Sparse classifiers for Automated Heart Wall Motion Abnormality Detection

Glenn Fung, Maleeha Qazi, Sriram Krishnan, Jinbo Bi, Bharat Rao
Siemens Medical Solutions
51 Valley Stream Parkway
Malvern, PA, USA
glenn.fung@siemens.com

Alan Katz
St. Francis Hospital
100 Port Washington Blvd
Roslyn, New York, USA

Abstract

Coronary Heart Disease is the single leading cause of death world-wide, with lack of early diagnosis being a key contributory factor. This disease can be diagnosed by measuring and scoring regional motion of the heart wall in echocardiography images of the left ventricle (LV) of the heart. 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. We describe a novel feature selection technique based on mathematical programming that results in a robust hyperplane-based classifier. The classifier depends only on a small subset of numerical feature extracted from dual-contours tracked through time. We verify the robustness of our system 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.

1. Introduction

Cardiovascular Disease (CVD) is a global epidemic that is the leading cause of death worldwide (17 mil. deaths) [16]. In the United States, CVD accounted for 38% of all deaths in 2002 [1] and was the primary or contributing cause in 60% of deaths. *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. It is well-known that early detection (along with prevention) is an excellent way of controlling CHD. CHD 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 indeed 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.

Unfortunately, echocardiograms are notoriously difficult to interpret, even for the best of physicians. Inter-observer studies have shown that even world-class experts agree on their diagnosis only 80% of the time [10], and intra-observer studies have shown a similar variation when the expert reads the same case twice at widely different points in time. There is a tremendous need for an automated "second-reader" system that can provide objective diagnostic assistance, particularly to the less-experienced cardiologist.

In this paper, we address the task of building a computer-aided diagnosis system that can automatically detect wall-motion abnormalities from echocardiograms.

The following section provides some medical background on cardiac ultrasound and the standard methodology used by cardiologists to score wall-motion abnormalities. We also describe our real-life dataset, which consists of echocardiograms used by cardiologists at St. Francis Heart Hospital to diagnose wall-motion abnormalities. The next section provides an overview of our proposed system which we built on top of an algorithm that detects and tracks the inner and outer cardiac walls [9, 17, 5, 6]. 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. Then we describe our methodology for feature selection and classification, followed by our experimental results. We conclude with some thoughts about our plans for future research.

2. Medical Background Knowledge

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 are using echocardiography. The Cardiac Imaging Committee of the Council on Clini-

cal 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 [4]. This is the standardization used in this project.

Accurate regional wall motion analysis of the left ventricle is an essential component of interpreting echocardiograms (echos). The left ventricle (LV) is divided into 17 myocardial segments as shown in Figure 1 (modified from reference [4]), which are fed by 3 coronary arteries: the left anterior descending(LAD) (feeds segments 1, 2, 7, 8, 13, 14, 17), right coronary artery (RCA) (feeds segments 3, 4, 9, 10, 15), and the left circumflex coronary artery(LCX) (feeds segments 5, 6, 11, 12, 16).

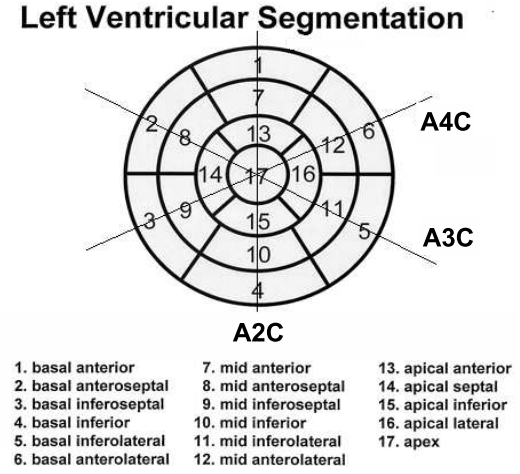

Figure 1. Display, on a circumferential polar plot, of the 17 myocardial segments and the recommended nomenclature for tomographic imaging of the heart. Modified from reference [4].

The echocardiograms are run through an algorithm which automatically detects and tracks both the endocardial and epicardial borders of the LV [5, 6]. Motion interferences (e.g. probe motion, patient movement, respiration, etc.) are compensated for using global motion estimation based on robust statistics outside the LV. Numerical feature vectors extracted from the dual-contours tracked through time form the basis for regional wall motion classification.

