

Identification and Diagnosis of Valvular Heart Diseases Using Time Frequency Analysis and Machine Learning Technique

Prabhu Vignesh Kumar R¹, Senthil Kumar K²

¹M.Tech, Computer Science and Engineering, SRM University, Chennai, India

²Asst professor .Department of Computer Science and Engineering SRM University, India,

Abstract— Valvular Heart Disease (VHD) causes a lot of asymptomatic pathological disorders like stroke, Mitral Stenosis and Regurgitation, Tricuspid Stenosis and Regurgitation. The normal heart sound recorded by Phonocardiogram which can also record pathological sound called Murmurs like Early diastolic murmur (EDM), late diastolic murmur (LDM), early systolic murmur (ESM) and late systolic murmur (LSM). The present diagnosis system uses moving pictures which increases the complexity of computational process and main memory utilization. In order to overcome these limitations, Time Frequency Analysis is proposed to convert the analog sound signals into program understandable expression by Renyi marginal entropy (RME) with help of Renyi entropy calculation. Time and location of murmur indicate the respective disorders. With the number of training data, Machine learning concept called Support vector machine (SVM) makes learning and classification process. The learning process is done by the algorithm, which can make the classification of input data by the knowledge of training data. According to the parameter given to the SVM, it makes classification of given data. And algorithm can remember the parameter and adopt itself with respect to the upcoming training data. Finally the test datum is classified and it was stored in database and the particular stored data was given as input to GUI.

Keywords— Index Terms— Valvular Heart Disease - Pathological sound - Time Frequency Analysis - Program Understandable Expression - Support Vector Machine - learns and classifies -Result of Test data in GUI.

I. INTRODUCTION

Valvular heart disease[4] is any disease process involving one or more of the valves of the heart (the aortic and mitral valves on the left and the pulmonary and tricuspid valves[4] on the right). Valve problems may be congenital (inborn) or acquired (due to another cause later in life). Treatment may be with medication but often (depending on the severity) involves valve repair or replacement (insertion of an artificial heart valve). Specific situations include those where additional demands are made on the circulation, such as in pregnancy. Valvular heart disease is characterized by damage to or a defect in one of the four heart valves: the mitral, aortic, tricuspid or pulmonary [4].

The mitral and tricuspid valves control the flow of blood between the atria and the ventricles (the upper and lower chambers of the heart). The pulmonary valve controls the flow of blood from the heart to the lungs, and the aortic

valve governs blood flow between the heart and the aorta, and thereby the blood vessels to the rest of the body.

Normally functioning valves ensure that blood flows with proper force in the proper direction at the proper time. In valvular heart disease, the valves become too narrow and hardened (stenotic) to open fully, or are unable to close completely (incompetent). To compensate for poor pumping action

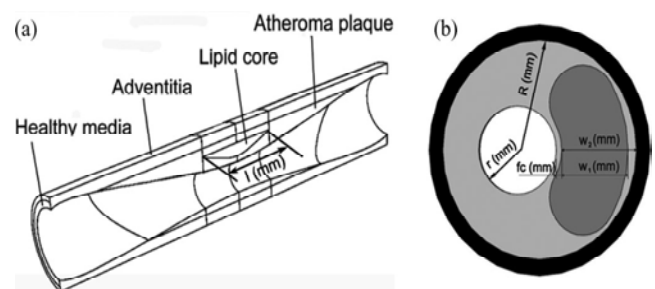


Fig.1 (Heart valve pathological facts)

The heart muscle enlarges and thickens, thereby losing elasticity and efficiency. In addition, in some cases, blood pooling in the chambers of the heart has a greater tendency to clot, increasing the risk of stroke or pulmonary embolism

II. INTRODUCTION ON HEART SIGNAL PROCESSING

The training data which are having heart murmurs can be given as the input to signal pre processing unit which can convert the analog sound signals into program understandable expression. The signal pre-processing unit has the following units

A. Fourier Transforms

The Fourier Transform [3] is extensively used in the field of Signal Processing. In fact, the Fourier Transform is probably the most important tool for analyzing signals in that entire field. A signal is any waveform (function of time). This could be anything in the real world - an electromagnetic wave, the voltage across a resistor versus time, the air pressure variance due to your speech (i.e. a sound wave), or the value of Apple Stock versus time. Signal Processing then, is the act of processing a signal to obtain more useful information, or to make the signal more useful.

