2

Conf-9410156--1

UCRL-JC-117704

CLASSIFICATION OF HEART VALVE SINGLE LEG SEPARATIONS FROM ACOUSTIC CLINICAL MEASUREMENTS

Gregory A. Clark
Brenda C. Bowman
Nicholas Boruta
Graham H. Thomas
Holger E. Jones
Michael R. Buhl

This paper was prepared for submittal to the
28th Annual Asilomar Conference on
Signals, Systems, and Computers
Pacific Grove, CA
October 31 - November 2, 1994

May 1994

Lawrence Livermore National Laboratory

DISCLAIMER

Twenty-Eighth Annual Asilomar Conference on Signals, Systems and Computers,
October 31-November 2, 1994, Pacific Grove, CA

Classification of Heart Valve Single Leg Separations from Acoustic Clinical Measurements

Gregory A. Clark, Brenda C. Bowman, Nicholas Boruta,
Graham H. Thomas, Holger E. Jones, and Michael R. Buhl

Lawrence Livermore National Laboratory
7000 East Avenue, L-156, Livermore, CA 94550

Please send correspondence to:
Dr. Gregory A. Clark
Lawrence Livermore National Laboratory
7000 East Avenue, L-156, Livermore, CA 94550
Telephone: (510)423-9759
FAX: (510)422-3013
E-Mail: clark9@llnl.gov

Abstract

Our system classifies the condition (intact or single leg separated) of in vivo Bjork-Shiley Convexo-Concave (BSCC) heart valves by processing acoustic measurements of clinical heart valve opening sounds. We use spectral features as inputs to a two-stage classifier, which first classifies individual heart beats, then classifies valves. Performance is measured by probability of detection and probability of false alarm, and by confidence intervals on the probability of correct classification. The novelty of the work lies in the application of advanced techniques to real heart valve data, and extensions of published algorithms that enhance their applicability. We show that even when given a very small number of training samples, the classifier can achieve a probability of correct classification of 100%.

"Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48."

Summary

1.0 Introduction

Prosthetic heart valves and the many great strides in valve design have been responsible for extending the life spans of many people with serious heart conditions. Even though the prosthetic valves are extremely reliable, they are eventually susceptible to the long-term fatigue and structural failure effects expected for mechanical devices operating over long periods of time. The purpose of our work is to classify the condition of in vivo Bjork-Shiley Convexo-Concave (BSCC) heart valves by processing acoustic measurements of heart valve sounds. The structural failures of interest for BSCC valves is called single leg separation (SLS). SLS can occur if the outlet strut cracks and separates from the main structure of the valve. We measure acoustic opening and closing sounds (waveforms) using high sensitivity contact microphones on the patient's thorax. For our analysis, we focus our processing and classification efforts on the opening sounds because they yield direct information about outlet stut condition with minimal distortion caused by energy radiated from the valve disc.

Our heart valve analysis system consists of algorithms for data acuisition, signal processing and signal classification. Data acqisition and signal processing are discussed in other papers. This paper concentrates on classification algorithms and results, with brief descriptions of the acquisition and signal processing provided as necessary to understand the inputs to the classifiers. We decribe the extraction and selection of spectral features from the spectral estimates computed by the signal processing system. We show that a two-stage classifier, which first classifies individual heart beats, and then classifies valves is very effective in correctly classifying the valves represented in our data set. We measure performance by constructing confusion matrices, receiver operating characteristic (ROC) curves showing probability of detection and probability of false alarm (sensitivity and 1-specificity), and by specifying statistical confidence intervals on the probability of correct classification. We show that even given a very small number of training samples, the classifier achieves probability of correct classification of 100% for a real test set.

2. Measurements and Signal Pre-Processing

In this paper, we discuss feature extraction, feature selection and classification techniques developed to analyze BSCC heart valves sounds and assign them to one of two classes; those which correspond to valves having the single leg separation (SLS) condition and those which are intact (INT). The structural condition of interest for BSCC valves, called single leg separation (SLS), can occur if the outlet strut cracks and/or separates from the main structure of the valve. We measure acoustic opening and closing sounds (beats or waveforms) using high sensitivity contact microphones on the patient's thorax. For our analysis, we focus our processing and classification efforts on the opening sounds because they yield direct information about outlet strut condition with minimal distortion caused by energy radiated from the valve disc.

The heart sounds are noisy transient waveforms with a duration of approximately 10 msec to 20 msec. The opening sounds have a much smaller amplitude than the closing sounds have, and the low signal-to-noise ratio for opening sounds complicates our analysis. The statistical variability of the opening sounos can be quite large, so we must be careful to screen out beats that are not statistically acceptable to be used as a classifier input. We have developed a beat monitor that performs this screening based upon spectral content of the waveforms.

3.0 Classification Using Supervised Learning

Our approach to classifying heart valve structural failures is depicted in Figure 1. Here we see that after the data have been acquired and processed, a set of feature vectors is extracted and processed by a classifier to decide on the condition of this valve. Our approach is based on estimating the spectrogram surface (power vs. frequency vs. beat or time), which displays the resonant peaks of the heart valve under investigation. Note that there is a separate spectrogram for the valve closing and opening beats. The spectrogram is obtained by first digitizing the acoustic data, preprocessing (filtering, smoothing, etc.) and then estimating the power/energy spectrum at each beat producing an individual slice of the spectrogram surface. Unfortunately, due to nonstationarities within the valve dynamics (disc closure mechanism) or the acoustic medium, these resonant frequencies vary as a function of time; thus, searching for a single "fixed" fracture frequency can be futile.

In our approach, a supervised classifier (probabilistic neural network) is trained to classify heart valves into two classes: Intact (INT) or Single Leg Separated (SLS). We train our classifier using clinical signals measured from valves that were explanted and examined after measurement, so the condition ("ground truth") of the valves is positively known. We test the classifier using similar signals from a "blinded" data set of signals measured from valves that were explanted. The processing steps are summarized here, and will be described in detail in the paper.

Preprocessing

The preprocessing techniques [3-9] serve several functions. First, they cut out heart beat opening sounds (see Fig. 2 for an example of the real data). Second, they screen out opening sounds that are not representative of what we call "good" beats that are of sufficient quality to warrant further processing. The algorithms screen out opening sounds (beats) that lie outside objective statistical bounds we defined for them. Third, we calculate ARMA and lattice models for the beats and create estimated signal spectra from them. These spectral estimates serve as the basic for the features we extract. For a detailed description, see [3-8].

