner (endocardial) contour of that frame. The frame corresponding to the maximal cavity area that is achieved at the end of diastolic phase (expansion phase of the heart) is the frame considered to be the beginning of systole. The frame corresponding to the minimal cavity area (achieved at the end of systolic phase) is the frame assumed to be the end of systole. For the time being, all features are computed based only on the systolic phase. However, the methods used to calculate the features are generally applicable for the diastolic phase as well. The following is a basic description of some of the features:

- Velocity features: determines how fast any pair of control points change in the x and y coordinate system per frame.
- Circumferential strain features: also called Fractional Shortening, measures how much the contour between any two control points shrinks in the systolic phase.
- Radial strain features: measures how much the contour shrinks for each control point between any two time frames.
- Local and Global Simpson Volume features: determine the volume as computed by the Simpson rule (with 50 disks) for each frame, and for the systolic phase of the heart as a whole (this gives the global EF, which captures the whole heart's contraction abnormalities).
- Segmental Volume features: determine the volume per segment per frame, and the segmental EF values (i.e local EFs, which aim to capture the local cardiac contraction abnormalities).

In general, the global version of certain features (e.g. radial strain, circumferential strain, etc) can be calculated by taking the mean, or standard deviation, of the 6 segment's respective feature values from any one view. All in all we had 120 local and global features for each of the 3 views (360 total), all of which were continuous.

5. DATA MINING

Our data mining process which forms the raw data (ultrasound videos) into the final outcome, a robust classifier that predicts/ assesses heart wall motion for the 16 segments of the heart, can be summarized as follows:

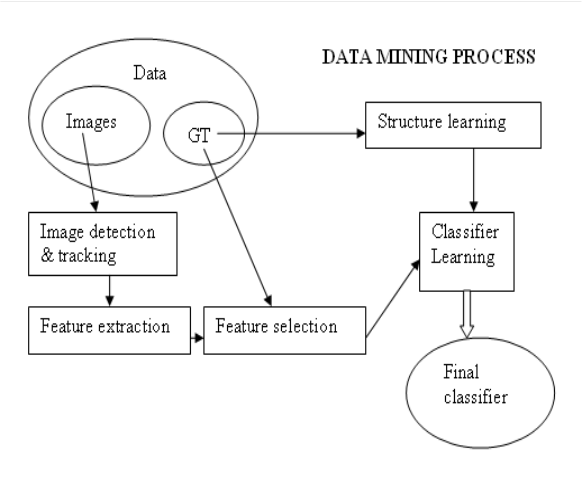

Figure 7: Diagram of process, from raw data to final classifier.

1. The image sequences are pre-processed to extract the area of interest (i.e. the LV). Intensity Normalization and an appropriate resizing are applied.
2. Given the initial contour, the border tracking is performed as described in section 4.1, item 2.
3. A set of features is extracted from the tracked contours as described in section 4.2.
4. Using the provided Ground Truth (assessments from the doctors), a network structure is learnt that represents the correlations among segments, this process is explained in detail in section 5.2.
5. For each segment, a subset of relevant features for classification is selected (section 5.1), as a result the classifiers we obtain only depend on a small number of numerical features extracted from the dual-contours tracked through time. Having classifiers depending on a small number of features not only improves performance time but also results in better generalization.
6. This subset of numerical features is provided as evidence for the networks' variables of interest (the 16 heart segments) to provide classification for each segment and the whole heart. The entire process is diagrammed in figure 7.

5.1 Feature Selection

One of the difficulties in constructing a classifier for this task is the problem of feature selection. It is a well-known fact that a reduction on classifier feature dependence can improve the classifier's generalization capability. However, the problem of selecting an "optimal" minimum subset of features from a large pool (in the order of hundreds) of potential original features is known to be NP-hard. In this paper we use a very simple but efficient filter-based approach. In general, Filter-based methods aim to select features based on simple auxiliary criteria, such as feature correlation, to remove redundant features. We used the Kolmogorov-Smirnov test (KS-test) [6, 10] to measure the maximum difference between the empirical cumulative distributions of the two classes according to a given feature. All the available features are ranked by the score given by the KS-test and only the features at the top of the list are selected for classification. Note that a high rank according to the KS-test implies that the corresponding empirical cumulative distributions of the two classes according to these features are considerably different, making these features desirable for discrimination between the two classes.

