

# Least Squares Method-Based Ellipse Model for Heart Sound Classification Research

Shuping Sun, Wei Song, Tingting Huang, Biqiang Zhang, Baojin Liu, Jie Wu, Hao Chen

Nanyang Institute of Technology, Nanyang Henan  
Email: shp\_sun@yeah.net, aamir\_song@163.com

Received: Mar. 4<sup>th</sup>, 2020; accepted: Mar. 20<sup>th</sup>, 2020; published: Mar. 27<sup>th</sup>, 2020

---

## Abstract

Heart sound classification is one of the important methods to diagnose heart disease, and its identification accuracy enables medical staff to provide more accurate treatment for patients. In view of this, in order to improve the recognition accuracy, this study proposes to generate classification boundary curve based on support vector machine (SVM). Considering the shape of classified boundary curve, a least square elliptical model based on classified boundary curve is established, which optimizes the complexity of heart disease diagnosis. Firstly, the classification boundary curve based on support vector machine is generated according to the frequency characteristics of the heart sound signal, and then the least square elliptical model based on the classification boundary curve is established to perform ellipse fitting on the classification boundary curve. To verify the effectiveness of elliptical model in heart sound recognition, an experimental analysis of online heart sound database and clinical heart sounds was carried out. 188 cases of atrial fibrillation, 181 cases of aortic regurgitation, 257 cases of mitral regurgitation, 325 cases of normal sound and 150 heart sounds of pulmonary stenosis were detected, and the classification accuracy was 91.7%, 98.8%, 99.8%, and 98.7%, proving that its classification has high precision.

## Keywords

Heart Sound Signal, Support Vector Machines, Least Squares, Ellipse Mode

---

# 基于最小二乘法椭圆拟合在心音识别中的应用研究

孙树平, 宋伟, 黄婷婷, 张弼强, 刘保进, 吴杰, 陈豪

南阳理工学院, 河南 南阳  
Email: shp\_sun@yeah.net, aamir\_song@163.com

收稿日期: 2020年3月4日; 录用日期: 2020年3月20日; 发布日期: 2020年3月27日

## 摘要

心音识别是诊断心脏疾病的重要方法之一，其识别精度使医护人员为患者提供更加准确的治疗手段。鉴于此，为提高识别精度，本研究提出了基于支持向量机(SVM)对心音信号特征生成分类边界曲线，考虑到分类边界曲线的形状，建立了基于分类边界曲线的最小二乘法椭圆模型，优化了心脏疾病诊断的复杂性。其算法步骤为：首先根据心音信号的频率特征生成基于支持向量机的分类边界曲线，然后建立基于分类边界曲线的最小二乘法椭圆模型对分类边界曲线进行椭圆拟合。为验证在心音识别应用中椭圆模型的有效性，对在线心音数据库和临床心音的实验分析，即对188例房颤、181例主动脉瓣返流、257例二尖瓣返流、325例正常声音和150例肺动脉瓣狭窄心音进行检测，其分类精度分别为91.7%、98.8%、98.4%、99.8%和98.7%，证明其分类具有较高的精度。

## 关键词

心音信号，支持向量机，最小二乘法，椭圆模型

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来心血管疾病的患病率和死亡率均呈现持续增长的趋势，并且心血管疾病死亡人数连续多年位居疾病死亡之首。据《中国心血管报告 2018》[1]显示，2018 年农村和城市因心血管疾病死亡人数占全部因疾病死亡人数的比例分别是 46% (见图 1(A))和 43% (见图 1(B))。同时报告指出我国目前大约有 2.9 亿患者，因此我国心血管病防治工作面临严峻挑战[2]。目前诊断心血管病的方法包括无创式(心电图、胸透 X 射线、心音听诊、超声成像和多普勒技术)和有创式(血管造影术等) [3]。由于心音听诊具有非创伤性、价格低廉、操作简单等优点，因此利用心音诊断心血管病得到广泛的研究[4]。而心音诊断的关键即诊断精度。