3. Data

The data is based on standard adult transthoracic B-mode ultrasound images collected from the four standard views: apical 4 chamber (A4C), apical 2 chamber (A2C), parasternal long axis (PLAX) or apical 3 chamber (A3C),

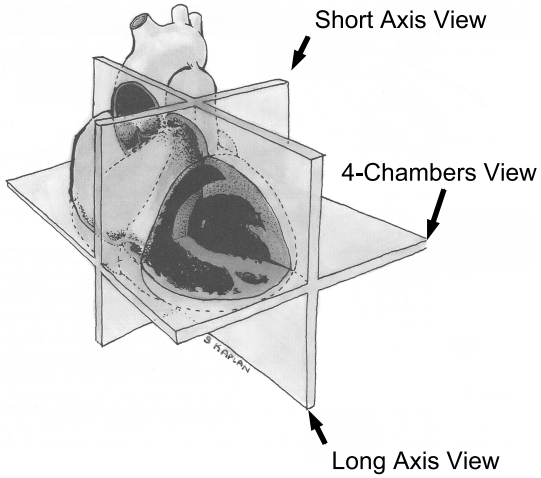

Figure 2. The three basic image planes used in transthoracic echocardiography. The ventricles have been cut away to show how these image planes intersect the left and right ventricles. *Dashed lines* **indicate the image planes at the great vessel and atrial levels. From reference [3]**

and parasternal short axis (PSAX) – shown in Figure 2 from reference [3]. Currently we are only utilizing two of the four possible views - A4C and A2C. These two views show 12 of the 16 total segments, but that is enough 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 of this work is to automatically provide an initial score or classification to determine whether a heart is normal or abnormal given the ultrasound.

The ultrasound data was collected from St. Francis Heart Hospital, Roslyn, NY, USA (abbrev: SF). The SF data consists of 141 cases that will be used for training, and 59 cases that are ear-marked as the final test set. All the cases have been labeled at the segment level by a group of trained cardiologists. The heart level classification labels can be obtained from the segment level labels by applying the following definition: A heart is considered abnormal if two or more segments are abnormal.

4. Methodology

The classification algorithm used in the system is based on a novel feature selection technique, which is in turn based on mathematical programming. As a result we obtain a hyperplane-based classifier that only depends on a sub-

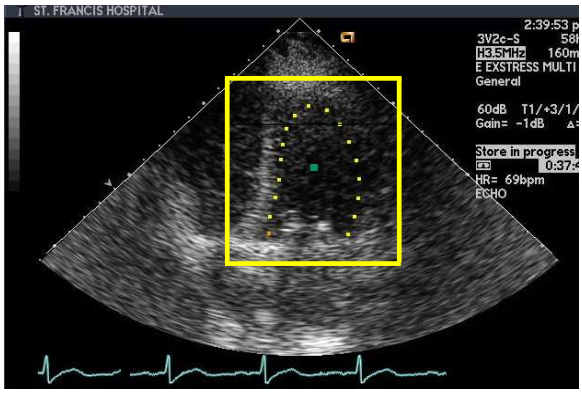

Figure 3. One frame from an A4C image clip with the yellow box showing the localized LV, and the yellow dots representing the control points along the detected contour.

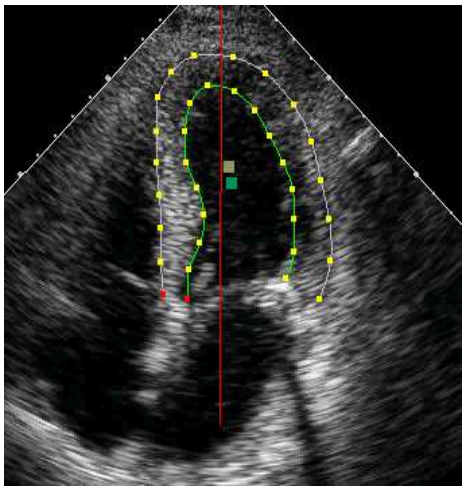

Figure 4. One frame from an A4C image clip with the outer and inner contour control points shown. The red vertical line shows use of global motion compensation, and the two squares denote the centers of the individual contours.

set of numerical features extracted from the dual-contours tracked through time, and these are then used to provide classification for each segment and the entire heart.