5.2 Bayesian Networks

A probabilistic graphical model comprises of a graph and parameters. The nodes in the graph correspond to random variables, and the (lack of) arcs represent conditional independences. The parameters quantify the dependence among variables. Graphical models provide a compact representation of the joint probability distributions over random variables. Bayesian Networks or Belief Networks (BNs) are directed graphical models (e.g., see [4] for an overview): the BN structure is a directed acyclic graph (DAG), and the BN parameters are conditional probabilities, represented by a contingency table in case of discrete variables. The directed edges in the graph can often be interpreted in an intuitive way, namely as causal relationships (requiring additional assumptions) [21].

Learning graphical model structures from data is computationally expensive. In fact, finding the optimal graph with respect to a given scoring function was shown to be NP-complete [11]. In the last decades, many efficient heuristic learning algorithms have been developed for learning BN structures (e.g., [21, 18]), while learning (general undirected) Markov models is computationally much more expensive. The BN structure can only be learned up to Markov equivalence in principle (e.g., [21]), so that the direction of some edges cannot be determined based solely on given data. Moreover, small data sets introduce increased model uncertainty, obscuring the (true) graph structure and edge-directions. The orientation of those edges can only be determined based on additional background knowledge. Once a BN model is learned, not only can its graph be interpreted as to gain new insights in the interdependences of the variables, but also the BN model (structure and conditional probabilities) can be used for quantitative predictions: given the values of some variables as evidence, the probability distributions of the remaining variables can be inferred, either exactly (e.g. junction tree algorithm [4]) or approximately (e.g., loopy belief propagation [23]).

We applied BNs to automatically detecting heart abnormalities to utilize the structural knowledge of the heart segments in hopes of improving the classification accuracy for the whole heart. In this paper, the following representation will be used

- numbers between 1 and 16 will be used to name the respective heart segment being referred to.
- arrows (for example: $\leftarrow$, $\rightarrow$) are used to denote the direction of causality. An arrow is drawn from the cause (parent) node to the effect (child) node. Example: $5 \leftarrow 6$ should be read as 6 causes/influences 5, here 5 is the child node and 6 is the parent node.
- a line connecting two segments denotes that a link exists between the two, but the directionality is indeterminable (see above). Example: 1 – 7.
- Two segments are called "neighbors" if they share a common boundary. For example: the neighbors of 1 are 6, 7, and 2 (as shown in Figure 4). 8 and 12 are not neighbors since they lie diagonal to 1 and share only a point of contact with 1. They could only be influenced by 1 via one of the neighbors.

5.2.1 Prior Domain Knowledge expressed as Rules

We derived prior knowledge from prior clinical research – Moir & Marwick [20] are well-known cardiologists who have provided some basic rules about reading echos in standard views. Refer to Figure 2 [7] to see the standard views. The sentence "...a lateral segment should be matched with a posterior wall finding before circumflex disease is reported" extracted from [20] can be translated into the following rules:

lateral segments: 6, 12
posterior wall segments: 5, 11
rules: $6 \leftarrow 5$ and $12 \leftarrow 11$

And the sentence "...the basal inferior wall should be matched to an adjacent abnormal segment (basal septum, mid-inferior) to reduce false positive interpretations" can be translated into the following rule:

basal inferior wall: 4
basal septum: 3
mid inferior: 10
rule: $4 \leftarrow 3$ OR $4 \leftarrow 10$

We also received input from our cardiologist collaborators. We find that two rules of thumb that most doctors use while reading a patient's echos have to do with neighboring segments and shared coronary artery territory. Neighboring segments influence each others behavior. One strong example is if one looks at the basal inferior segment in only the A2C view and believes it to be abnormal, then the probability of it being abnormal increases if its nearest neighbors in the A2C or PSAX views are seen to be abnormal (ABN). This can be represented as:

basal inferior wall: 4
nearest neighbors: 10, or 3 and 5 (5 can be eliminated as a choice since it does not share blood territory with 4 as do 3 and 10)
rule: $4 \leftarrow 3$ OR $4 \leftarrow 10$
(NOTE: this same rule is also found in the above mentioned article [20])

Also if two segments are fed by the same coronary artery, one is ABN and the other unknown, then it is better to err on the side of caution and classify the unknown segment as ABN.