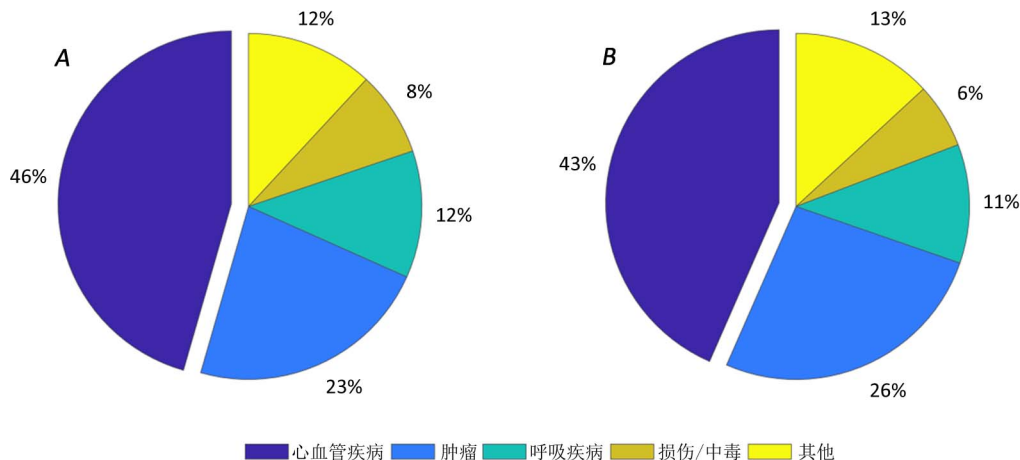


Figure 1. The main cause of death in rural and urban residents in China in 2018

图 1. 2018 年中国农村和城市居民主要疾病死因构成

关于心音诊断,国内外诸多学者广泛采用支持向量机技术对心音信号进行诊断识别[5] [6] [7],然而采用支持向量机生成的分类函数过度依赖于数据点的个数,也就是说,随着学习样本数据量增加,将导致测试样本计算量增大,进而不利于大数据检测。鉴于此,本研究提出一种基于最小二乘法为支持向量机所获取的分类函数建立数学模型进而化简其复杂程度实现对心音信号识别。本研究主要内容包括心音特征提取、基于最小二乘法建立支持向量机椭圆模型、实验分析和结论,其总体结构组织如图2所示。

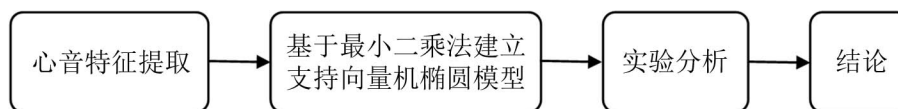


Figure 2. Overall structure flow chart

图2. 总体结构组织流程图

## 2. 心音特征提取

### 2.1. 心音信号的采集与预处理

本研究采用 3M-3200 电子听诊器[8]对心音数据进行采集:根据心音信号的频带范围,采样频率  $F_s$  设置为 44.1 kHz;基于所分析心血管病病理信息有效成分,主动脉瓣区设置为听诊区对心音信号进行采集。电子听诊器采集到的原始心音信号波形如图3(A)所示,其近似为周期信号,每个周期包含两个成分第一心音( $S_1$ )和第二心音( $S_2$ ),同时不可避免包含大量背景噪声。根据心音预处理研究[9]所描述,基于小波变换采用 dB10 作为母小波可为心音信号预处理提供一种有效的途径,对心音信号进行降噪处理,保存 21.5 Hz 和 689 Hz 之间的频率成分,其降噪效果如图3所示,其中图3(A)为滤波之前的原始波形图,图3(B)为滤波后降噪效果图。对比图3(A),从图3(B)中可以明显观察到  $S_1$  和  $S_2$  的有效信息,而  $S_1$  和  $S_2$  之间的噪声得到有效的处理。