4.1. Image processing

The first step toward classification of the heart involves automatic contour generation of the LV [9]. 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 the algorithm we use 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: border detection and border tracking. Border detection involves localizing the LV on multiple frames of the image clip (shown in Figure 3 as a box drawn around the LV), and then detecting the LV's shape within that box. Border tracking involves tracking this LV border from one frame to the next through the entire movie clip. 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 4 as a vertical red line near the center of the image. 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: global and local ejection fraction (EF) ratio, radial displacement, circumferential strain, velocity, thickness, thickening, timing, eigenmotion, curvature, and bending energy. Some of these features, including timing, eigenmotion, curvature, local EF ratio and bending energy, are based on the 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 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:

- Timing-based features: examine the synchronousness

of the cardiac motion, i.e. whether all the points along the LV move consistently or not.

- Eigenmotion-based features: determine the most significant moving direction of a point and the amount of it's motion in that direction.

- Curvature-based features: Are mainly aimed at detecting abnormalities at the apex. It is also useful in identifying more general abnormalities associated with cardiac shapes. If a segment is dead, it may still move because it is connected to other segments, but we can observe that it's shape will largely remain unchanged during the cardiac cycles. Curvature can capture this type of information.

- Local EF ratio features: Are aimed at capturing local cardiac contraction abnormalities.

- Bending energy features: Assuming that the provided contour is made of elastic material and moving under tension, then the bending energy associated with the contour may be used to capture the cardiac contraction strength of a segment or the whole LV.

- Circumferential strain features: also called Fractional Shortening, measures how much the contour between any two control points shrinks in the systolic phase.

In general, the global version of certain features (e.g. radial displacement, radial velocity, 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 192 local and global features, all of which were continuous.

4.3. Classification and 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 improves the classifier 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. Recently, Mika et al, proposed a novel mathematical programming formulation for Fisher's Linear Discriminant using kernels [14, 13], this new formulations included a regularization term similar to the used in the standard SVM formulation [12]. We will make use of Mika's formulation but by using the 1-norm instead of the 2-norm we will obtained solutions that are moire sparse and hence depend on a smaller number of features. The next section describe the details of the approach.

4.4. Fisher's Linear Discriminant

Let $A_i \in R^{d\times l}$ be a matrix containing the l training data points on d-dimensional space and l_i the number of labeled samples for class w_i, $i \in \{\pm\}$. FLD [7]is the projection α, which maximizes,

$$J(\alpha) = \frac{\alpha^T S_B \alpha}{\alpha^T S_W \alpha} \tag{1}$$

where

$$S_B = (m_+ - m_-)(m_+ - m_-)^T$$

$$S_W = \sum_{i\in\{\pm\}} \frac{1}{l_i}\left(A_i - m_i e_{l_i}^T\right)\left(A_i - m_i e_{l_i}^T\right)^T$$

are the between and within class scatter matrices respectively and $m_i = \frac{1}{l_i} A_i e_{l_i}$ is the mean of class w_i and e_{l_i} is an l_i dimensional vector of ones. Transforming the above problem into a convex quadratic programming problem provides us some algorithmic advantages. First notice that if α is a solution to (1), then so is any scalar multiple of it. Therefore to avoid multiplicity of solutions, we impose the constraint $\alpha^T S_B \alpha = b^2$, which is equivalent to $\alpha^T (m_+ - m_-) = b$ where b is some arbitrary positive scalar. Then the optimization problem (1) becomes,

$$\begin{array}{ll} s.t. & \min_{\alpha\in R^d} \quad \alpha^T S_W \alpha \\ & \alpha^T (m_+ - m_-) = b \end{array} \tag{2}$$

For binary classification problems the solution of this problem is

$$\alpha^* = \frac{b S_W^{-1}(m_+ - m_-)}{(m_+ - m_-)^T S_W^{-1}(m_+ - m_-)} \tag{3}$$

According to this expansion since S_W^{-1} is positive definite unless the difference of the class means along a given feature is zero all features contributes to the final discriminant.