As for the location of a segment in a view (2) influencing it's chances of being a true or false positive one can logically conclude the following:

LAD: basal (1,2), mid-cavity (7,8), apical (13,14)
– All on right hand side of standard view images (Figure 2), thus are common sites of false positives and should not be used as a parent to another segment.

RCA: basal (3,4), mid-cavity (9,10), apical (15)
– All on left hand side of standard view images (Figure 2), thus are most likely to be true positives and do not need other segments as parents.

LCX: basal (5,6), mid-cavity (11,12), apical (16)
– Segments found on right hand sides of standard view images (6,12) should be backed up by segments found on left hand side of images (5,11) (Figure 2).
Rules: $6 \leftarrow 5$, $12 \leftarrow 11$.
(NOTE: these are the same rules as found in the above mentioned article [20])

5.2.2 Learning Bayesian Network structure from data

Three different algorithms were used in the attempt to find a structure between the heart segments from the data. The resulting structures was validated from our cardiac knowledge (as discussed in Section 2). We applied multiple methods to learn the structure because to create a structure by hand using only physical knowledge of the heart would have resulted in a cyclic graph. All the methods used found structures that actually mirrored the physical relationships within the heart but without creating cycles. The three algorithms that were used to find a structure between the heart segments are:

1. Chow and Liu algorithm [12]: An implementation of the dependence tree finding algorithm proposed by Chow and Liu was applied to the 5-class problem, with all 16 segments, and it came up with the undirected links shown in Table 1. All, but one, of these links make logical sense since the linked segments are neighbors and/or share blood territory. The link 10 – 5 does not make sense because the two segments fail both requirements. Also the link 14 – 16 does not technically satisfy both requirements, but it still makes some sense since the two segments are 'neighbors' in the standard views (and since we are not considering the 17th segment - the apex - they would be neighbors were that segment to be removed).

Table 1: Edges/Links learnt by different structure learning algorithms

16 seg Chow & Liu	16 seg HUGIN NPC	16 seg BNT
1 – 7, 7 – 13, 2 – 8, 8 – 9, 3 – 9, 9 – 14, 6 – 12, 12 – 16, 4 – 10, 10 – 5, 10 – 15, 5 – 11, 14 – 16, 14 – 15, 13 – 16	$1 \leftarrow 7$, $13 \leftarrow 7$, $2 \leftarrow 8$, $9 \leftarrow 3$, $9 \leftarrow 14$, $6 \leftarrow 12$, $16 \leftarrow 12$, $4 \leftarrow 10$, $15 \leftarrow 10$, $11 \leftarrow 5$, $16 \leftarrow 14$, $15 \leftarrow 14$, $3 \leftarrow 4$, $15 \leftarrow 13$, $3 \leftarrow 2$, $5 \leftarrow 6$	$7 \leftarrow 1$, $13 \leftarrow 7$, $8 \leftarrow 2$, $9 \leftarrow 8$, $9 \leftarrow 3$, $14 \leftarrow 9$, $12 \leftarrow 6$, $16 \leftarrow 12$, $10 \leftarrow 4$,$15 \leftarrow 10$, $11 \leftarrow 5$, $15 \leftarrow 14$, $16 \leftarrow 13$, $4 \leftarrow 3$, $12 \leftarrow 7$, $10 \leftarrow 9$, $14 \leftarrow 13$, $6 \leftarrow 1$, $3 \leftarrow 2$, $6 \leftarrow 5$, $2 \leftarrow 1$, $5 \leftarrow 4$, $13 \leftarrow 8$

2. Hugin Lite (demo version of Hugin Expert) [1]: We used both of Hugin Expert's structure learning algorithms on our binary class data: PC and NPC (Necessary Path Condition). The demo is limited to 50 states and 500 cases for learning, hence we could not attempt to use the 5-class data. The Hugin PC algorithm is a variant of the original PC algorithm due to Sprites, Glymour & Scheines (2000) [21], and is a constraint-based learning algorithm (cf. Hugin manual for details). The basic idea of such like algorithms is to derive a set of conditional independence and dependence statements (CIDs) by statistical tests. The NPC algorithm improves upon the PC algorithm as it can resolve inconsistencies within the derived set of CIDs. The implementation of the NPC algorithm in Hugin Lite also allows for user interaction to decide the directionality of those links that could not be determined based on the data only. We thus used the NPC algorithm for most of our experiments. The PC algorithm was applied mostly to check what would be learnt without human interaction.