Feature Extraction

Features are chosen initially based upon judgment obtained from knowledge of the processes and measurements involved. We use features of the estimated frequency spectra of individual opening sounds. We use ARMA and lattice model spectral estimates [3-9]. The features used in our current prototype system are the magnitudes of the spectral estimates in a given small band of frequencies. The width of the band is allowed to vary to produce a large number of features, one at each width. We then use feature selection algorithms to choose the subset of optimal features (see Figs. 3 and 4).

Feature Selection

We use both automatic feature selection procedures, along with manual validation to apply human judgment. We use algorithms for automatically searching through the set of features and ranking them in order of importance. For example, the sequential forward selection algorithm ranks the features one by one in order of a statistical measure of distance between cluster centers in feature space. The branch and bound algorithms produce a list of the optimal set of features (given a number of features to choose a priori) [9-13, 16]. After using feature selection algorithms, we generally perform a manual inspection of the one- and two-dimensional cluster plots in feature space to further reduce the feature set, to gain physical insight and to allow the insertion of valuable human judgment into the process. We choose the number of features according to the well-known rule of thumb that says that the number of independent training samples (feature vectors) should be greater than or equal to approximately 5 times the number of features contained in a feature vector. Thus the number of features we can use is limited by the number of training samples (valves) available [16, 19].

Classification

We use a probabilistic neural network (PNN) classifier [17]. The PNN is a Bayes classifier which estimates the conditional probability density functions (pdf's) for the input feature vector, given the class to which it belongs (see Figs. 5-7). It is a Parzen window estimator, which converges to the Bayes optimal decision boundary as the number of training samples approaches infinity. A classifier smoothing parameter must be chosen, and we created an automatic optimization algorithm for choosing it, based upon searching for the largest value of the probability of correct classification. We use a two-stage classifier, in which the first stage uses a PNN to classify beats and the second stage uses hypothesis testing with a threshold to detect SLS or INT valves. Future plans include using a PNN with other features in the second stage (see Figs. 8 and 9).

4.0 Classification Results

We processed the training and test data described above, and obtained the following performance. For training, we found the the system is very robust, even for a small samples size (10 training samples). The ROC curve derived from the confusion matrix (see Fig. 10) shows P(Detection) = 1.0 and P(False Alarm) = 0. However, we must keep in mind that our confidence in these estimates is not high, because the sample size is so small. To evaluate this confidence, we compute

$$P(\text{Correct Classification}) = .5[P(\text{Detection}) + 1 - P(\text{False Alarm})]$$

for this problem [16]. We show that P(Correct Classification) has a binomial distribution, and we can define a 95% confidence interval about the estimated P(Correct Classification). This confidence interval is a function of the number of training samples N, and we show how the confidence interval tightens as N increases. For our data set, N= 10. For our estimated value of P(Correct Classification) = 1.0, we see that the 95% confidence interval has a lower bound of 71% and an upper bound of 100%. If we were to have 1000 samples, the confidence interval bounds are 99.6% and 100%. If we were to have N = 100 samples, the bounds are 96% and 100%. See Figs. 11-13. Clearly, we can improve our performance if we can obtain more heart valves to study. See Figs. 14-19 for details of the classification results.

5.0 Conclusions

Our classification system is very robust, even though the number of samples is small. However, the P(correct classification) has a very wide confidence interval, because the training sample size is so small. Ongoing and future work includes continued study using data from more heart valves, as they continue to be provided by the clinical team explanting and studying valves. Within the next year, we expect to obtain data from from approximately 100 explanted valves. For this sample size, we expect to obtain very tight and meaningful confidence intervals for P(correct classification).

Acknowledgments

The authors gratefully acknowledge the many helpful contributions provided by Dr. James V. Candy of LLNL, Dr. David Wieting of the Shiley Heart Valve Research Center (SHVRC), Rebecca Inderbitzen of SHVRC.

References

1. Chandler, J. G., "Broken Heart Valves," *Emergency Medicine,* Cahners Pub, 1992.

2. Plemons, T., Hovenga, M., Reeder, H., Dow, J. J., Interbitzen, B., Chia, R., and D. Wieting, "Classification of Bjork-Shiley Convexo-Concave valve status by detection of the intact outlet strut resonant frequency," *IEEE Confr. Medicine & Biology,* 1993.

3. Mullenhoff, Carmen, "Signal Processing of Shiley Heart Valve Data for Fracture Detection", *LLNL-Report*, UCRL-ID-113760, 1993.

4. M. R. Buhl, G. A. Clark, J. V. Candy, G. H. Thomas, "Detection of 'Single-Leg Separated' Heart Valves Using Statistical Pattern Recognition With the Nearest Neighbor Classifier", *LLNL-Report*, UCRL-ID-114802, July 16,1993.

5. Crawford, Susan L., and Thomas, Graham H., "In-Vivo Classification of the Bjork-Shiley Convexo-Concave Heart Valve from Acoustic Signatures", *LLNL-Report*, UCRL-ID-114819, 1993.

6. Candy, J., Barnes, F., Heart valve processing: a feasibility study. *LLNL Report.* , UCRL-ID-107630, 1991.

7. J. V. Candy, H. E. Jones, "Processing of Prosthetic Heart Valve Sounds for Single Leg Separation Classification," Lawrence Livermore National Laboratory report UCRL-JC-116175, to appear in the Journal of the Acoustical Society of America, 1994.

8. J. V. Candy, H. E. Jones, "Classification of Prosthetic Heart Valve Sounds, a Parametric Approach," Lawrence Livermore National Laboratory report UCRL-JC-107630, submitted to the Journal of the Acoustical Society of America, 1994.

9. G. A. Clark, S. K. Sengupta, M. R. Buhl, R. J. Sherwood, P. C. Schaich, N. Bull, R. J. Kane, and M. J. Barth, "Detecting Buried Objects by Fusing Dual-Band Infrared Images," *Invited paper*, Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 1-3, 1993.

10. G. A. Clark, S. K. Sengupta, P. C. Schaich, R. J. Sherwood, M. R. Buhl, J. E. Hernandez, D. J. Fields, and M. R. Carter, "Data Fusion for the Detection of Buried Land Mines," International Symposium on Substance Identification Technologies, European Optical Society (EOS)/The International Society for Optical Engineering (SPIE), Congress Centre Innsbruck, Innsbruck, Austria, October 4-8, 1993.