If a given feature in the training set is redundant, its contribution to the final discriminant would be artificial and not desirable. As a linear classifier FLD is well-suited to handle features of this sort provided that they do not dominate the feature set, that is, the ratio of redundant to relevant features is not significant. Although the contribution of a single redundant feature to the final discriminant would be negligible when several of these features are available at the same time, the overall impact could be quite significant leading to poor prediction accuracy. Apart from this impact, in the context of FLD these undesirable features also pose numerical constraints on the computation of S_W^{-1} especially when the number of training samples is limited. Indeed, when the

number of features, d is higher than the number of training samples, l, S_W becomes ill-conditioned and its inverse does not exist. Hence eliminating the irrelevant and redundant features may provide a two-fold boost on the performance. In what follows we propose a sparse formulation of FLD. The proposed approach incorporates a regularization constraint on the conventional algorithm and seeks to eliminate those features with limited impact on the objective function.

4.5. Sparse Fisher's Linear Discriminant via linear programming

We propose a formulation similar to the one used for 1-norm SVM classifiers [2] where the 1-norm is introduced for both measuring the classification error and regulation. The use of the 1-norm instead of the 2-norm leads to linear programming formulations where very sparse solutions can be obtained. Our objective is to formulate an algorithm that can be seen as an approximation to (1) and that provides a sparse projection vector α. In order achieve this we add a regularization term to the objective function of (2):

$$\begin{array}{ll} \min_{\alpha \in R^d} & \nu \alpha^T S_W \alpha + \|\alpha\|_1 \\ s.t. & \\ & \alpha^T (m_+ - m_-) = b \end{array} \tag{4}$$

Where ν is the trade-off between $J(\alpha)$ maximization and regularization or sparsity of the projection vector α. The price to pay for sparsity of the solution is that unlike (2), there is no a closed form solution for the constrained quadratic in (4), furthermore the parameter ν introduced in (4) has to be chosen by means of a tuning set which requires the problem to be solved several times and that can be computationally demanding. In order to address this issue we propose next, a linear programming formulation that can be interpreted as an approximation to (4) and that results in sparser solutions than (4). Lets consider the following matrix:

$$H^T = \left[\frac{1}{\sqrt{l_+}} \left(A_+ - m_+ e_{l_+}^T \right)^T \quad \frac{1}{\sqrt{l_-}} \left(A_- - m_- e_{l_-}^T \right) \right]$$

From (1) we have that $S_w = H^T H$, then:

$$\begin{array}{rcl} \alpha^T S_W \alpha & = & \alpha^T H^T H \alpha \\ & = & (H\alpha)^T (H\alpha) \\ & = & \|H\alpha\|_2^2 \end{array} \tag{5}$$

Hence, quadratic programming problem (4) can be rewritten as:

$$\begin{array}{ll} \min_{\alpha \in R^d} & \nu \|H\alpha\|_2^2 + \|\alpha\|_1 \\ s.t. & \\ & \alpha^T (m_+ - m_-) = b \end{array} \tag{6}$$

We can now use the 1-norm instead of the 2-norm in the objective function of (6) to obtain the following linear programming formulation that can be solved more efficiently and gives sparser solutions:

$$\begin{array}{ll} \min_{\alpha \in R^d} & \nu \|H\alpha\|_1 + \|\alpha\|_1 \\ s.t. & \\ & \alpha^T (m_+ - m_-) = b \end{array} \tag{7}$$

That this problem is indeed a linear program, can be easily seen from the equivalent formulation:

$$\begin{array}{ll} \min_{\alpha \in R^d} & \nu e's + e't \\ s.t. & \\ & \alpha^T (m_+ - m_-) = b \\ & -s \leq H\alpha \leq s \\ & -t \leq \alpha \leq t \end{array} \tag{8}$$

Next, we propose an algorithm based on formulation (8) and equation (3) that provides accurate FLD classifiers depending on a minimal set of features.