Multiple rounds of structure learning were performed.

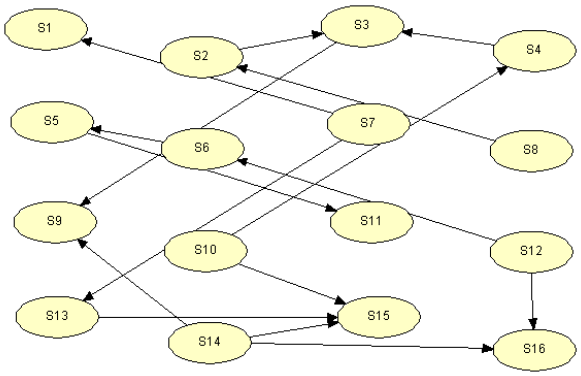

Figure 8: Final resultant network learnt using Hugin Lite's NPC algorithm and all 16 heart segments.

Here we describe the final structure that will be tested in the next section. We wanted to find the most common structure for all the training data (Erasmus + another hospital). This was done with all 16 segments using the NPC algorithm. 1 network was created using the Erasmus training cases, 1 network with 169 training cases from the other hospital, and 5 networks from the 2216 Erasmus non-image cases (each network created from a random sample of 500 cases). An edge consensus of all networks was done to create the final resultant network seen in Figure 8. When one overlays the structural relationships learnt (i.e. arcs in the BN) on the circular diagram of the heart segments as seen in Figure 9 it is easier to see how the edges relate to the physical relationships within the heart. Comparing these results with the Chow and Liu algorithm results, most of the relationships that made sense were covered, except 8 – 9 and 13 – 16. And all the edges found between the segments match up with the actual physical relationships within the heart. Results can be seen in Table 1.

3. BNT algorithm [2]: We also applied the open-source code created by Kevin Murphy to learn the most common structure for all the training data with all 16 segments. For this we used the multi-class labels (1 - 5) instead of the binary labels because of the illogical results given by the program when provided the binary data and many different topological orderings of the segments (a requirement of a program). When using the multi-class labels the program always gave the same structure regardless of the topological ordering of the segments. This structure can be seen in Figure 10 Overlaying the structural relationships learnt on the circular diagram of the heart segments (see Figure 11) shows the logical sense of them with respect to the physical heart. When compared with the results of the Chow and Liu algorithm all of the links, save 14 – 16, that made logical sense were covered. And comparing the results to the 16 segment NPC algorithm results gives the following result (Hugin Lite learnt 16 edges, while BNT learnt 23 total edges): 7 edges kept the same, 7 edges had their directionality reversed, 9 new edges learnt by BNT, 2 edges not learnt by BNT. Results can be seen in Table 1. The two edges which were not learnt by BNT do not tech-

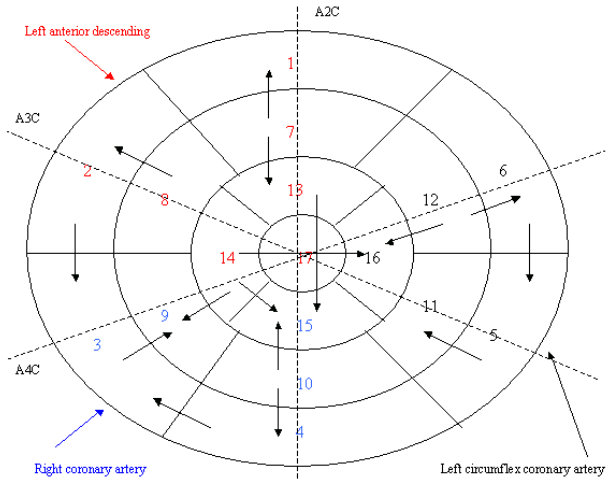

Figure 9: Structural relationships learnt using Hugin Lite's NPC algorithm and all 16 heart segments, overlayed on physical heart diagram.