How can a signal be made better? Suppose that you are listening to a recording, and there is a low-pitched hum in the background. By applying a low-frequency filter, we can

eliminate the hum. Or suppose you have a digital photograph, and it is very noisy (that is, there are random specs of light everywhere).

Signal processing and Fourier transforms to filter out this undesirable "noise". So in this noise and heart murmur should be separated so here the following transforms [3] are used.

- The Quadratic Time Frequency Distributions
- Wigner Ville Distribution
- Short Time Fourier Transform
- The normalized Re'nyi marginal entropy (RME)

Each blocks are involves into the process of making unique parameters of the signal which can be used for further process.

B. Introduction on Learning And Classification

All tables and figures will be processed as images. You need to embed the images in the paper itself. Please don't send the images as separate files.

C. Support Vector Machine

The parametric functions which are obtained from the signal preprocessing are given as the input to learning and classification process. In the learning and classification process two major parts [3] are there.

- Learning process
- Classification process.

The learning process is done by the algorithm called support vector machine, which can make the classification of input data by the knowledge of training data. According to the parameter given to the SVM it makes classification of given data. And algorithm can remember the parameter and adopt itself according to the upcoming training data. The classification of data can be stored in the database for the further learning process. So Vapnik & Chervonenk dimensions (VC dimensions) [3] can fix the criteria for the classification by the knowledge of learning process.

D. Material and theoretical background

The analysis has been conducted on real biomedical data Obtained from [14]. The recordings used include both normal and abnormal PCG signals. In this paper, HSs are categorized into five classes: normal (NL), early aortic stenosis (EAS), late aortic stenosis (LAS), PS and mitral regurgitation (MR). We have used six PCG records in WAV format for each type of diseases. Each PCG signal contains 4096 or 8192 samples and the sampling frequencies used are given in Table 1. As an example, Fig. 1 shows the time-amplitude display of different cases of abnormal HS. It can be seen that the first component (S1) and the second component (S2) are easily identified but detached or mixed with murmur.

E. Time-frequency distributions

The quadratic TFD constitutes a powerful tool in the anal sis of non-stationary signals, that is, signals whose frequency co tent varies with time. The quadratic class can be expressed as [15]

$$C_x(t,f) = \iiint e^{j2\pi(n-t-f\tau)} g_x(v,\tau) x\left(u+\frac{\tau}{2}\right) x^*\left(u-\frac{\tau}{2}\right) dv d\tau dt \tag{1}$$

Where t is the time, f is the frequency, t is the time-Lag, n is the Doppler frequency and g (n, t) is the Doppler-Lag kernel of the distribution, x (t) is the analytical form of the signal under consideration, x*(t) its complex conjugate. All the integrals are from -1 to +1, unless otherwise stated. A choice of a particular kernel function yields a particular quadratic TFD with its own specificities [15]. In particular, if the kernel of distribution is unity go (n, t) = 1, we obtain the WVD written as

$$W_x(t, f) = \int x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \tag{2}$$

The WVD satisfies a large number of desirable mathematical properties, particularly, it preserves time and frequency shift and satisfies the marginal properties but caused interference terms. These interference terms are difficult since they may overlap with auto-terms or signal terms and thus make it difficult to visualize the time-frequency results.

$$S(\tau, f)^2 = \int_{-\infty}^{\infty} x(\tau)h(\tau-t)e^{-j2\pi ft} d\tau \tag{3}$$

The STFT defined as a windowed Fourier transform is given by where x (t) is the signal under consideration and h (t) is the Analysis window function. Another distribution called the spectrogram (SPEC), considered as a smoothed version of the WVD is defined as the squared magnitude of the STFT

$$|S(\tau, f)|^2 = \left| \int_{-\infty}^{\infty} x(\tau)h(\tau-t)e^{-j2\pi ft} d\tau \right|^2 \tag{4}$$