11. G. A. Clark, S. K. Sengupta, R. J. Sherwood, J. E. Hernandez, M. R. Buhl, P. C. Schaich, R. J. Kane, M. J. Barth, and N. K. DelGrande, "Sensor Feature Fusion for Detecting Buried Objects," *Invited paper,* SPIE's 1993 International Symposium and Exhibition on

Optical Engineering and Photonics in Aerospace and Remote Sensing, Conference on Underground and Obscured Imaging and Detection, Orlando, Florida, April 12-16, 1993.

12. N. K. DelGrande, P. F. Durbin, D. E. Perkins, G. A. Clark, R. J. Sherwood, J. E. Hernandez, and A. B. Shapiro, "Dual-Band Infrared Capabilities for Imaging Buried Object Sites, " SPIE's 1993 International Symposium and Exhibition on Optical Engineering and Photonics in Aerospace and Remote Sensing, Conference on Underground and Obscured Imaging and Detection, Orlando, Florida, April 12-16, 1993.

13. G. A. Clark, J. E. Hernandez, S. K. Sengupta, R. J. Sherwood, P. C. Schaich, M.R. Buhl, R. J. Kane, M.J. Barth, and N. K. DelGrande, "Computer Vision and Sensor Fusion for Detecting Buried Objects," Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 26-28, 1992.

14. G. A. Clark, J. E. Hernandez, N.K. DelGrande, R. J. Sherwood, S.Y. Lu, P. C. Schaich, and P.F. Durbin, "Computer Vision for Locating Buried Objects," Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 4-6, 1991.

15. Kohonen, T. Self-Organization an Associative Memory, Springer-Verlag, 3rd edition, 1989.

16. Duda R. O., Hart, P. E. *Pattern Classification and Scene Analysis.* New York, NY: Wiley; 1973.

17. Specht, D. F., "Probablistic Neural Networks," Vol 3, *Neural Networks,* 1990.

18. Candy, J. *Signal Processing: The Modern Approach.* New York, N.Y.: McGraw Hill, 1988.

19. Devivjer, P. A. and Kittler, J., Pattern Recognition Theory and Applications, Springer Verlag, 1986, p. 7.

20. Jain AK. Fundamentals of Digital Image Processing. Englewood, NJ: Prentice-Hall

21. Rabindra, P., Sengupta, S. K., and Welch, R. M., "An Interactive Hybrid Expert System for Polar Cloud and Surface Classification, " 1992, to be published.

22. Young, T. Y., and Fu, K-S, Handbook of Pattern Recognition and Image Processing, Academic Press, 1986.

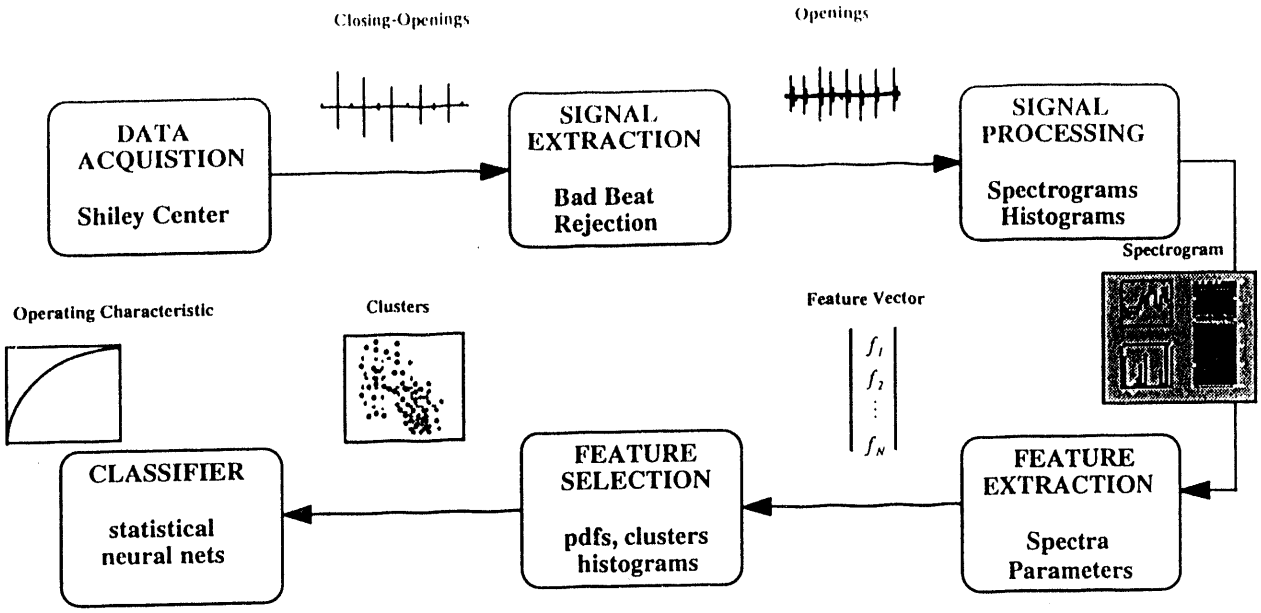


Figure 1

Block diagram of the overall heart beat analysis system.

