Murmur Separation and Classification from Heart Sound Using Constrained Singular Spectrum Analysis and Wavelet Transform

Yuanyang Qi* and Saeid Sanei†
* Imperial College London, London, UK
E-mail: yuanyang.qi23@imperial.ac.uk Tel/Fax: +44-7422951891
† Imperial College London, London, UK
E-mail: s.sanei@imperial.ac.uk Tel/Fax: +44-7919098002

Abstract—**Detection and classification of heart murmurs are crucial for the early and timely diagnosis and treatment of cardiovascular diseases. However, heart murmurs often overlap with normal heart sounds, making their detection difficult. In this paper, we propose a two-stage method for separating murmurs before their classification. The first stage employs Constrained Singular Spectrum Analysis to exploit the distinct statistical characteristics of heart sounds and murmurs, facilitating their initial separation. The second stage leverages the capability of Wavelet Transform to effectively localize sound components in both time and frequency domains, which is essential for separating subtle murmurs that are not completely removed during the first stage. After these two stages, normal heart sounds are precisely extracted, and the remaining audio components, predominantly murmurs, are effectively classified. Our results demonstrate that pre-separating murmurs enhances the classification performance, achieving an accuracy of over 99% after using transfer learning. The proposed method can improve the clarity and accuracy of heart sound analysis, thereby facilitating better diagnostic evaluations and interventions in clinical settings.**

I. Introduction

Cardiovascular disease is one of the leading threats to human health. Murmurs are very important indicators of such diseases. Therefore, detecting heart murmurs is crucial for the early and timely diagnosis and treatment of cardiovascular conditions [1]. A heart murmur is an abnormal sound heard during the cardiac cycle, typically audible with the aid of a stethoscope, and is often described as a whooshing or swishing noise [2]. This sound results from turbulent blood flow through the heart. Different types of murmurs are associated with various underlying heart diseases [3]. For instance, systolic murmurs, such as those associated with aortic stenosis or mitral regurgitation, occur during ventricular contraction and often indicate abnormalities in valve structure or function. Diastolic murmurs, like those linked to aortic regurgitation or mitral stenosis, manifest during ventricular relaxation and suggest issues such as valve incompetence or narrowing [4]. Fig. 1 illustrates the normal heart sound and different types of murmurs. Often, murmurs are obscured by normal heart sounds and other noises [5]. Extensive experience and training for physicians are needed to identify heart murmurs.

In recent years, various computer-aided technologies have been applied to classify murmurs. In [6] the authors presented a method using wavelet transform and Hilbert phase envelope to classify abnormal heart signal sounds into ten classes depending on the occurrence moment. Many machine learning approaches have also been proposed to detect and classify normal heart sounds and heart murmurs, such as the K-Nearest Neighbour (KNN) algorithm, Support Vector Machines (SVM) algorithm [7], Convolution Neural Network (CNN) [8], Residual Networks (ResNet) [9], and Attention Mechanism [10]. However, they don't attempt the murmur separation method before classification. In [11], [12] the authors have used deep learning methods for source separation. Although the separation results are competitive, their methods require a large data size for training [13].

In this paper, we introduce and employ signal processing and machine learning methods to separate and classify different types of murmurs. Normal heart sounds generally have slow variation and hence a lower Zero Crossing Rate (ZCR) [14] and more peaky distribution, denoted by higher kurtosis [15] compared with murmurs. We use Constrained Singular Spectrum Analysis (CSSA) with ZCR and kurtosis to initially separate the normal heart sound from the murmur. Subsequently, we apply Wavelet Transform (WT) for further extraction of the murmur to achieve high separation precision. Finally, we use machine learning techniques to classify the separated murmurs into their respective types.

II. Methodology

A. Separation of Heart Sound and Murmur

Fig. 2 demonstrates the overall separation process. The method is a two-stage process for the separation of murmurs from original sounds. In the first stage, CSSA is used to initially segregate the normal heart sound from the original sound. This stage can filter out the regular heartbeats, although some murmurs may remain mixed within the extracted sound. The second stage focuses on refining the normal heart sound obtained from the first stage using WT to achieve a high-precision separation of the normal heart sounds.