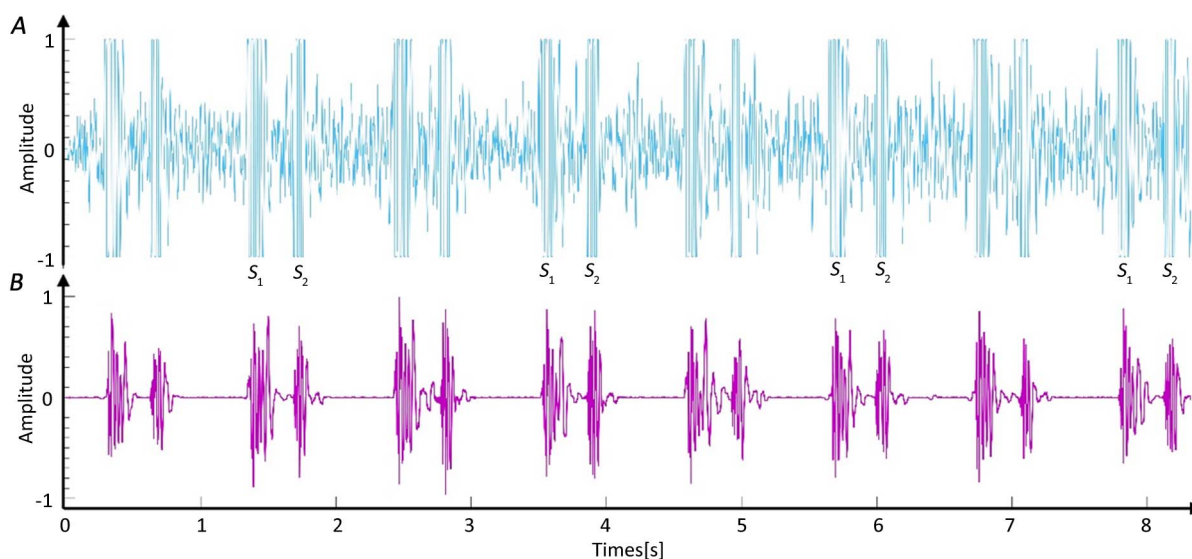


Figure 3. Comparison of heart sound signal before and after processing

图3. 心音信号处理前后对比

### 2.2. 心音信号的特征提取

由研究[9]知,利用  $S_1$  与  $S_2$  时间域距离结合频率域宽度作为特征能够有效区分正常和室间隔缺损心音

信号,其定义方法如公式(1)~(5),特征提取结果如图4所示,(A)为时间域特征 $[T_{12}, T_{11}]$ 定义方法,(B)为从 $E_f$ 中提取的频域特征 $[F_g, F_w]$ ,其计算方法如公式(2)所示。图中 $E_T$ 指从心音 $S_T$ 中提取的包络线, $G_{S1}$ 和 $G_{S2}$ 是第一心音和第二心音的重心, $L_f$ 和 $R_f$ 是包络与阈值( $H_T$ )为0.2时的左右交叉点。