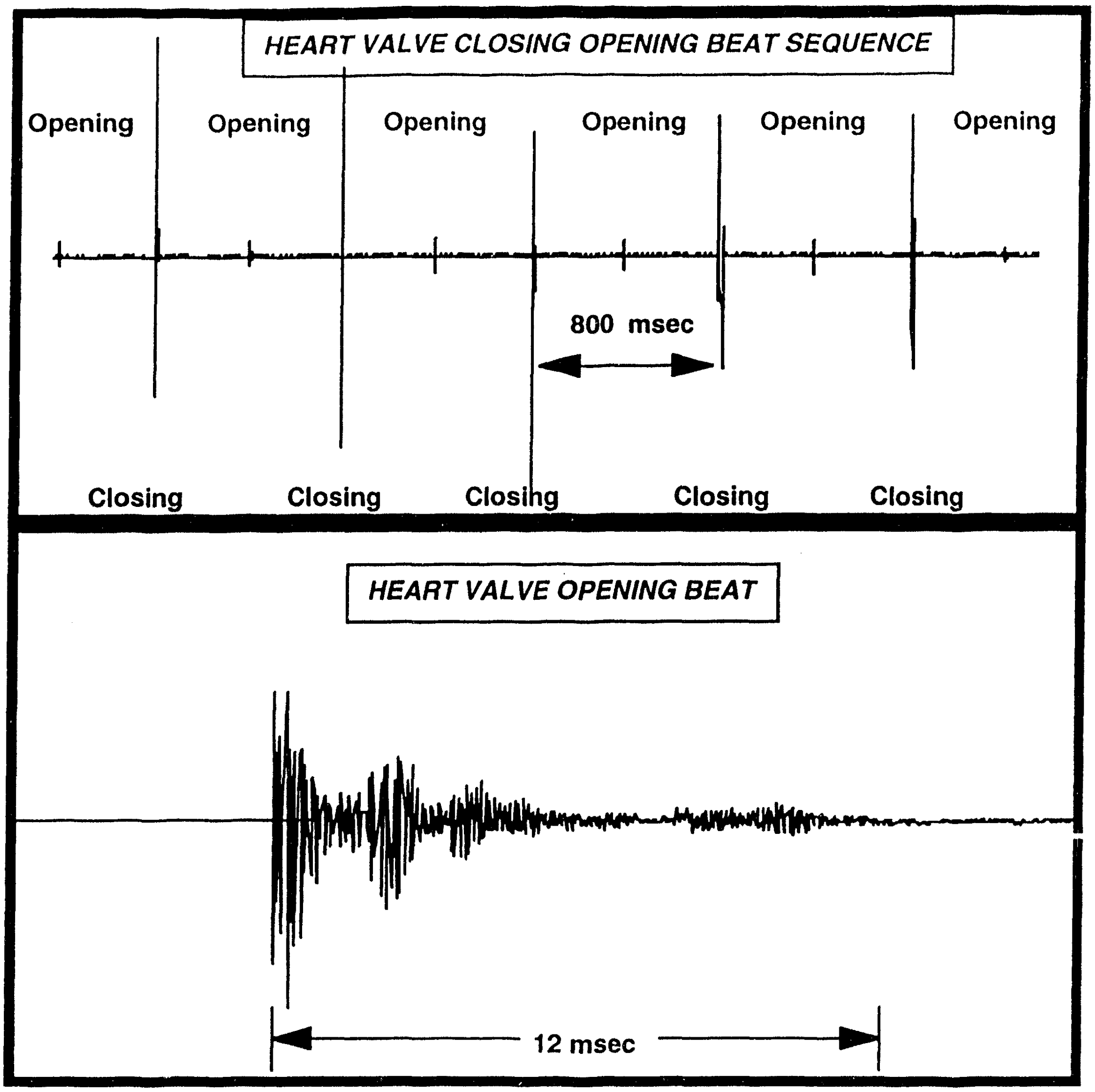
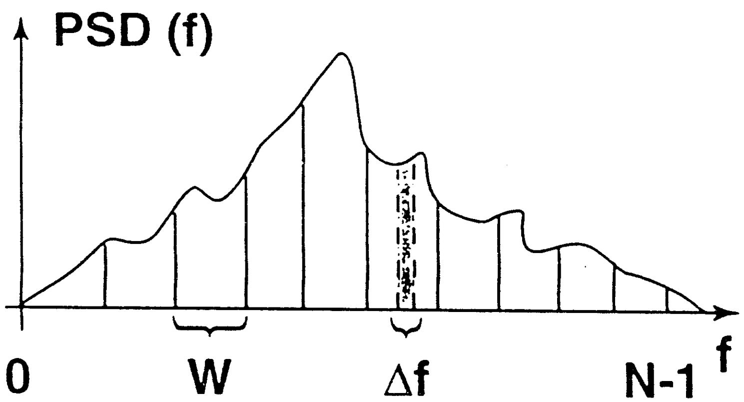


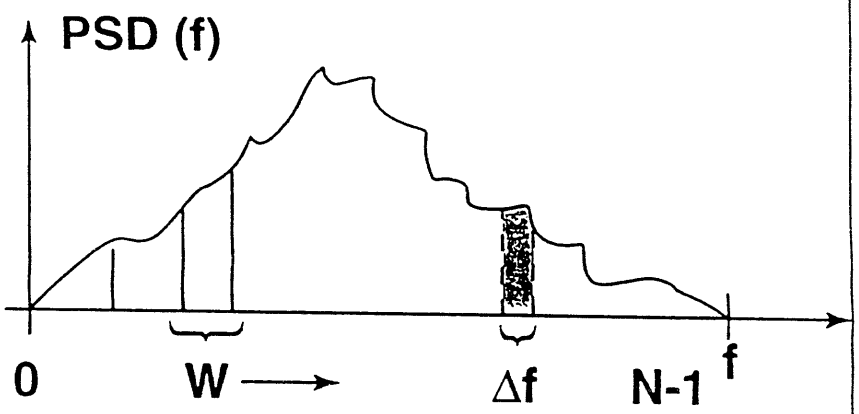
Figure 2
The raw data are cut into sections called "beats," consisting of opening sounds.

We use both fixed window features and sliding window features

Fixed adjacent window features



Sliding window features



- W = Frequency window width
- Δf = Frequency bin width
- We get number bands = $\dfrac{N}{W}$
- Fixed window bands form a subset of the sliding window bands