The implementation of the (3b) needs the use of two parameters: nfft the number of points used in the FFT to compute the spectrogram and 'nh' the number of points (or the length) of the analysis window function. Indeed, the time-frequency resolution o tainted is governed by the choice of analysis window. Therefore an evaluation of Re'nyi entropy provides the best combination between the two parameters generating a high resolution time-frequency. The expression of the SPEC can be written as a 2D convolution of the WVD of the signal Wx (t, f) by the WVD of the analysis window [15]

$$C_x(t, f) = \iint W_x(s, \xi) W_h(t-s, f-\xi) ds d\xi \tag{5}$$

The time frequency resolution trade-off of the spectrogram is controlled by the size of the analysis window h (t). A large kernel provides a narrow-band spectrogram, whereas a small kernel provides a wide-band spectrogram. So a narrowband spectrogram (good frequency resolution smoothes more in time than a wide-band spectrogram good time resolution).

F. Re nyi entropy

Here we have used a local measure of time-frequency signal concentration to find the optimal size of the kernel which gives good concentration everywhere in time-frequency plane and provides high concentration of

different components at different locations. The Reni entropy of order α is defined as

$$R_\alpha = \frac{1}{1-\alpha} \log_2 \iint C_x^\alpha(t, f) dt df \quad (6)$$

Where the rank of the Re'nyi measure with $\alpha \geq 2$. This criterion has been used to evaluate the complexity of a signal in the time-frequency plane in [16]. The normalization operation assures that the TFD behaves like a probability density function. Then minimizing the Re'nyi entropy for a given TFD is equivalent to maximizing its concentration, peachiness and resolution [17]. The best parameters of the kernel for TFD with respect to the minimal value of the Re'nyi will give a good localization of the energy. Recently, the performance of minimum entropy kernels for best TFDs and component counting has also been demonstrated [18]. In our applications, we have used the third order of Re'nyi entropy with volume normalized. Moreover the discrete time formulation of the third Re'nyi entropy for TFD with normalization volume is given by

$$RV_3 = -\frac{1}{2} \log_2 \sum_{k=-K}^K \sum_{n=-N}^N \left(\frac{C_x(n, k)}{\sum_{m=-K}^K \sum_{l=-N}^N |C_x(m, l)|} \right)^3 \quad (7)$$

Where n and k are variables for discrete time and discrete frequency, respectively, $(2N + 1)$ and $(2K + 1)$ are number of samples in time and frequency, respectively. Moreover it is proved in [18] that the third-order Re'nyi entropy for a uniformly distributed 2D random vector is given by

$$RV_3(P_{\text{uniform}}) = \log_2(2N + 1)(2K + 1) \quad (8)$$

Where P -(uniform) is the probability distribution function. Consequently, all TFD $C_x(t, f)$ possess the upper bound given by

$$E[RV_3(C_x(n, k))] < \log_2(2N + 1)(2K + 1) \quad (9)$$

These bounds may be used as a direct way of detecting main components of PCG signals and pathological murmur.

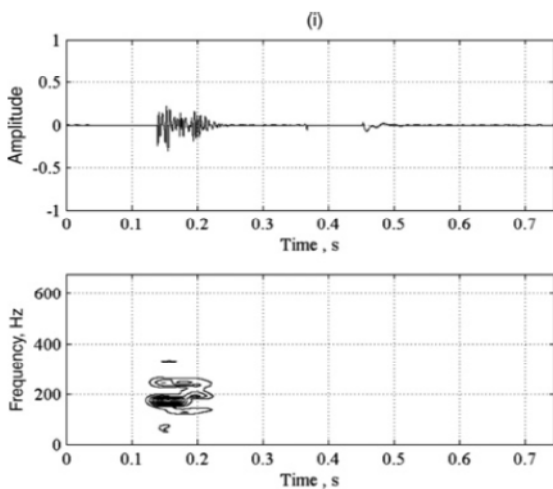


Fig.3 (Time and frequency form of single murmur)

The Re'nyi entropy is integrated with respect to frequency, in order to marginalize the signal's energy distribution over time. The normalized Re'nyi marginal entropy (RME) of third order is given by

$$RV_3(n) = -\frac{1}{2} \log_2 \sum_{k=-K}^{+K} \left(\frac{C_x(n, k)}{\sum_{m=-K}^K \sum_{l=-N}^N |C_x(m, l)|} \right)^3 \quad (10)$$

Where $(2N + 1)$ and $(2K + 1)$ are the number of samples in time and frequency, respectively.