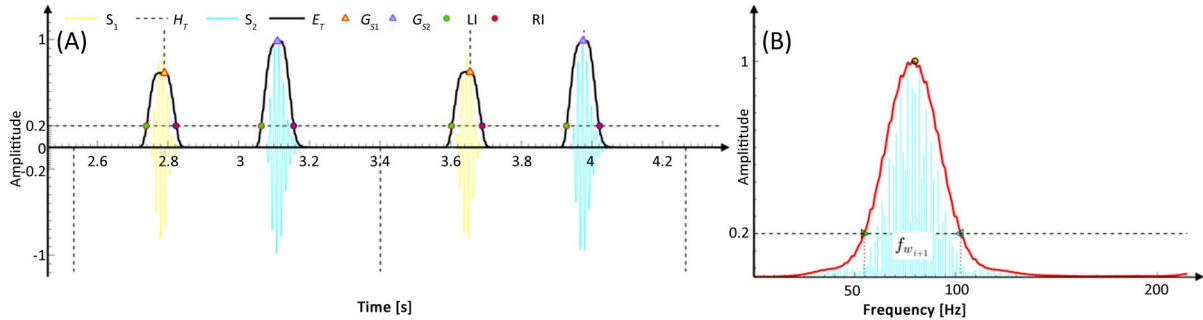


Figure 4. Heart sound signal feature map  
图4. 心音信号特征图

$$\begin{cases} T_{12} = G_{S2} - G_{S1} \\ T_{11} = G_{S2+1} - G_{S1} \end{cases} \quad (1)$$

$$\begin{cases} F_w = R_f - L_f \\ F_g = \frac{\sum_{l=0}^{[N/2]} l \times E_f^2[l]}{\sum_{l=0}^{[N/2]} E_f^2[l]} \end{cases} \quad (2)$$

其中

$$E_f[k] = \frac{1}{2W_f + 1} \sum_{l=k-W_f}^{k+W_f} |S_f[l]| \quad (3)$$

$$\begin{cases} S_f[l] = \sum_{n=1}^{N-1} S_p[n] \exp\left(-j \frac{2\pi}{N} nl\right) \\ l = 0, \dots, N-1 \end{cases} \quad (4)$$

$$w_f = 0.5 \times 16 \times F_s / N \quad (5)$$

其中,  $N$  为心音周期长度,  $L_f$  和  $R_f$  是包络与阈值( $H_T$ )为0.2时的左右交叉点。

为消除特征之间相关性与减小特征维度,本研究采用主成分分析提取前两个主成分作为其心音特征以取代复杂的信息量,其 $[y_1, y_2]$ 的散点图如图5所示。

### 3. 基于最小二乘法建立支持向量机椭圆模型

#### 3.1. 支持向量机

支持向量机(SVM)是由 Vladimir Naumovich Vapnik 开发的一种功能强大、应用广泛的机器学习技术[10],其已成功应用于心音信号分析中[11]。实际上,SVM就是最大限度找到样本数据最优分离超平面(线),关于SVM典型例证如图6所示,其中红线表示由SVM得到的最优函数分类曲线,+与\*表示两类不同样本数据,圈中点即数据的支持向量。