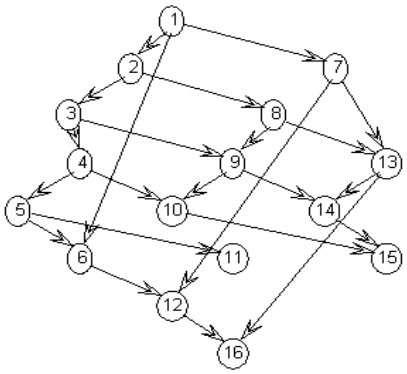

Figure 10: Structure learnt by BNT using multi-class labels.

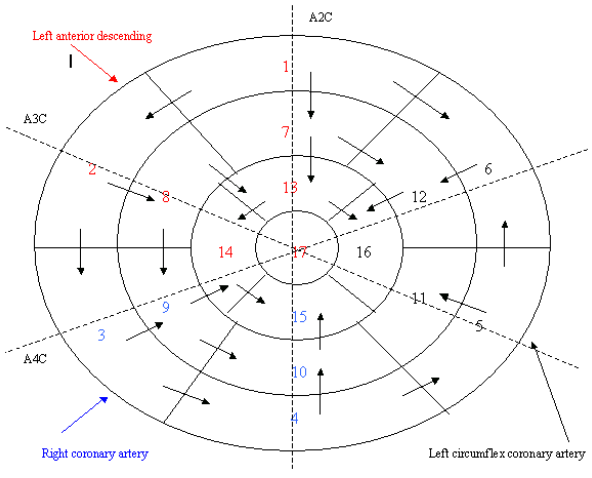

Figure 11: Structural relationships learnt by BNT using multi-class labels overlayed on physical heart diagram

nically satisfy the requirements of having the segments involved be neighbors and/or share blood territory, but in each case the segments are 'neighbors' in different standard views (and since we are not considering the 17th segment - the apex - they would be neighbors were that segment to be removed).

6. TESTING LEARNT BN MODELS

To test the BNs we learnt, we transferred the networks learnt to Netica [3] and did parameter training using the 220 Erasmus training cases with images, and 2216 Erasmus cases without images (i.e. a subset of the same cases used to do structure learning). The test set was 125 cases held out specifically for this purpose. The testing process involved entering all available evidence into the compiled network, running inference, and querying all variables of interest simultaneously. When evidence in entered into a compiled network, Netica automatically incorporates the new information into the network (via junction tree propagation). In our case, the evidence given to the network were the 6 best segment specific features found by the Kolmogorov-Smirnov test (all continuous attributes), and the variables of interest that were queried were the nodes that represented each heart segment (S1, S2, ... , S16; all discrete attributes). After the structure between the segments was learnt off of data via different methods described above, we manually added the evidence variables using a naive Bayesian methodology (each node representing a segment specific feature was influenced only by the corresponding segment's node in the network, e.g. S2 → s2_A3C_segvol_ef). Experiments done to learn a better structure between the segment nodes and the new evidence nodes showed that this simple approach was indeed the most logical and gave the best results.

All continuous valued features were discretized using Recursive Minimal Entropy Partitioning [15], a method based on minimal entropy heuristic, presented in [8, 16]. This is a supervised algorithm that uses class information entropy of candidate partitions to select bin boundaries for discretization.

We tested 2 networks (the last two mentioned in the learning section) using the subset of selected features described above:

1. The 16 segment network learnt by Hugin Lite using the NPC algorithm and binary segment GT
2. The 16 segment network learnt by BNT using the 5-class GT labels

6.1 Numerical Experiments

We have obtained many different answers from doctors about the trade off between false positives (FP, i.e. wrongly labeling the heart abnormal) verses false negatives (FN, i.e. wrongly labeling the heart normal); this depends on the clinical context. If this system is used as an initial reader, then having too many FPs or FNs will cause the doctors to shut off the system because it will be deemed unreliable. But as we plan to use our system as a "second reader" to validate a physician read, the main focus should be to keep the FN rate low. In general, if you have a high FP rate then you are sending too many patients for additional, more expensive tests, which would lead to higher costs for health insurance. A high FN rate could mean that a patient might go undiagnosed if the doctor using the system is not well trained and also misses potential abnormalities. In our context, the "cost" of a FN is thus higher than a FP. By focusing on keeping the FN rate low, we lower the risk of missing