G. Description of the method

Our proposed method for PCG analysis is based on the RME estimated on the time frequency results. The entropy profile of the RME in co junction with time-frequency analysis may be used to find an appropriate threshold to decide the existence of murmurs regions that always occur between the main components S1 and S2. We propose the upper bound of the Re'nyi entropy and investigate its use as an appropriate threshold to determine the boundary between signal events.

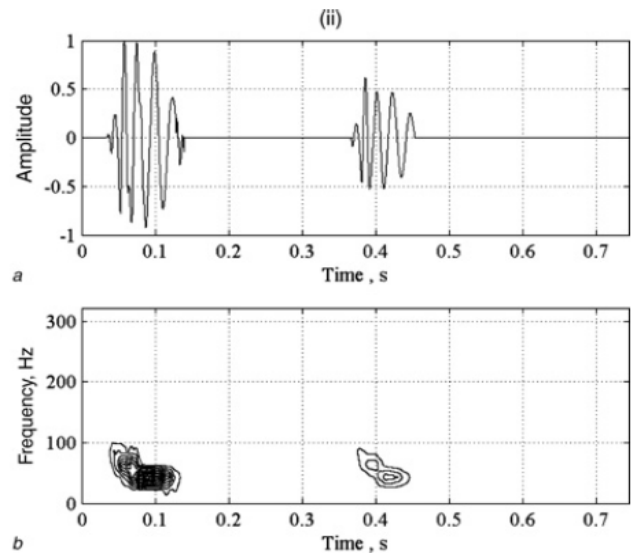


Fig.4 (Time frequency form of multiple murmurs)

Method also expects to improve the result in the time-frequency plane after analysing the segmented signal by using appropriate parameters of the Kernel faction. The method has recently been applied on speech signal [19] and PCG signals [20] and may be useful especially in case of abnormal HS containing various murmurs. The proposed method may be divided into four steps:

Step 1 (time frequency analysis): The first step is to evaluate the time frequency representation of the PCG signal given by the spectrogram using the (3b). It is important to use the best parameter of the window analysing during the analysis based on that equation. To this end, the Re'nyi entropy is used as a measurement of the complexity of the signals, which permits to obtain the optimal window length for the analysis.

Step 2 (estimation of RME): In the second stage, we estimate the RME given by the (7) and use it to find the boundaries between the main components and the murmurs of the PCG signal. He advantage is that the value of this statistical measure is high for murmurs considered as random event and relatively low for the main components considered as quasi-organized event [18].

Step 3 (threshold): The third step consists of the detection of the ending point of the murmur and the starting point of S1 and S2 component based on a threshold of the RME profile.

Once a threshold has been determined, samples of the RME with values below the threshold (corresponding to sample of the signal in time representation) are considered samples of the main components, and everything above the threshold is sample of the murmur. The result provides the segment tin of the signal and permits the characterization of the main components and the murmur both in time and frequency.

Step 4 (enhancement of time-frequency representation): After achieving the segmentation between the main component and the murmur of the PCG signal, the fourth step provides the enhancement of the time-frequency representation. I deed, the use of appropriate window for the main components and the murmur gives best visualization in joint time frequency domain. So, we adopt a narrow window for the murmur considered as random event and wide window for main component considered as quasi-organized event.

III. LEARNING USING SUPPORT VECTOR MACHINES

H. Linear Support Vector Machines

Consider the problem of separating the set of training vectors belonging to two separate classes $(x_1, y_1), \dots, (x_l, y_l)$, where $x_i \in R^n$ a feature vector is and $y_i \in \{-1, +1\}$ a class label, with a hyper plane of equation $w \cdot x + b = 0$. of all the boundaries Determined by w and b , the one that maximizes the margin (Fig.1.Left) would generalize well as opposed to other possible separating hyper planes.