Algorithm 1 *Sparse Linear Fisher Discriminant*
Given the training dataset $\{A_-, A_+\}$ *and a set of values* $N = \{10^{-5}, 10^{-4}, \ldots, 10^5\}$ *for the parameter* ν *do:*

1. *For each* $\nu \in N$ *calculate cross-validation performance using the linear programming formulation (8).*
2. *Let* ν^* *the value for which formulation (8) gives the best cross-validation performance. Let's call* $\hat{\alpha}$ *the obtained sparse projection.*
3. *Select the subset* $\hat{F}$ *of the feature set* F *defined by* $f_i \in \hat{F} \Leftrightarrow \hat{\alpha}_i \neq 0$*, this is, select the features corresponding to nonzero components of the projection* $\hat{\alpha}$*.*
4. *Solve original quadratic programming problem (1) with close form solution (3) considering only the feature subset* $\hat{F}$ *to obtain a final projection* α^* *that depends on only the "small" feature subset* $\hat{F}$*.*

5. Numerical Experiments

In order to empirically demonstrate the effectiveness of the proposed approach, we compared our feature selection algorithm, Sparse LFD (**SLFD**) to three other well-known classification algorithms: The first algorithm is a very popular publicly available implementation of SVM, $\mathbf{SVM}_{light}$ [11]. This formulation does not incorporate feature selection and produces classifiers that often depend on all the input features. The purpose of the comparison is to show that a feature selection method improves generalization performance on this dataset. The second method included in our numerical comparisons is the Automatic Relevance Determination (**ARD**) algorithm [15] which is one of the

Table 1. Results including *AUC* Area under the ROC curve for the testing set and number of features selected for the four methods: SLFD, SVM$_{light}$,ARD and LFD. Best results in bold

Algorithm	*AUC*	# of features
SLFD	**89.6 %**	**3**
SVM$_{light}$	87.4 %	79*
ARD	85.8%	13
LFD	87.4%	79*

* classifier uses all the features.

most successful Bayesian methods for feature selection and sparse learning. It finds the relevance of features by optimizing the model marginal likelihood, also known as the evidence. The third approach consists of applying the standard **LFD** algorithm [7] without feature selection. All the classifiers were trained using 141 cases, and were tested on 59 cases. For the methods that needed parameters to be tuned, i.e. our algorithm and SVM$_{light}$, the model parameters were tuned by the means of leave-one-patient-out (LOPO) [8] cross validation on the training set.

We have gotten many different answers from doctors as to what they feel the cost of a false positive (FP, i.e. wrongly labeling the heart abnormal) or false negative (FN, i.e. wrongly labeling the heart normal) happens to be. If this system is used as an initial reader then too many FPs or FNs will cause the doctors to shut off the system because it is too unreliable. But as a validation system the main focus is 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. For us, 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 abnormalities and leave the final diagnosis to the expertise of the doctor. Taking this into account, we decided that the best way to evaluate the classifier performance is to measure the area under the ROC curve (AUC).

For each algorithm, Table (1) shows the AUC for the testing set and the number of features that the corresponding classifier depends on. As it can be seen from the results, our method obtained the ROC with the largest area and only depended on three features. This very low feature dependence is very important in our application since the features used for classification have to be calculated in real time.

6. Final Results

The three features selected by **SLFD** were:

1. A feature that measures the motion along the significant directions of movement of the wall of hearts
2. A feature that measures the correlation between the estimated area of the heart cavity and the distance between the walls of the heart to the center of mass of the heart.
3. The estimated ejection fraction of the heart.

The Receiver Operating Characteristic (ROC) curve on the testing set for the final classifier is shown in figure 5. The area under the ROC curve for the testing set was 0.896. The LOPO cross-validation performance for the final model was 7 false positives and 17 false negatives out of 81 positives (abnormals) and 60 negatives (normals), i.e., 88.3% of the normal hearts and 79.0% of the abnormal hearts were correctly classified. On the testing set we got 3 false positives and 6 false negatives out of 39 positives (abnormals) and 20 negatives (normals), i.e., 85.0% of the normal hearts and 84.6% of the abnormal hearts were correctly classified. A 3D plot depicting the final classifier and the testing set is shown in Figure 6. The clinical results were presented and published at the American College of Cardiology meeting in March 2005 under the title:"Clinical Evaluation of a Novel Automatic Real-Time Myocardial Tracking and Wall Motion Scoring Algorithm for Echocardiography Introduction".