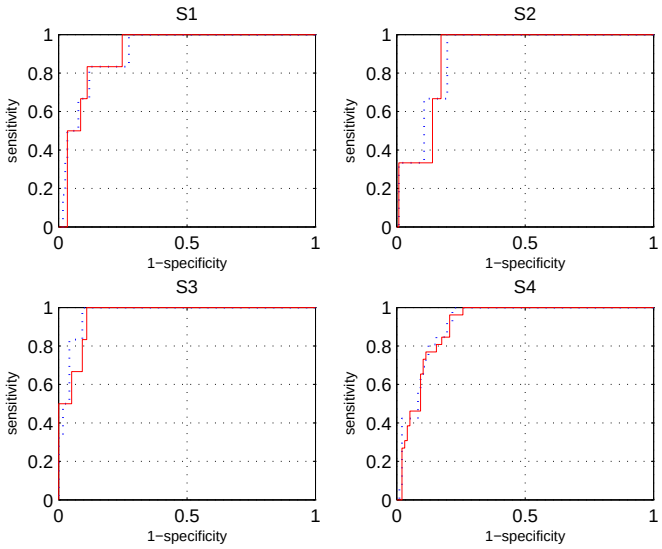

Figure 12: Test ROC curves for segments 1 through 4: BNT (dotted line) and HUGIN (solid line).

abnormalities and leave the final diagnosis to the expertise of the doctor. However, to take all variations into account, we decided that the most objective way to evaluate the classifier performance is to measure the area under the ROC curve (AUC).

Both our classifiers were trained using 220 cases, and were tested on 125 cases. Their results can be compared per segment against one another as shown by the ROC curves in Figures 12, 13, 14, and 15. Our feature selection resulted in each segment being dependent on six features (both global and local). The same subset of features were used for both classifiers. For each algorithm, Table 2 shows the AUCs for the testing set. As the results show, both classifiers did equally well on each segment, and on the whole achieved high sensitivity, greater than or equal to 80%, for 80% specificity on most segments.

Table 2: AUCs for test set

Segment	16 seg HUGIN NPC	16 seg BNT
S1	0.90883	0.90883
S2	0.89344	0.89617
S3	0.95798	0.96779
S4	0.9096	0.91673
S5	0.83913	0.84806
S6	0.99593	0.98374
S7	0.87192	0.86453
S8	0.82301	0.8208
S9	0.97436	0.96474
S10	0.92561	0.92763
S11	0.83764	0.84508
S12	0.97967	0.98374
S13	0.97434	0.97168
S14	0.92056	0.92552
S15	0.94682	0.94818
S16	0.94	0.945

7. FUTURE WORK

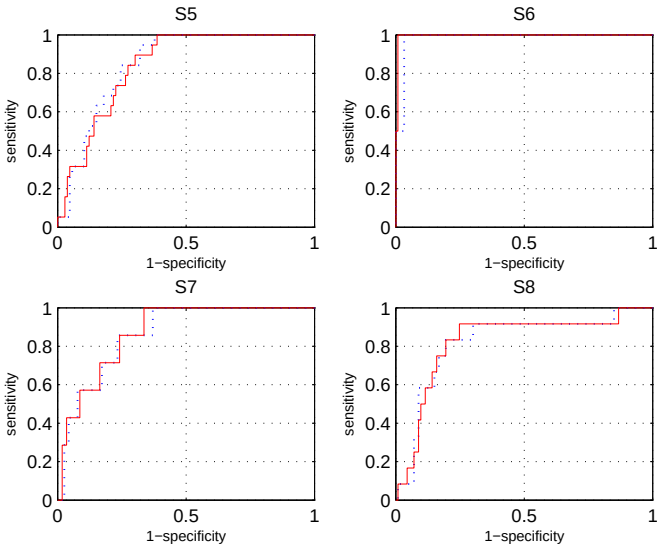

Figure 13: Test ROC curves for segments 5 through 8: BNT (dotted line) and HUGIN (solid line).