Singular Spectrum Analysis (SSA) is a non-parametric method particularly advantageous for analyzing non-linear and

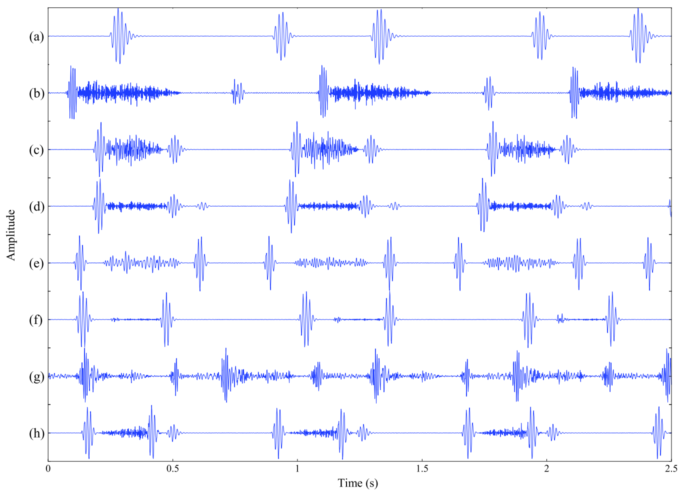

Fig. 1. Normal Heart sound (a) and heart sounds with (b) aortic regurgitation murmur, (c) aortic stenosis murmur, (d) mitral regurgitation murmur, (e) mitral stenosis murmur, (f) mitral valve prolapse murmur, (g) pericardial friction murmur, and (h) pulmonary stenosis murmur.

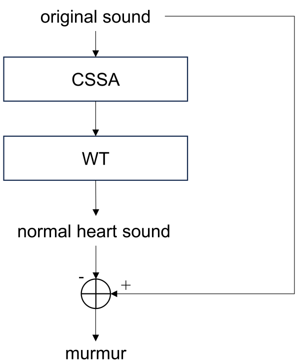

Fig. 2. The overall flowchart for separating heart sound and murmur.

non-stationary time series data. It has many applications such as noise reduction, signal reconstruction, forecasting future trends and biomedical signals [16]. Using SSA, the signal is decomposed into multiple components first, and then the desired signals are reconstructed according to some constraints, referred to as CSSA. The normal heart sounds tend to have lower ZCR and higher kurtosis compared with the murmurs. Therefore, setting ZCR or kurtosis as constraints in the CSSA process is an approach to differentiating these sounds. The above process can be summarized as follows.

After selecting a fixed window length L, the original signal x is transformed from a one-dimensional signal of length N into an $L \times K$ trajectory matrix X, where $K = N - L + 1$.

$$X = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{K-1} \\ x_1 & x_2 & x_4 & \cdots & x_K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L-1} & x_L & x_{L+1} & \cdots & x_{N-1} \end{bmatrix} \tag{1}$$

In the next step, the covariance of the Hankel matrix X, $S = XX^T$, is decomposed into $U\Lambda U^T$ where Λ is a diagonal matrix consisting of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_L$ $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L)$ and $U = [U_1, U_2, \cdots, U_L]$ is the orthonormal eigenvectors of S. Here, $X = \sum_{i=1}^{d} X_i$ where $d = \text{Rank}(X)$. After applying diagonal averaging, the original signal x is converted into a $N \times d$ matrix R, with each component R_i representing a one-dimensional signal of length N.

$$R = \begin{bmatrix} R_0 & R1 & R2 & \cdots & R_{d-1} \end{bmatrix} \tag{2}$$

Next, we apply the constraint along with W, a column vector filled with binary values, to selectively extract the desired component from R.

$$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{d-1} \end{bmatrix}, \; w_i = \{0, 1\} \tag{3}$$

This selection process allows us to reconstruct the final signal x_r, which is obtained by multiplying R by W, denoted as $R \times W$.

$$x_r = R \times W \tag{4}$$

CSSA with ZCR: ZCR measures how frequently the signal changes between positive and negative values. For each component R_i of the signal, we calculate the ZCR. If the computed rate is below or equal to a predefined threshold, then we assign w_i a value of 1; otherwise, w_i is set to 0. CSSA with ZCR enables selective signal reconstruction based on the dynamic characteristics of each component.

$$w_i = \begin{cases} 1 & \text{ZCR}(R_i) \leq \text{threshold}, \\ 0 & \text{ZCR}(R_i) > \text{threshold}. \end{cases} \tag{5}$$