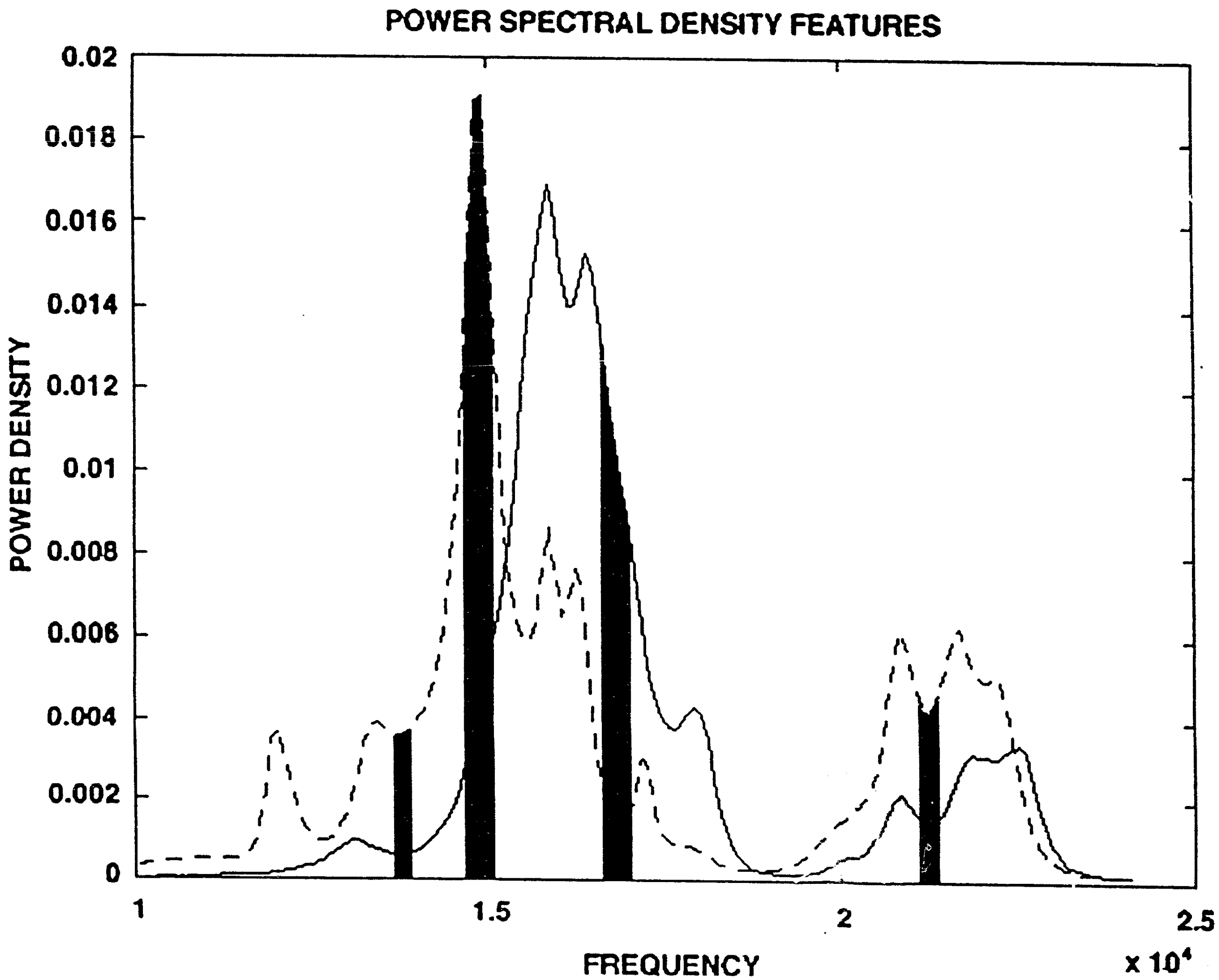
- The window slides
- $\hat{\mu}(f)$ = Mean in window
  $$= \frac{1}{W} \sum_{c=1}^{W} PSD_i(f)$$
- Number bands = N
- Sliding window bands may perform better because they account for all shifts in window centers

0394/S/GAC

FIG 3

Power spectral density estimates averaged over all beats for a single valve—overlay shows features chosen by the feature selector

- Solid Line: Intact Valve
- Dashed Line: SLS Valve



0394/9/GAC

FIG 4

We Use the PNN (Probabilistic Neural Net) for Bayesian Classification

Recall the Bayes Decision Rule for a Two-Class Problem:

$$d(\underline{X}) = \theta_A \quad \text{if} \quad h_A\, l_A\, f_A(\underline{X}) > h_B\, l_B\, f_B(\underline{X})$$

$$d(\underline{X}) = \theta_B \quad \text{if} \quad h_A\, l_A\, f_A(\underline{X}) > h_B\, l_B\, f_B(\underline{X})$$

where

θ = state of nature = θ_A or θ_B

$d(\underline{X})$ = decision based on $\underline{X}$

$f_A(\underline{X})$ = pdf for class A

$f_B(\underline{X})$ = pdf for class B

$\underline{X} = [x_1\, x_2 \ldots, x_p]^T$ = measurements

l_A, l_B = loss functions for A, B

h_A, h_B = prior probability of occurrence for patterns from classes A, B, $h_B = 1 - h_A$

FIG. 5

0394/12/GAC

The PNN estimates probability density functions from training data

$$f_A(\underline{X}) = \frac{1}{(2\pi)^{\frac{p}{2}}\ \sigma^p}\left(\sum_{i=1}^{m}\right)\exp\left\{-\frac{(\underline{X}-\underline{X}_{Ai})^T(\underline{X}-\underline{X}_{Ai})}{2\,\sigma^2}\right\}$$

i = Pattern number

m = Number of training patterns

$\underline{X}$ = Pattern (feature) vector under test

$\underline{X}_{Ai}$ = i th training pattern from class A

σ = Smoothing parameter

p = Dimension of $\underline{X}$

$\sigma \longrightarrow 0$ $\Rightarrow$ PNN = Nearest neighbor classifier

$\Rightarrow$ Estimated pdf has distinct modes corresponding to locations of training samples

$\sigma \longrightarrow \infty$ $\Rightarrow$ PNN = Hyperplane (linear) classifier