#### 3.2. 分类边界曲线的生成

分类边界曲线生成的主要步骤如下:首先由心音信号生成特征 $[y_1, y_2]$ 的散点图,然后根据高斯函数

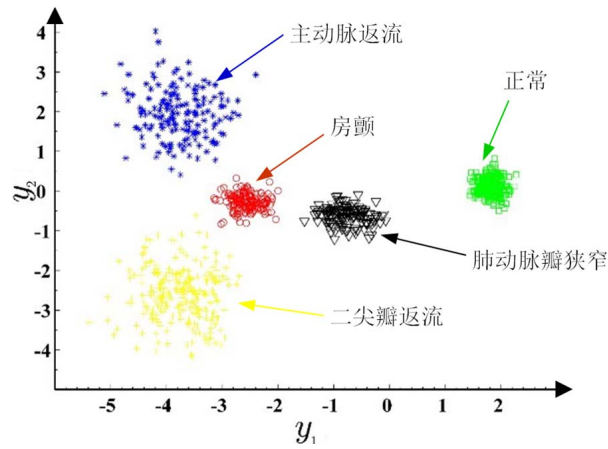


Figure 5. Feature  $[y_1, y_2]$  scatter diagram  
图 5. 特征 $[y_1, y_2]$ 散点图

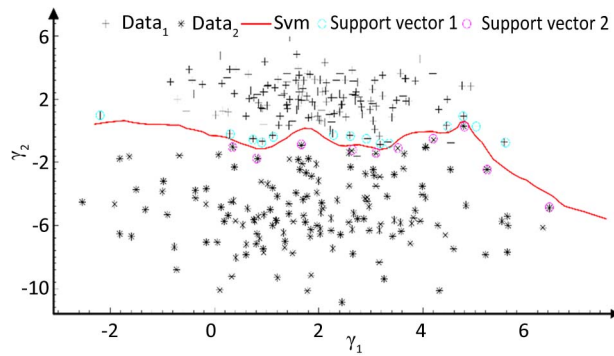


Figure 6. Optimal separation hyperplane  
图 6. 最优分离超平面

作为 SVM 技术的核函数，最后利用高斯核函数对输入向量  $x$  进行边界曲线分类。基于高斯核函数将输入向量  $x$  的分类边界曲线分为两类：正类  $\lambda^{(+)}$  和负类  $\lambda^{(-)}$ 。表示为

$$F(x, \tau) = \sum_{i=1}^N w_i^* y_i e^{\left( \frac{\|x - \lambda_i\|^2}{2\tau^2} \right)} + d^* \quad (6)$$

其中

$$w^* = \arg \max \left[ \sum_{i=1}^N -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j e^{\left( \frac{\|x - \lambda_i\|^2}{2\tau^2} \right)} \right] \quad (7)$$

$$d^* = y_m - \sum_{i=1}^N w_{i=1}^* y_i e^{\left( \frac{\|x - \lambda_i\|^2}{2\tau^2} \right)}, w_m^* \neq 0 \quad (8)$$

$$x_i = \begin{cases} \lambda^{(-)}, & \text{如果 } F(x_i) \leq 0 \\ \lambda^{(+)}, & \text{其他} \end{cases} \quad (9)$$

因此，对于检测样本  $x_i$ ，基于所生成的分类边界曲线的分类结果可以通过方程(6)~(9)确定。根据支持向量机边界曲线生成 SVM 分类边界曲线如图 7(A)所示。虽然 SVM 技术的分类边界曲线具有较高精确

度, 但根据公式(6)可知, 由于分类边界曲线数学表达式过于复杂, 而且计算量大, 为了克服这一问题, 本研究提出了基于最小二乘法建立支持向量法椭圆模型。

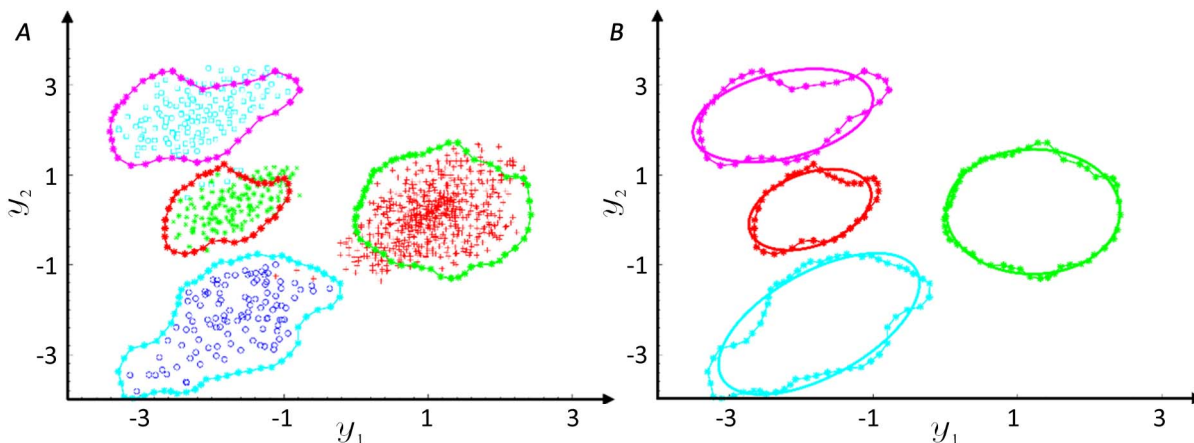


Figure 7. SVM classification boundary curve  
图 7. SVM 分类边界曲线

### 3.3. 基于最小二乘法的椭圆模型算法

为了简化分类边界曲线并使其参数化, 考虑到边界曲线的椭圆模型(EMs), 所以本研究提出了一种基于最小二乘法的椭圆模型。

椭圆一般方程为

$$f(y_1, y_2) = y_1^2 + A_1 y_1 y_2 + A_2 y_2^2 + A_3 y_1 + A_4 y_2 + A_5 \tag{10}$$