CSSA with Kurtosis: Kurtosis is a statistical measure that describes the distribution's peakedness relative to a normal distribution. In this method, we optimize a vector of parameters W to maximise the kurtosis of $R \times W$. This task constitutes a nonlinear integer programming problem, which can effectively be solved using Genetic Algorithm (GA). Such an algorithm is effective for solving complex optimization challenges involving the nonlinear cost function and constraint. Fig. 3 illustrates the process by which the CSSA with kurtosis separates the heart sounds from murmurs. CSSA with kurtosis enables selective signal reconstruction based on the statistical characteristics of each component represented by its distribution.

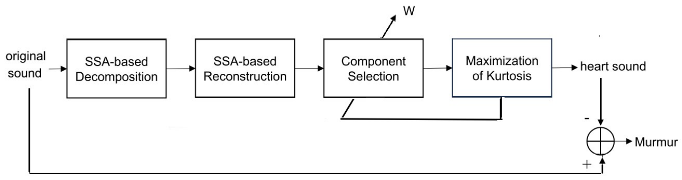

Fig. 3. The process for CSSA with kurtosis.

$$W = \max_{W} \text{ kurtosis}(R \times W) \tag{6}$$

In the first stage of separating heart sound and murmur, we employ CSSA both with ZCR and kurtosis. Each method is utilized to separate the normal heart sound, leaving the murmur as the residual component. Following this, we calculate the correlation between the separated normal heart sound and the murmur. The method—either CSSA with ZCR or CSSA with kurtosis—that yields the lower correlation between these components is then selected as the input to the second stage of the analysis. Fig. 4 shows the flowchart of CSSA in the first stage. This approach ensures that the method leading to the least similarity between the normal sound and the murmur is advanced, enhancing the effectiveness of the separation process.

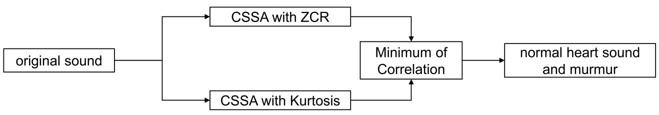

Fig. 4. The flowchart of CSSA in the first stage of separation.

In the second stage of separating heart sound and murmur, we employ WT to meticulously filter out the residual murmur from the heart sound obtained in the first stage. WT is known for its time-frequency analysis capabilities, making it particularly effective for analyzing non-stationary signals, where the signal characteristics change over time [17]. This makes WT an important tool in various applications, such as signal compression, noise reduction, and feature extraction. WT works by decomposing a signal into a set of wavelets, which are functions that can capture both frequency and temporal location information simultaneously. This decomposition allows for a multi-resolution analysis of the signal, enabling the identification of features that may not be apparent in the original time-domain representation. Given that our signal is discrete, we employ the Discrete Wavelet Transform (DWT) to distinguish the residual murmurs from the normal heart sounds identified initially. During the application of DWT, the heart sound signal is decomposed into various frequency components at different scales. The filtering process involves thresholding and reconstructing the signal to ensure that the residual murmur is effectively separated from the normal heart sounds.

After completion of these two stages, the process results in precise extraction of normal heart sounds. The use of CSSA and DWT ensures that both stages complement each other, leading to a more refined separation process. The separated murmurs, are then separated and can be used for classification. This separation is crucial as it allows for a focused analysis of murmur characteristics without the interference of normal heart sounds. This two-stage method enhances the clarity and accuracy of heart sound analysis, facilitating better diagnostic evaluations and interventions in clinical settings.

B. Classification of Murmur

Once the murmur has been separated, we proceed to extract its features and employ machine learning for classification. For this purpose, we utilize four distinct classification methods: SVM [18], Random Forest (RF) [19], CNN [20] and transfer learning [21].

SVM is a supervised machine learning algorithm and its goal is to find an optimal hyperplane that can best separate different classes by maximizing the margin between the closest data points of each class, known as support vectors. RF is a bagging method within ensemble learning and can enhance the accuracy and stability of predictions by constructing multiple decision trees and aggregating their outputs by majority voting for classification. By training each tree on different subsets of data and features, Random Forest enhances model robustness and accuracy and reduces overfitting. For SVM and RF, we use Mel Frequency Cepstral Coefficients (MFCCs) as the feature input.