A canonical hyper plane [13] has the constraint for parameters w and b . $\min_x \sum y_i (w \cdot x_i + b) = 1$. A separating hyper plane in canonical form must satisfy the following constraints $y_i [(w \cdot x_i) + b] \geq 1, i = 1, \dots, l$. the margin is $2/\|w\|$ according to its definition. Hence the hyper plane that optimally separates the data is the one that minimizes $\Phi(w) = \frac{1}{2} \|w\|^2$ the solution to the optimization problem can be obtained as follows [13]

$$\bar{a} = \arg \max_a \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{11}$$

First, find the maximization solution to the following problem
Subject to

$$\alpha_i \geq 0 (i = 1, \dots, l), \sum_{i=1}^l \alpha_i y_i = 0 \tag{12}$$

Then calculate

$$\bar{w} = \sum_{i=1}^l \alpha_i y_i x_i, \bar{b} = -\frac{1}{2} w \cdot [x_+ + x_-] \tag{13}$$

Where x_+ is a support vector of the “+” class and x_- is a support vector of the “-” class. Now, a new data point is classified by the sign of

$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + \bar{b} \right) \tag{14}$$

Vapnik [8] is considered as the pioneer in introducing the concept of optimum separating hyper plane of a sample of data in a classification problem, which is the core of the SVM method (SVM).

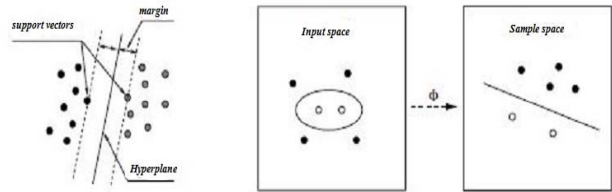


Fig.2 (left-Hyper plane, Right-Input and sample space)

Different historical facts can be highlighted in the development of SVMs.

- 1) The feature space generation from input space by the transformation. By the reverse transformation, the linear boundaries of the separating hyper planes in the feature space result in nonlinear boundaries in the input space. This transformation is called the kernel trick.
- 2) The appearance of soft-margin algorithm for problems where perfect separability is not reachable (problems with noise in the sample data).
- 3) The SVM generalization to regression problems by Vapnik’s ϵ -insensitive loss function [28].

Given a sample of data $\{(x_i, y_i)\}_{i=1}^n$, the SVM problem [28] can be formulated as follows:

$$\min_{w,b,\xi} \frac{1}{2} \left\{ \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi'_i \right\} \tag{15}$$

$$\begin{cases} \langle w, \psi(x_i) \rangle + b - y_i \geq \epsilon + \xi_i \\ y_i - (\langle w, \psi(x_i) \rangle + b) \geq \epsilon + \xi'_i \\ \xi_i, \xi'_i \geq 0 \end{cases}$$

Where ξ_i, ξ'_i are slack variables that ensure that the solution does not contain, within the band of radius ϵ , all the points (x_i, y_i) of the sample (thus avoiding possible outliers and over fitting), where the parameter ν expresses the importance of the slack variables in each point, and where $\psi: X \rightarrow Z$ is a transformation of the input space into a new space Z usually of larger dimension, where we define an inner product by means of a positive definite function k (kernel)

$$\langle \psi(x), \psi(x') \rangle = \sum_i \psi_i(x) \psi_i(x') = k(x, x') \tag{16}$$

The kernel function can be linear, polynomial, radial basis, or sigmoid. The linear and the polynomial kernel functions are used in problems without high nonlinearity;

however, the segment id and the radial basis kernel functions are indicated for problems with high nonlinearity, as is the case of this research. Therefore, radial basis kernel function has been selected due to its better performance against the sigmoid kernel function. The solution, which can be obtained from the dual problem, is a linear combination of a subset of sample points denominate support vectors (s.v.) and it can be written as follows:

$$w = \sum_{s.v} \beta_i \psi(x_i) \beta \psi \Rightarrow$$

$$f_{w,b}(x) = \sum_{s.v} \beta_i \langle \psi(x_i), \psi(x) \rangle + b = \sum_{s.v} \beta_i k(x_i, x) + b$$

----- (17)

It is possible to introduce a parameter in the SVM regression model (Nu-SVR) in order to control the number of support vectors determined [29]. A tenfold cross validation has been implemented in order to determine the optimal SVM parameters according to the best-fit criterion.