用矩阵方法表示为

$$EM(A, [y_1, y_2]) = [y_1^2, y_1 y_2, y_2^2, y_1, y_2, 1] \times [1, A]^T = 0 \tag{11}$$

具有特定椭圆约束

$$A_1^2 - 4A_2 < 0 \tag{12}$$

其中  $A = [A_1, A_2, A_3, A_4, A_5]$  是椭圆 EM 的系数,  $[y_1, y_2]$  是椭圆 EM 上一点坐标。通过公式(13)计算  $A$

$$\min_A \sum_{i=1}^I EM^2(A, [y_1, y_2]_i) \tag{13}$$

其中  $[y_1, y_2]_i (i = 1, 2, \dots, I)$  为分类边界曲线  $F$  上的点。为了解决多变量约束非线性问题的最小化问题, 利用 MATLAB 优化函数 fmincon 进行拟合, 对优化问题进行求解, 并将求解器设置为内点算法。根据以上公式(10)~(12)获得识别心血管疾病心音信号的椭圆模型, 其对应于图 7(B)中绘制的分类边界曲线。基于最小二乘法获得对应于图 7(B)的椭圆模型优化结果绘制于表 1 中, 其中  $EM_{Type}$  表示每种心音信号的椭圆模型。

Table 1. Ellipse model optimization results  
表 1. 椭圆模型优化结果

椭圆模型	对应于图 7 边界曲线的椭圆模型优化结果
	$EM : y_1^2 + A_1 y_1 y_2 + A_2 y_2^2 + A_3 y_1 + A_4 y_2 + A_5 = 0$
$EM_{AF}$	$y_1^2 - 0.097176 y_1 y_2 + 0.79601 y_2^2 + 5.0586 y_1 + 0.15736 y_2 + 5.9105 = 0$



## Continued

$EM_{AR}$	$y_1^2 - 0.066649y_1y_2 + 0.57374y_2^2 + 7.815y_1 - 2.60321y_2 + 15.8407 = 0$
$EM_{NM}$	$y_1^2 - 0.049832y_1y_2 + 0.41581y_2^2 - 3.6507y_1 - 0.005841y_2 + 3.0193 = 0$
$EM_{MR}$	$y_1^2 - 0.304642y_1y_2 + 0.57816y_2^2 + 7.0621y_1 + 1.8328y_2 + 14.3707 = 0$
$EM_{PS}$	$y_1^2 + 0.131112y_1y_2 + 1.26492y_2^2 + 1.5956y_1 + 1.7412y_2 + 0.32955 = 0$

## 4. 实验分析

通过对在线心音数据库和临床心音的实验分析,对 188 例房颤、181 例主动脉瓣返流、257 例二尖瓣返流、325 例正常声音和 150 例肺动脉瓣狭窄心音进行分类检测,通过方程(14)来求出检测结果分类准确率,其中  $TP$  表示真正类,表示  $FP$  假正类,表示  $FN$  假负类,  $TN$  真负类。本研究与其他研究中提出的支持向量机方法、基于主成分分析离散隐含马尔可夫模型(PCA-DHMM)方法[12]和多级基础选择(MLBS)方法[13]进行了比较,结果如表 2,相比其他学者提出的方法准确率有明显提高。

$$CA(\%) = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

**Table 2.** Classification accuracy comparison  
**表 2.** 分类准确率比较

准确度(%)	分类准确率				
	AF	AR	MR	NM	PS
PCA-DHMM	91.1	81.1	92.0	96.6	97.0
MLBS	90.3	90.2	97.8	98.0	91.8
SVM	90.3	83.4	96.6	96.3	83.0
LS-EM	91.7	98.8	98.4	99.8	98.7