CNN is a deep learning algorithm and our proposed architecture is shown in Fig. 5. This CNN comprises two convolutional layers and two max pooling layers which help in extracting and downsampling spatial hierarchies of features, two fully connected layers that synthesize these features into predictions, and three dropout layers strategically placed to prevent overfitting by randomly omitting a subset of features during training. For CNN, we use the Mel Spectrogram as the feature input. The Mel Spectrogram is considered as a coloured image which has been resized to 224×224 for three colour channels of red, blue and green.

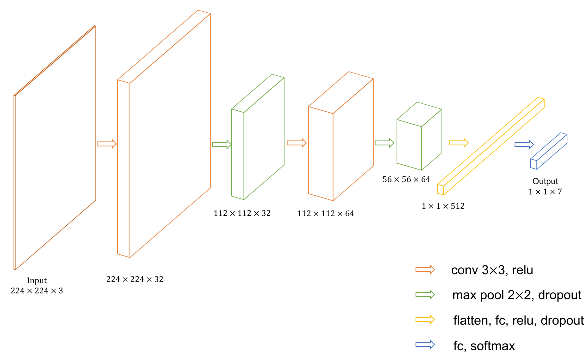

Fig. 5. The architecture of the designed CNN.

Transfer learning aims to learn from diverse and comprehensive data and then apply the learned knowledge to a new task. We use Wav2cec 2.0 [22] as the pre-trained model which is trained on a large amount of audio data. Wav2Vec 2.0 is an end-to-end self-supervised learning model that directly takes raw audio waveforms as input [23]. After pretraining, the model can be fine-tuned to perform classification directly. Fig. 6 shows the architecture of transfer learning using Wav2vec 2.0 as the pre-trained model.

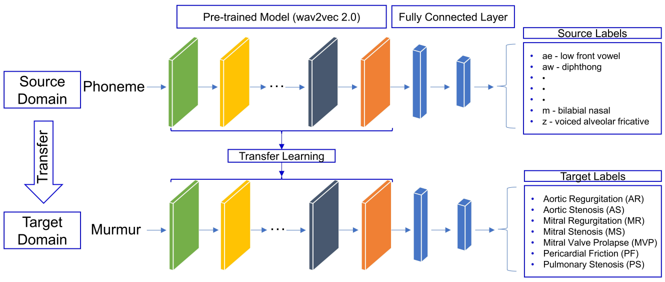

Fig. 6. The architecture of transfer learning using Wav2vec 2.0 as the pre-trained model.

III. Experiment and Results

Data: We utilized the publicly available dataset from [24] that includes four classes of abnormalities: Aortic Stenosis, Mitral Regurgitation, Mitral Stenosis, and Mitral Valve Prolapse. Additionally, we expanded the data by collecting data for three more abnormal classes: Aortic Regurgitation, Pericardial Friction, and Pulmonary Stenosis from online open sources, such as Google.

Separation: Fig. 7 shows the separation results of seven types of murmurs using CSSA and WT.