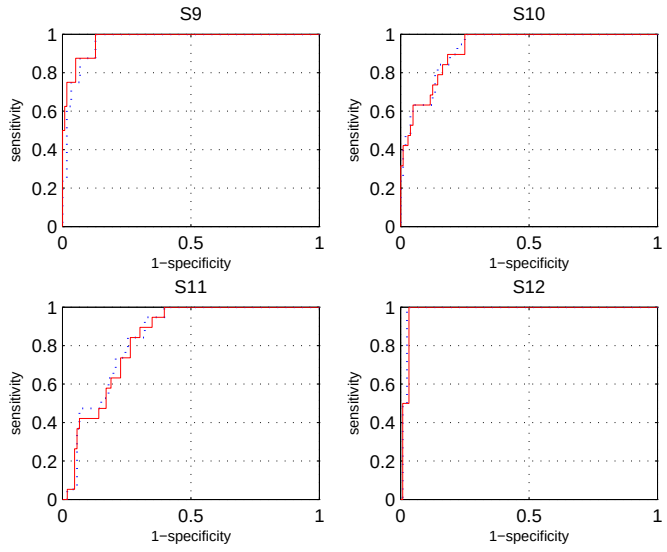

Figure 14: Test ROC curves for segments 9 through 12: BNT (dotted line) and HUGIN (solid line).

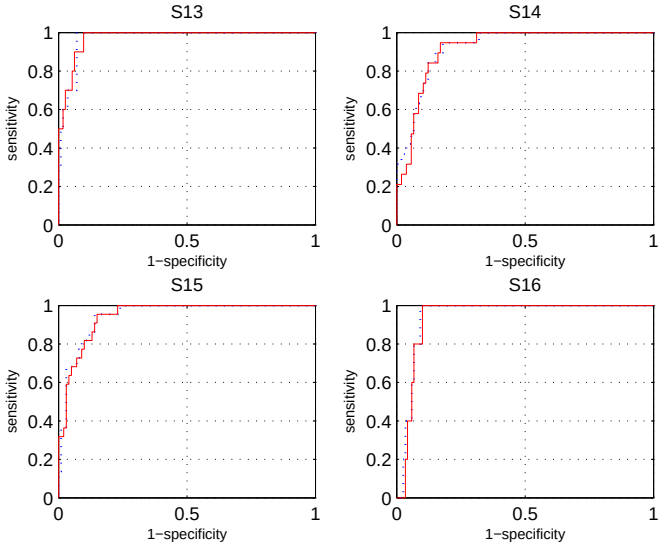

Figure 15: Test ROC curves for segments 13 through 16: BNT (dotted line) and HUGIN (solid line).

In the future we plan on expanding our classification to identify different levels of CHD severity at the segment level (Levels 1-5: 1 = normal, 2 = hypo-kinetic, 3 = a-kinetic, 4 = dys-kinetic, 5 = aneurysm), incorporating the use of other standard echocardiography views (for example: parasternal short axis (PSAX), parasternal long axis(PLAX)), and including images from other levels of stress (e.g.: peak dose). We would also like to apply a ranking algorithm to take advantage of multi-class scores for classification.

8. CONCLUSION

Cardiovascular Disease (CVD) is the single largest contributor to "Disability Adjusted Life Years" (DALYs), which can be thought of as "healthy years of life lost". This measure indicates the total burden of a disease verses the deaths that result from it. CVD accounts for 10% of DALY's in low and middle income nations, and 18% of DALY's in high-income nations. This burden is projected to rise from 47 million DALYs globally in 1990 to 82 million DALYs in 2020 [22]. Over 60% of the global burden of this disease occurs in developing countries. Hence the WHO and CDC agree that "CVD is no longer an epidemic. It is a pandemic". In the United States, CVD accounted for 37.3% of all deaths in 2003 (or 1 of every 2.7 deaths) and was the primary or contributing cause in about 58% of all deaths that year [5]. Ultrasound is the best diagnostic tool to detect this disease and is used routinely for this purpose.

In this paper we address the task of building an objective classification application for heart wall motion analysis, based on features calculated off of echocardiograms. The simple but effective feature selection technique that was used results in classifiers depending on only a small subset of the calculated features, and their limited number makes it easier to explain the final classifier to physicians in order to get their feedback. Although we only had a relatively small number of cases with images and ground truth, we were able to leverage a large number of additional cases - namely data without images, but with ground truth only - to improve performance. It was interesting to note that both Bayesian Network classifiers

tested performed equally well on the task and gave greater or equal to 80% sensitivity for the same amount of specificity.