I. Kernel Support Vector Machines

In linearly non-separable but nonlinearly (better) separable case, the SVM replaces the inner product $x \cdot y$ by a kernel function and then constructs an optimal separating hyper plane in the mapped space. According to the Mercer theorem [13], the kernel function implicitly maps the input vectors, via an associated with the kernel, into a high dimensional feature space in which the Mapped data is linearly separable (Fig.1.Right). This provides a way to address the curse of dimensionality [13]. Possible choices of kernel functions include (1) Polynomial $K(x,y) = ((x \cdot y + 1))^d$ where the parameter d is the degree of the polynomial; (2)

Gaussian Radial Basis Function $K(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right)$ where the parameter σ is the width of the Gaussian function. However, Exponential Radial Basis Function (ERBF)

$K(x, y) = \exp\left(-\frac{|x-y|}{2\sigma^2}\right)$ has been empirically observed to perform

better than above three [6]. For a given kernel function, the classifier is given by

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + \bar{b}\right)$$

---(18)

J. Multi-Classes

Classification of multi-classes can be achieved by combining all the two-class SVMs. There are two common schemes for this purpose: one-against-all and the one against- one. We use the later and construct a bottom-up binary tree for classification. By comparison between each pair, one class number is chosen resending the “winner” of the current two classes.

The selected classes (from the lowest level of the binary tree) will come to the upper level for another round of tests. Finally, a unique class label will appear on the top of the tree. Denote the number of classes as c , the SVMs learn $\frac{c(c-1)}{2}$ discrimination functions in the training stage, and

carry out comparisons of $c-1$ time under the fixed binary tree structure. If c is not equal to a power of 2, we can decompose c as $c = 2^{n_1} + 2^{n_2} + \dots + 2^{n_i}$ where $n_1 \geq n_2 \geq \dots \geq n_i$ because any natural number (even or odd) can be decomposed into finite positive integers which are the power of 2. If you take a high capacity set of functions get low training error. But you might “over fit”. If you take a very simple set of models, you have low complexity, but won’t get low training error.

Capacity of hyper planes can be figured out by Vapnik & Chervonenk dimensions and it is also showed as follows Consider hyper planes $(w \cdot x) = 0$ where w is normalized set of points X_+ such that $\min |w \cdot x_i| = 1$.

The set of decision functions $f_w(x) = \text{sign}(w \cdot x)$ defined on X^* such that $|w| \leq A$ has a VC dimension satisfying $h \leq R^2 A^2$ where R is the radius of the smallest sphere around the origin containing x^* [12]. Where x^* is set of Training data points incident on hyper plane. Each class are named as different Valvular Heart diseases in order to obtain the reference of test data’s class.

IV. CONCLUSION

Test data which is identified the particular class can be displayed in graphical user interface with the detail of murmur occurred, that is where the murmur occurred, position and duration of murmur are displayed in GUI.

ACKNOWLEDGMENT

The authors are grateful to express sincere thanks and gratitude to our Professor and HOD, Dr.E.Poovammal, Department of Computer Science and Engineering, SRM University for her encouragement and the facilities that were offered to me for carrying out this project. And we take this opportunity to thank our Director, Dr.C.Muthamizchelvan, Ph.D, Faculty of Engineering and Technology, SRM University for providing us with excellent infrastructure that is required for the development of our project.

REFERENCE

- [1] C. Cortes and V. Vapnik. Support vector networks. Machine Learning, 20:273–297, 1995.
- [2] B. Feiten and S. Günzel. “Automatic indexing of a sound database using self-organizing neural nets”. Computer Music Journal, 18(3):53–65, 1994.
- [3] B. Feiten and T. Ungvary. “Organizing sounds with neural nets”. In Proceedings 1991 International Computer Music Conference, San Francisco, 1991.
- [4] J. Foote. “Content-based retrieval of music and audio”. In C. C. J. Kuo et al., editors, Multimedia Storage and Archiving Systems II, Proc. of SPIE, volume 3229, pages 138–147, 1997.
- [5] Foster, W. Schloss, and A. J. Rockmore. “Towards an intelligent editor of digital audio: Signal processing methods”. Computer Music Journal, 6(1):42–51, 1982.
- [6] S. Gunn. “Support vector machines for classification and regression”. Technical report, ISIS, May 1998.
- [7] S. Z. Li. “Content-based classification and retrieval of audio using the nearest feature line method”. IEEE Transactions on Speech and Audio Processing, September 2000.
- [8] S. Z. Li, K. L. Chan, and C. L. Wang. “The the nearest feature line method in image classification and retrieval”. IEEE Transactions on Pattern Analysis and Machine Intelligence accepted 2000.