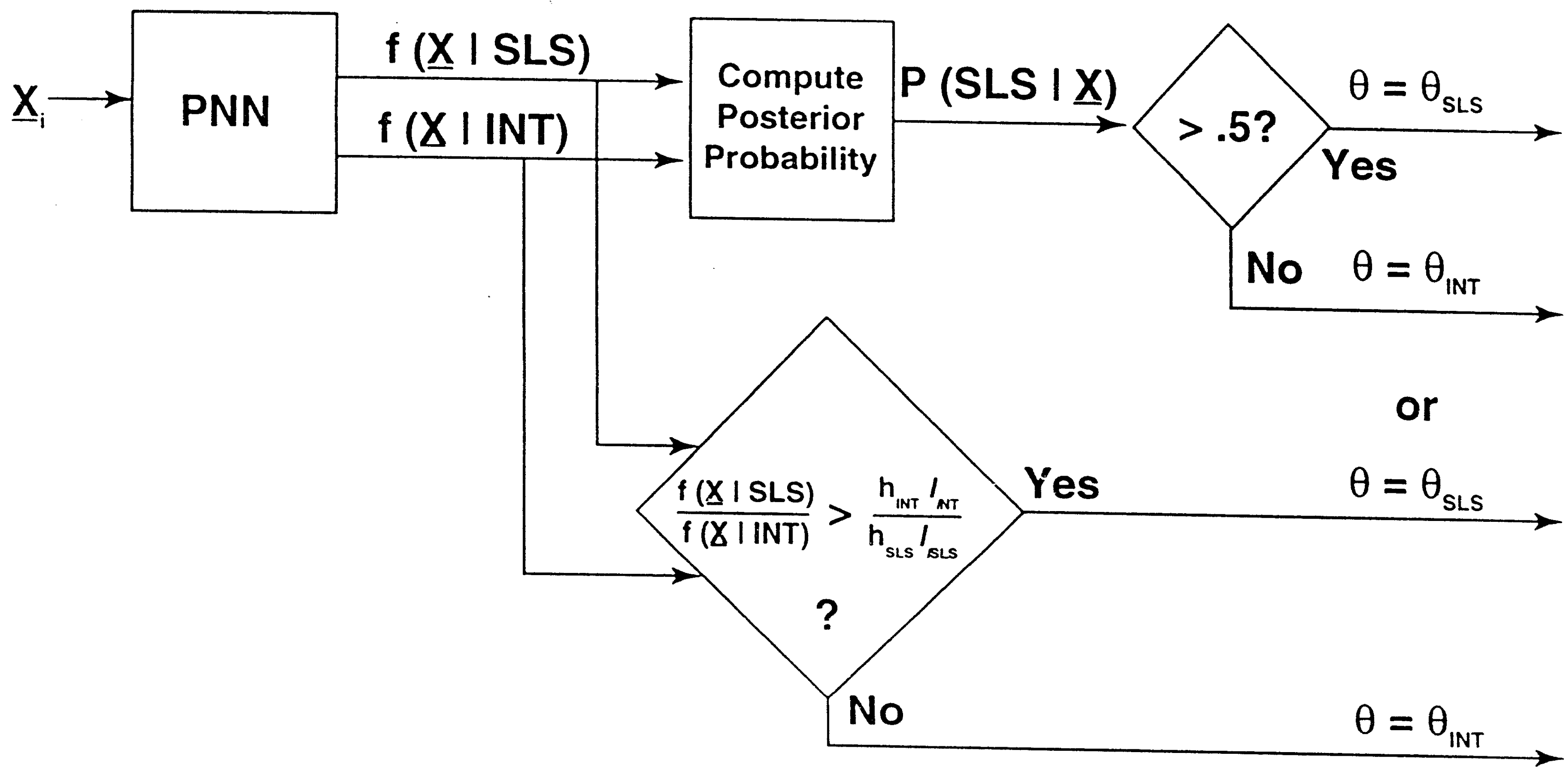
$\Rightarrow$ Broad smoothing, interpolation

$\Rightarrow$ Estimated pdf approaches Gaussian

0394/13/GAC

FIG. 6

For a given feature vector $\underline{X}_i$ the PNN produces pdf's / posterior probabilities and a decision

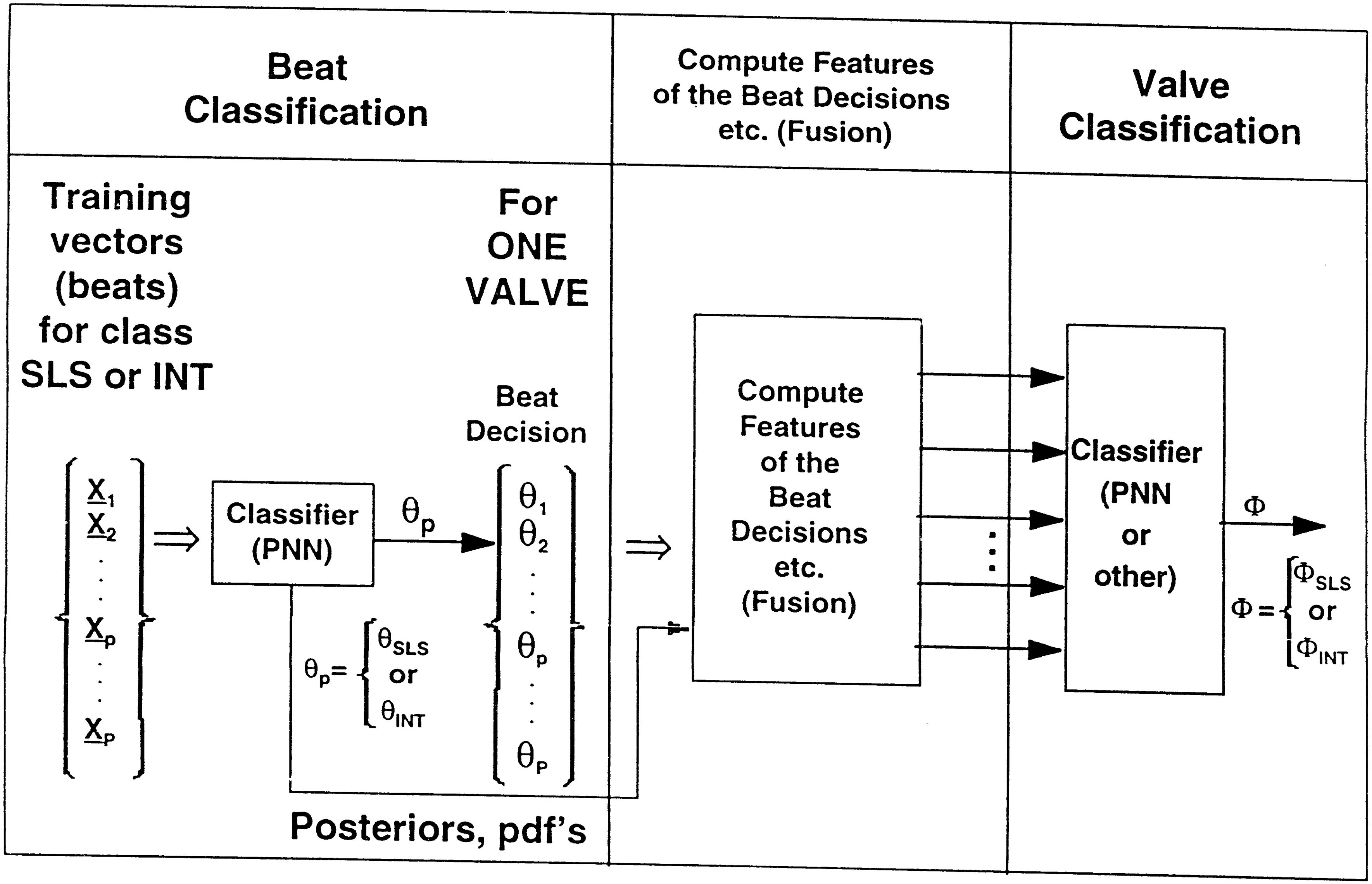


$$P(SLS \mid \underline{X}) = \frac{h_{SLS} f(\underline{X} \mid SLS)}{h_{SLS} f(\underline{X} \mid SLS) + h_{INT} f(\underline{X} \mid INT)}$$