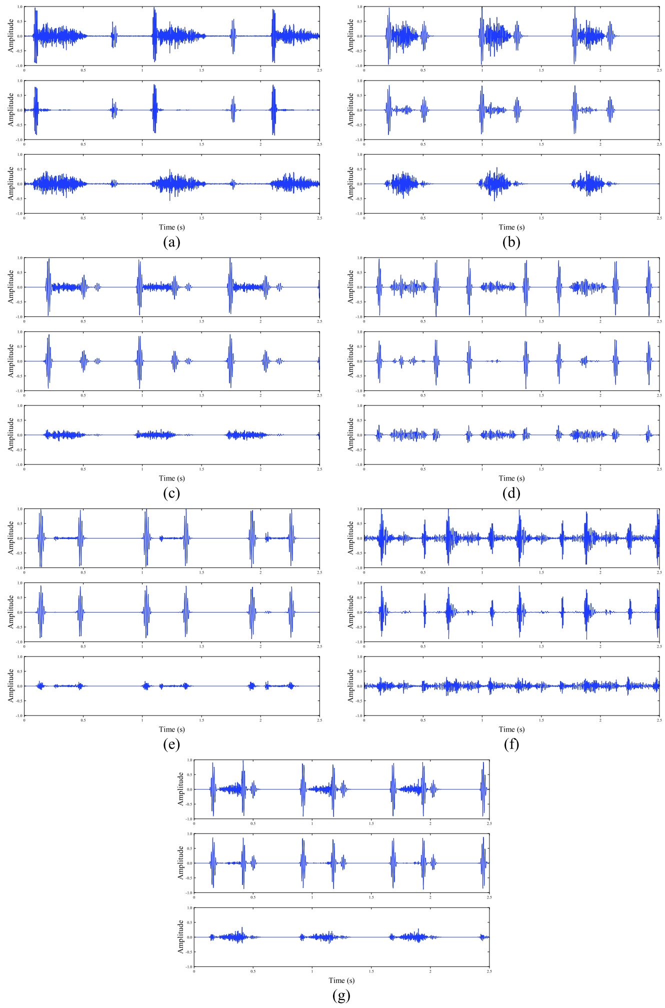

Fig. 7. Separation of seven types of murmurs from the original sounds: the top signal is the original sound, the middle signal is the separated normal heart sound, and the bottom signal is the separated murmur; (a) aortic regurgitation murmur, (b) aortic stenosis murmur, (c) mitral regurgitation murmur, (d) mitral stenosis murmur, (e) mitral valve prolapse murmur, (f) pericardial friction murmur, and (g) pulmonary stenosis murmur.

Classification: Accuracy, F1 Score, Sensitivity, and Specificity are widely used to evaluate the performance of classification models, providing a comprehensive assessment of a model's predictive accuracy and precision in handling both positive and negative classifications.

Accuracy is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

where TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives.

F1 Score is the harmonic mean of Precision and Recall, calculated as:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

where Precision is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

and Recall (also known as Sensitivity) is calculated as:

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN} \tag{10}$$

Specificity is calculated as:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{11}$$

Together, these metrics provide a detailed understanding of a model's performance, allowing evaluation and fine-tuning of models for optimal predictive accuracy and precision in classification tasks.

Tables I and II respectively present the results for two different cases: (i) classification using the original, unseparated sound, and (ii) classification of murmur signals following their separation using CSSA and WT. The results from these tables clearly demonstrate that the pre-separation of murmurs significantly enhances the performance metrics, achieving higher Accuracy, F1 Score, Sensitivity, and Specificity.

TABLE I
ACCURACY, F1 SCORE, SENSITIVITY, AND SPECIFICITY METRICS FOR THE CLASSIFICATION OF SEVEN MURMUR TYPES (AS, AR, MR, MS, MVP, PF, PS) USING SVM, RF, CNN, AND TRANSFER LEARNING; THE ORIGINAL SOUND HAS BEEN USED HERE.

	SVM	RF	CNN	Transfer Learning
Accuracy	91.27%	94.76%	97.38%	99.13%
F1 Score	92.03%	94.68%	97.33%	98.97%
Sensitivity	91.56%	94.20%	97.47%	98.86%
Specificity	98.50%	99.10%	99.56%	99.85%

TABLE II
ACCURACY, F1 SCORE, SENSITIVITY, AND SPECIFICITY METRICS FOR THE CLASSIFICATION OF SEVEN MURMUR TYPES (AS, AR, MR, MS, MVP, PF, PS) USING SVM, RF, AND TRANSFER LEARNING; THE SEPARATED MURMURS USING CSSA AND WT HAS BEEN USED HERE.

	SVM	RF	CNN	Transfer Learning
Accuracy	94.76%	96.94%	98.69%	99.56%
F1 Score	94.88%	97.42%	98.73%	99.49%
Sensitivity	94.53%	97.29%	98.93%	99.47%
Specificity	99.10%	99.49%	99.78%	99.92%

IV. CONCLUSIONS

In this study, we introduced a novel two-stage method to separate murmurs from heart sounds before applying murmur classification. The initial stage involves utilizing CSSA to decompose the heart sound signal, followed by reconstructing the signal using ZCR and kurtosis as constraints. Subsequently, WT is applied in the second stage to further extract the murmur. This two-stage approach is designed to maximize the precision of murmur separation, thereby improving the performance of subsequent classification tasks. Our experimental results have been promising, showing that pre-separating murmurs can enhance classification performance, with accuracy levels exceeding 99% after using transfer learning. This methodological advancement has the potential to greatly improve the quality and effectiveness of medical care, enabling more accurate and reliable diagnosis of heart conditions.