9. ADDITIONAL AUTHORS

Additional authors: Romer Rosales (Siemens Medical Solutions, email: romer.rosales@siemens.com), Harald Steck (Siemens Medical Solutions, email: harald.steck@siemens.com), R. Bharat Rao (Siemens Medical Solutions, email: bharat.rao@siemens.com), and Dr. Don Poldermans (Erasmus University Medical Center, Rotterdam, email: d.poldermans@erasmusmc.nl).

10. REFERENCES

[1] Web site for Hugin Lite: http://www.hugin.com/.

[2] Web site for Bayes Net Toolbox for Matlab: http://bnt.sourceforge.net (by Kevin Murphy).

[3] Web site for Netica: http://www.norsys.com/.

[4] *Probabilistic Networks and Expert Systems*. Springer, 1999.

[5] American Heart Association. Heart disease and stroke statistics – 2006 update. 2006. URL:http://www.americanheart.org/presenter.jhtml?identifier=1200026.

[6] D. W. Biesiada J. Feature selection for high-dimensional data: A kolmogorov-smirnov correlation-based filter solution. *CORES 2005 4th International Conference on Computer Recognition Systems.*, pages 95 – 105, 22-25 May 2005. Rydzyna. Advances in Soft Computing, Computer Recognition Systems.

[7] M. Catherine M. Otto. *Textbook of Clinical Echocardiography, 2nd edition*. W.B. Saunders Company, Philadelphia, PA, 2000. ISBN 0-7216-7669-3.

[8] J. Catlett. On changing continuous attributes into ordered discrete attributes. *Proceedings of the European Working Session on Learning, Berlin, Germany*, pages 164 – 178, 1991.

[9] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, , and M. S. Verani. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *American Heart Association Circulation*, 105:539 – 542, Jan 2002. http://circ.ahajournals.org/cgi/content/full/105/4/539; accessed Dec 8, 2004.

[10] R. L. I. Chakravarti and J. Roy. *Handbook of Methods of Applied Statistics*. JohnWiley and Sons, Chichester, 1967.

[11] D. M. Chickering. Learning Bayesian networks is NP-Complete. In D. Fisher and H. Lenz, editors, *Learning from Data: Artificial Intelligence and Statistics V*, pages 121–130. Springer-Verlag, 1996.

[12] C. K. Chow and C. N. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, IT-14, NO. 3:462 – 467, May 1968.

[13] D. Comaniciu. Nonparametric information fusion for motion estimation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, I:59 – 66, 2003.

[14] D. Comaniciu, X. S. Zhou, and S. Krishnan. Robust real-time tracking of myocardial border: An information fusion approach. *IEEE Trans. Medical Imaging*, 23, NO. 7:849 – 860, 2004.

[15] J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. In *International Conference on Machine Learning*, pages 194–202, 1995.

[16] U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 1022 – 1027, 1993.

[17] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, 2005*.

[18] D. Heckerman. A tutorial on learning with bayesian networks. Technical report, 1995.

[19] R. Hoffmann, T. Marwick, D. Poldermans, H. Lethen, R. Ciani, P. van der Meer, H.-P. Tries, P. Gianfagna, P. Fioretti, J. Bax, M. Katz, R. Erbel, and P. Hanrath. Refinements in stress echocardiographic techniques improve inter-institutional agreement in interpretation of dobutamine stress echocardiograms. *European Heart Journal*, 23, issue 10:821 – 829, May 2002. doi:10.1053/euhj.2001.2968, available online at http://www.idealibrary.com.

[20] S. Moir and T. H. Marwick. Combination of contrast with stress echocardiography: A practical guide to methods and interpretation. *Cardiovascular Ultrasound*, 2:15, 2004. URL: http://cardiovascularultrasound.com/content/2/1/15; accessed Nov 5, 2004. doi:10.1186/1476-7120-2-15. Authors from the Division of Medicine, University of Queensland, Brisbane, Australia.

[21] C. G. P. Spirtes and R. Scheines. *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning. MIT Press, second edition, 2000.

[22] World Health Organization. The atlas of heart disease and stroke. 2004. http://www.who.int/cardiovascular_diseases/resources/atlas/en/index.html.

[23] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *NIPS*, pages 689–95, 2000.

[24] X. S. Zhou, D. Comaniciu, and A. Gupta. An information fusion framework for robust shape tracking. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 27, NO. 1:115 – 129, January 2005.