0394/14/GAC

FIG. 7

We use a two-stage classification scheme: classify beats first, valves second



0394/18/GAC

FIG. 8

Our current two-stage classifier uses one feature in the second stage

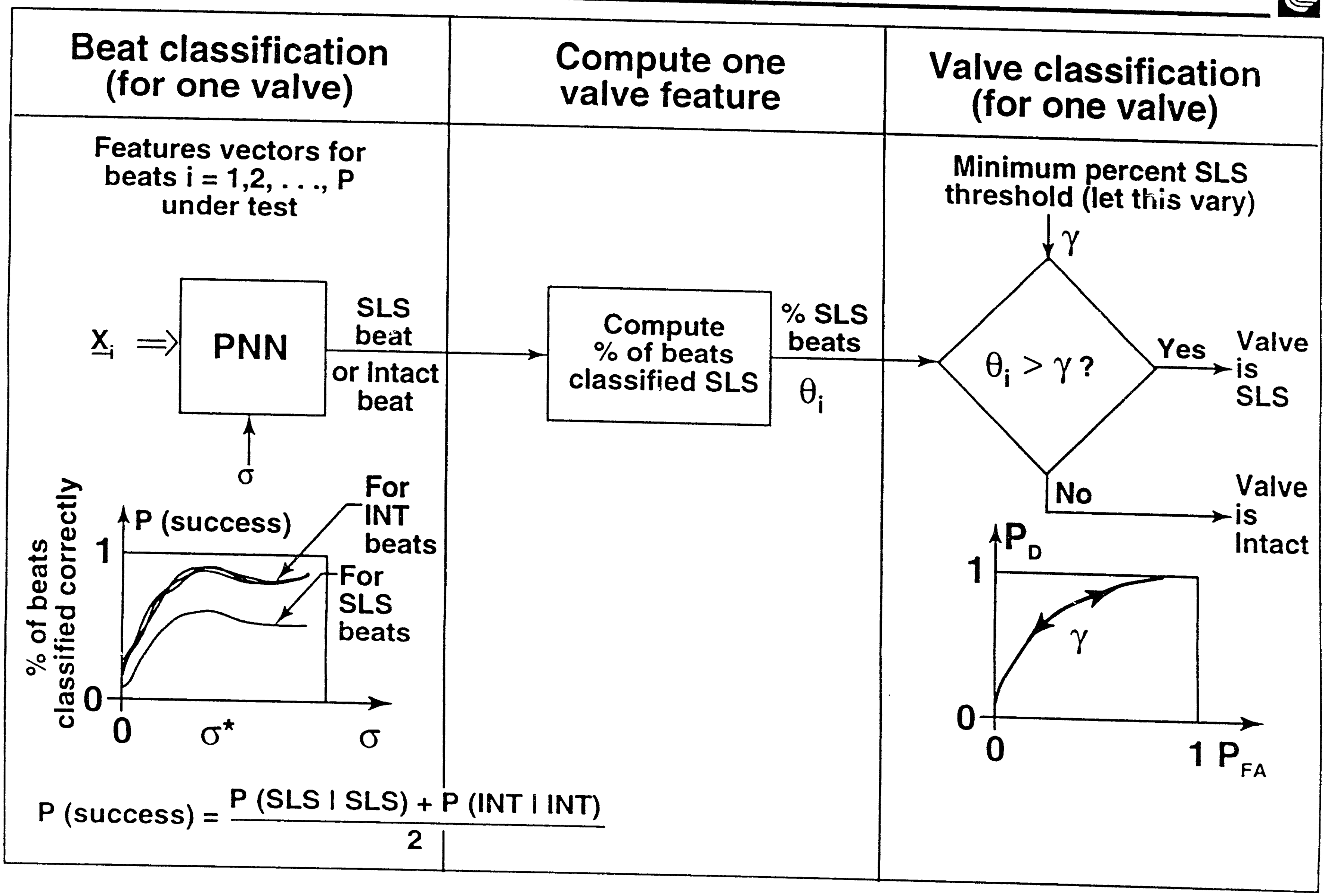


FIG. 9

0394/19/GAC

The confusion matrix is a useful tool generated from training results

		Truth	
		SLS	INTACT
Classifier Results	SLS	$\hat{P}$ (SLS | SLS) = P (detection) = sensitivity	$\hat{P}$ (SLS | INT) = P (false alarm)
	INT	$\hat{P}$ (INT | SLS) = P (miss)	$\hat{P}$ (INT | INT) = specificity = P (spec)

$$\hat{P}\text{(detection)} + \hat{P}\text{(miss)} = 1$$

$$\hat{P}\text{(false alarm)} + \hat{P}\text{(spec)} = 1$$

$$\hat{P}\text{(correct classification)} = \tfrac{1}{2}\,[\hat{P}(\text{SLS} \mid \text{SLS}) + \hat{P}(\text{INT} \mid \text{INT})] = \tfrac{1}{2}\,[\text{sensitivity} + \text{specificity}]$$