- [9] S. Z. Li and J. Lu. "Face recognition using the nearest feature line method". *IEEE Transactions on Neural Networks*, 10(2):439–443, March 1999.
- [10] Z. Liu, J. Huang, Y. Wang, and T. Chen. "Audio feature extraction and analysis for scene classification". In *IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing*, 1997. <http://vision.poly.edu:8080/paper/audiommsp.html>
- [11] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.
- [12] S. Pfeiffer, S. Fischer, and e.W. E. "Automatic audio content analysis". Tech. Rep. No. 96-008, University of Mannheim, Mannheim, Germany, April 1996. <ftp://pi4.informatik.unimannheim>.
- [13] N. Vapnik. *Statistical learning theory*. John Wiley & Sons, New York, 1998.
- [14] E. Wold, T. Blum, D. Keislar, and J. Wheaton. "Content based classification, search and retrieval of audio". *IEEE Multimedia Magazine*, 3(3):27–36, 1996 <http://musclefish.musclefish.com/ieemm96/>.
- [15] Rangayyan, R.M., Lehner, R.J.: 'Phonocardiogram signal analysis: a review', *CRC Crit. Rev. Biomed. Eng.*, 1988, 15, (3), pp. 211–236
- [16] Obaidat, M.S.: 'Phonocardiogram signal analysis: techniques and performance comparison', *J. Med. Eng. Technol.*, 1993, 17, (6), pp. 221–227
- [17] Shamsollahi, M.B., Senhadji, L., Chen, D., Durand, and L.-G.: 'Modified signal dependent time frequency representation for analysis of the simulated first heart sound'. *Proc. 19th Int. Conf. on IEEE/EMBS*, Chicago, IL, USA, 30 October–2 November, 1997, pp. 1313–1315
- [18] Debjais, F., Durand, L.G.: 'Time-frequency analysis of heart murmurs – part II: optimization of time-frequency representation and performances evaluation', *Med. Biol. Eng. Compute.*, 1997, 35, pp. 480–485
- [19] Lee Sang Min Lee, J.J., Kim, Icy., Min, H.K., Hong, S.H.: 'Comparison between short time Fourier and wavelet transform for feature extraction of heart sound'. *IEEE TENCON*, 1999, pp. 1547–1550
- [20] Djebbari, A., Reguig, F.B.: 'Short-time Fourier transforms analysis of the phonocardiogram signal'. *Proc. ICECS*, 2000, pp. 844–847
- [21] Santos, M.A.R., Souza, M.N.: 'Detection of first and second cardiac sounds based on time-frequency analyses. *Proc. 23rd Annual EMBS Int. Conf.*, Istanbul, Turkey, 25–28 October 2001, pp. 1915–1918
- [22] Zhidong, Z., Zhijin, Z., Yuquan, C.: 'Time-frequency Analysis of heart sound based on HHT'. *Proc. Int. Conf. on Communications, Circuits and Systems*, 2005, pp. 926–928
- [23] Wang, P., Kim, Y., Liang, L.H., Soh, C.B.: 'First heart sounds detection for phonocardiogram segmentation'. *Proc. IEEE Eng. Med. Bio. 27th Annual Conf.*, Shanghai, China, 1–4 September 2005, pp. 5519–5522
- [24] Gill, D., Gavrieli, N., Intrator, N.: 'Detection and identification of heart sounds using homomorphism envelope and self-organizing probabilistic model', *Compute. Cardiol.*, 2005, 32, pp. 957–960
- [25] Al-Naami, B., Chebil, J., Torry, J.N.: 'Identification of aortic stenosis disease using discrete wavelet packet analyses, *Compute. Cardiol.*, 2005, 32, pp. 667–670
- [26] MacCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, pp. 115–133, 1943.
- [27] M. Stone, "Cross-validated choice and assessment of statistical predictions," *J. Roy. Statist. Soc. Series B*, vol. 36, pp. 111–147, 1974.
- [28] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. Vapnik, *Support Vector Regression Machines*. Cambridge, MA: MIT Press, 1997.
- [29] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Compute.*, vol. 12, pp. 1207–1245, 2000