## 5. 结论

本研究通过最小二乘法的椭圆拟合算法为支持向量机所获取的分类函数建立椭圆模型。为验证该系统的可行性,本研究对 188 例房颤、181 例主动脉瓣返流、257 例二尖瓣返流、325 例正常声音和 150 例肺动脉瓣狭窄心音进行检测,与之前学者提出的基于 SVM 的分类边界曲线方法、离散隐含马尔可夫模型和多级基础选择方法进行了比较。结果表明相比其他学者提出的方法,本研究提出的椭圆模型在心音诊断时具有计算简单、诊断准确率更高的优点。因此利用电子听诊器或移动设备进行计算机辅助听诊,并结合基于最小二乘法的椭圆拟合算法,可以作为一种智能的、经济的、能够自动检测心血管疾病的诊断方法。

## 参考文献

- [1] 胡盛寿,高润霖,刘力生,等.《中国心血管病报告 2018》概要[J].中国循环杂志,2019,34(3):209-220.
- [2] 许莉莉,等.基于最小二乘支持向量机的心音分类识别研究[J].中国医疗设备,2017,32(4):38-41.
- [3] Sun, S. (2015) An Innovative Intelligent System Based on Automatic Diagnostic Feature Extraction for Diagnosing Heart Diseases. *Knowledge-Based Systems, Elsevier*, **75**, 224-238. <https://doi.org/10.1016/j.knosys.2014.12.001>
- [4] Amiri, A.M. and Armano, G. (2013) Heart Sound Analysis for Diagnosis of Heart Diseases in Newborns. *APCBEE*

- 
- Procedia*, **7**, 109-116. <https://doi.org/10.1016/j.apcbee.2013.08.020>
- [5] Sengur, A. (2012) Support Vector Machine Ensembles for Intelligent Diagnosis of Valvular Heart Disease. *Journal of Medical Systems*, **36**, 2649-2655. <https://doi.org/10.1007/s10916-011-9740-z>
- [6] Lubaib, P. and Muneer, K.V.A. (2016) The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques. *Procedia Technology*, **24**, 1024-1031. <https://doi.org/10.1016/j.protcy.2016.05.225>
- [7] Li, J., Ke, L. and Du, Q. (2019) Classification of Heart Sounds Based on the Wavelet Fractal and Twin Support Vector Machine. *Entropy*, **21**, 472. <https://doi.org/10.3390/e21050472>
- [8] <https://www.littmann.com>
- [9] Sun, S., Wang, H., Chang, Z., *et al.* (2019) On the Mahalanobis Distance Classification Criterion for a Ventricular Septal Defect Diagnosis System. *IEEE Sensors Journal*, **19**, 2665-2674. <https://doi.org/10.1109/JSEN.2018.2882582>
- [10] 林香亮, 袁瑞, 孙玉秋, 王超, 陈长胜. 支持向量机的基本理论和研究进展[J]. 长江大学学报(自科版), 2018, 15(17): 48-53.
- [11] Sun, S., Wang, H. and Jiang, Z. (2014) Segmentation-Based Heart Sound Feature Extraction Combined with Classifier Models for a VSD Diagnosis System. *Expert Systems with Applications*, **41**, 1769-1780. <https://doi.org/10.1016/j.eswa.2013.08.076>
- [12] Sarac, R. (2012) Hidden Markov Model-Based Classification of Heart Valve Disease with PCA for Dimension Reduction. *Engineering Applications of Artificial Intelligence*, **25**, 1523-1528. <https://doi.org/10.1016/j.engappai.2012.07.005>
- [13] Safara, F., Doraisamy, S. and Azman, A. (2013) Multi-Level Basis Selection of Wavelet Packet Decomposition Tree for Heart Sound Classification. *Computers in Biology and Medicine*, **43**, 1407-1414. <https://doi.org/10.1016/j.compbiomed.2013.06.016>