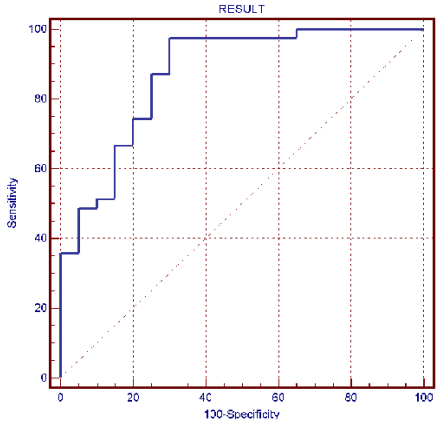

Figure 5. ROC curve for the testing set

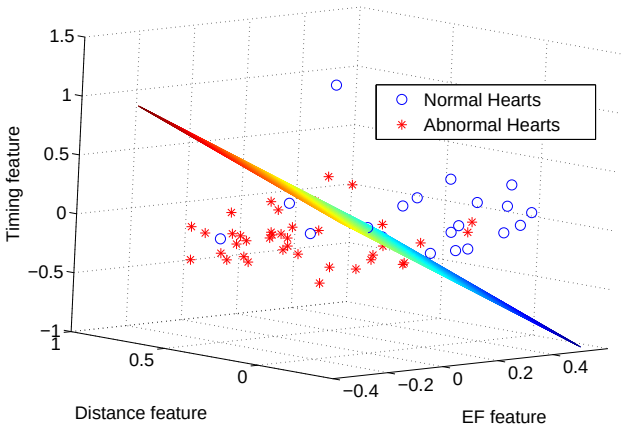

Figure 6. Final hyperplane classifier in three dimensions, circles represent normal hearts and stars represent abnormal hearts in the testing set

7. Future work

In the future we plan on expanding our classification to identify different levels of CHD severity (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: apical 3 chamber (A3C), parasternal short axis (PSAX), parasternal long axis(PLAX)), and including images from other levels of stress. Comparisons of our proposed SLFD algorithm on other publicly available datasets and medical applications are also planned to further explore the potential of the algorithm.

References

[1] American Heart Association. Heart disease and stroke statistics 2005 update. 2005. URL:http://www.americanheart.org/downloadable/heart.

[2] P. S. Bradley and O. L. Mangasarian. Feature selection via concave minimization and support vector machines. In J. Shavlik, editor, *Machine Learning Proceedings of the Fifteenth International Conference(ICML '98)*, pages 82–90, San Francisco, California, 1998. Morgan Kaufmann.

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

[4] M. Cerqueira, N. Weissman, V. Dilsizian, A. Jacobs, S. Kaul, W. Laskey, D. Pennell, J. Rumberger, T. Ryan, and M. Verani. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *American Heart Association Circulation*, 105:539 – 542, Jan 2002. URL: http://circ.ahajournals.org/cgi/content/full/105/4/539.

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

[6] 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.

[7] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2001.

[8] M. Dundar, G. Fung, L. Bogoni, M. Macari, A. Megibow, and B. Rao. A methodology for training and validating a cad system and potential pitfalls. In *CARS 2004, Computer Assisted Radiology and Surgery, Proceedings*, Chicago, USA, 2004. Elsevier.

[9] 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*.

[10] 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.

[11] T. Joachims. SVM^{light}, 2002. http://svmlight.joachims.org.

[12] O. L. Mangasarian. Generalized support vector machines. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 135–146, Cambridge, MA, 2000. MIT Press. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-14.ps.

[13] S. Mika, G. Rätsch, and K.-R. Müller. A mathematical programming approach to the kernel fisher algorithm. In *NIPS*, pages 591–597, 2000.

[14] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller. Fisher discriminant analysis with kernels. In Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, editors, *Neural Networks for Signal Processing IX*, pages 41–48. IEEE, 1999.

[15] Y. Qi, T. P. Minka, R. W. Picard, and Z. Ghahramani. Predictive automatic relevance determination by expectation propagation. In *Proceedings of Twenty-first International Conference on Machine Learning*, 2004.

[16] World Health Organization. The atlas of global heart disease and stroke. 2004. URL:http://www.who.int/ cardiovascular_diseases/ resources/atlas/.

[17] 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.