$$\hat{P}\text{(error)} = 1 - \hat{P}\text{(correct classification)}$$

FIG. 10

0394/15/GAC

We performed two experiments

Experiment IA

1. Training: Used 17 confirmed sessions for 10 valves, all with normal heart conditions

#SLS	Sessions	= 10	(6 valves)
# INT	Sessions	= 7	(4 valves)
	Total	= 17	

2. Testing: Used 11 confirmed sessions, with mixed heart conditions

#SLS	Sessions	= 7	(7 valves)
# INT	Sessions	= 4	(3 valves)
	Total	= 11	

Experiment II

1. Training: (Same as Experiment I)

#SLS	Sessions	= 10	(6 valves)
# INT	Sessions	= 7	(4 valves)
	Total	= 17	

2. Testing: Used 20 sessions from various sources, with mixed heart conditions, 3 valves were aortic

#SLS	Sessions	= 11
# INT	Sessions	= 9
	Total	= 20

FIG 11

0394/32/GAC

The confidence interval bounds are calculated, then displayed in graphical form

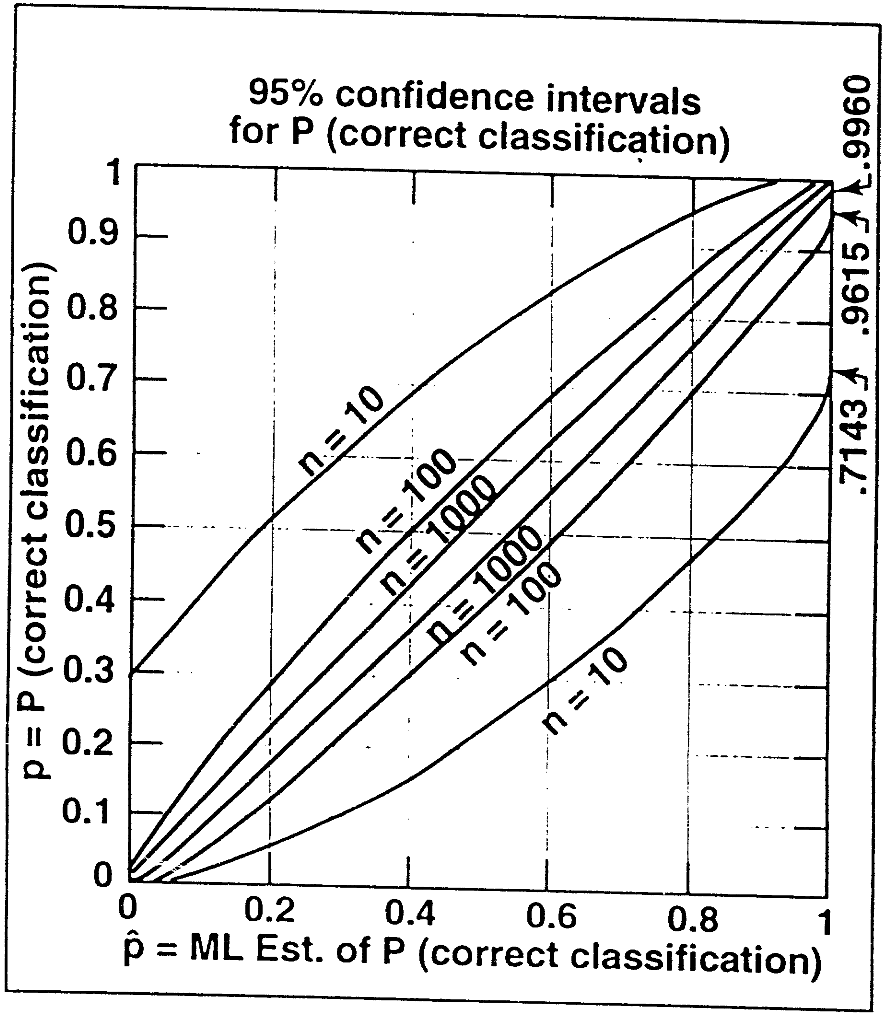


FIG 12

0394/28/GAC

We want to provide the bounds for the case in which $\hat{p} = 1$ (no incorrect classification)

- We will provide the bounds for whatever $\hat{p}$ we obtain during training (from our confusion matrix)
- However, Shiley has requested our "confidence" when $\hat{p} = 1$, so we can specify it as a function of n.

For example:

n	Lower Bound	Upper Bound
10	.7143	1.0
100	.9615	1.0
1000	.9960	1.0

This quantifies why we need a large sample size!

FIG. 13

0394.29/GAC

Experiment I: Summary of performance specifications based on training data

- Number of training samples: $n = 17$
- Test statistic: θ = % beats classified SLS
- Decision threshold: $\gamma = .5$
- $\hat{P}_D = 1, \hat{P}_{FA} = 0$
- Maximum likelihood estimate of P(correct classification) = $\hat{p} = 1$
- Confidence interval about the true P(correct classification) = p

$$P\,(.8095 < p < 1.0) = .95$$

FIG 19

0394/38/GAC

Experiment I: Classification results (testing) for blinded data from 10 confirmed valves were excellent

Tape Log #	True Class	Classification	θ	d
024/211	INT	INT	0	-.50
052	SLS	SLS	100	.50
091	INT	INT	0	-.50
106	SLS	SLS	100	-.50
221	INT	INT	7	-.43
395	SLS	SLS	68	.18
440	SLS	SLS	56	.06
451	SLS	SLS	100	.50
480	SLS	SLS	100	.50
8707	SLS	SLS	99	.49

FIG. [illegible]

0394/39/GAC

Experiment I: Summary of test results– Confusion matrix developed after truth was known

#SLS sessions = 7, #INT sessions = 4

Truth / Classification	SLS	INT
SLS	7/7 = 1	0/4 = 0
INT	0/7 = 0	4/4 = 1

$$P(cc) = \frac{1}{2}\,[\,1 + 1\,] = 1$$

FIG. 16

0394.40/GAC

Experiment II: Classification results for blinded data from confirmed intact valves

Tape Log #	Truth	Classification	Whiteness Test (%)	θ_i	d_i
AZL001	INT	INT	5	26	-.24
AZL003	INT	INT (3 features)	11	15	-.35
AZL004 (aortic)	INT	SLS	5	97	.47
AZL012	INT	*NC	–	–	–
AZL013	INT	SLS	18	100	.50
AZL014	INT	INT (3 features)	13	47	-.03
AZL025-1 (aortic)	INT	SLS (data?)	16	84	.16
AZL025-2 (aortic)	INT	*NC	–	–	–
GOT-01 (aortic)	INT	*NC	–	–	–

Leuven data are different, need to train on Leuven data

*NC = no classification due to insufficient data

0394/41/GAC

Experiment II: Classification results (testing) for data from confirmed SLS valves

Tape Log #	Truth	Classification	Whiteness Test (%)	θ_i	d_i
BC131	SLS	SLS	5	100	.50
BC409	SLS	SLS	5	100	.50
BC490	SLS	SLS	13	100	.50
BC610	SLS	SLS	11	100	.50
BC824	SLS	SLS	13	100	.50
BC832	SLS	SLS	5	98	.48
BC1026	SLS	NC	–	–	–
BC1031	SLS	SLS	5	100	.50
USA-A	SLS	SLS	5	100	.50
USA-B	SLS	SLS	10	100	.50
USA-C	SLS	SLS	5	100	.50

0394/42/GAC

Experiment II: Summary of test results– confusion matrix developed after truth was known

SLS sessions = 11, but only 10 were classified

INT sessions = 9, but only 6 were classified

Classification \ Truth	SLS	INT
SLS	10/10 = 1	3/6 = .5
INT	0/10 = 0	3/6 = .5

$$P(cc) = \frac{1}{2}\ [\,1 + .5\,] = .75$$