REFERENCES

[1] C. Xu, X. Li, X. Zhang, *et al.*, "Cardiac murmur grading and risk analysis of cardiac diseases based on adaptable heterogeneous-modality multi-task learning," *Health Information Science and Systems*, vol. 12, no. 1, p. 2, 2023.

[2] M. M. Milani, P. E. Abas, and L. C. De Silva, "A critical review of heart sound signal segmentation algorithms," *Smart Health*, vol. 24, p. 100 283, 2022.

[3] J. Prince, J. Maidens, S. Kieu, *et al.*, "Deep learning algorithms to detect murmurs associated with structural heart disease," *Journal of the American Heart Association*, vol. 12, no. 20, e030377, 2023.

[4] C. J. Riley, "Assessment of the heart," *History and Physical for the Pediatric Dental Patient: Establishing a Systematic Approach for Procedural Sedation*, pp. 43–60, 2024.

[5] A. F. Gündüz and A. Karci, "Heart sound classification for murmur abnormality detection using an ensemble approach based on traditional classifiers and feature sets," *Computer Science*, vol. 5, no. 1, pp. 1–13, 2022.

[6] V. Nivitha Varghees and K. I. Ramachandran, "Heart murmur detection and classification using wavelet transform and hilbert phase envelope," in *2015 Twenty First National Conference on Communications (NCC)*, 2015, pp. 1–6.

[7] A. D. Levin, A. Ragazzi, S. L. Szot, and T. Ning, "A machine learning approach to heart murmur detection and classification," in *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2020, pp. 521–525.

[8] S. Sudarsanan and J. Aravinth, "Classification of heart murmur using cnn," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2020, pp. 818–822.

[9] B. Walker, F. Krones, I. Kiskin, G. Parsons, T. Lyons, and A. Mahdi, "Dual bayesian resnet: A deep learning approach to heart murmur detection," in *2022 Computing in Cardiology (CinC)*, vol. 498, 2022, pp. 1–4.

[10] V. Vakamullu, M. Mishra, and A. Mukherjee, "Real-time heart murmur classification using attention based deep learning approach," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2022, pp. 1–6.

[11] P. Chandna, M. Miron, J. Janer, and E. Gómez, "Monoaural audio source separation using deep convolutional neural networks," in *Latent Variable Analysis and Signal Separation: 13th International Conference, LVA/ICA 2017, Grenoble, France, February 21-23, 2017, Proceedings 13*, Springer, 2017, pp. 258–266.

[12] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Joint optimization of masks and deep recurrent neural networks for monaural source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2136–2147, 2015.

[13] K.-H. Tsai, W.-C. Wang, C.-H. Cheng, *et al.*, "Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 3203–3214, 2020.

[14] H. Tang, M. Wang, Y. Hu, B. Guo, and T. Li, "Automated signal quality assessment for heart sound signal by novel features and evaluation in open public datasets," *BioMed Research International*, vol. 2021, no. 1, p. 7 565 398, 2021.

[15] S. Sanei, M. Ghodsi, and H. Hassani, "An adaptive singular spectrum analysis approach to murmur detection from heart sounds," *Medical engineering & physics*, vol. 33, no. 3, pp. 362–367, 2011.

[16] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals*. CRC press, 2015.

[17] C. S. Burrus, R. A. Gopinath, and H. Guo, "Wavelets and wavelet transforms," *rice university, houston edition*, vol. 98, 1998.

[18] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[19] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[20] Y. LeCun, B. Boser, J. S. Denker, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[21] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, 2010, pp. 242–264.

[22] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems*, vol. 33, pp. 12 449–12 460, 2020.

[23] D. Krishna, P. Wang, and B. Bozza, "Using large self-supervised models for low-resource speech recognition.," in *Interspeech*, 2021, pp. 2436–2440.

[24] Yaseen, G.-Y. Son, and S. Kwon, "Classification of Heart Sound Signal Using Multiple Features," *Applied Sciences*, vol. 8, no. 12, p. 2344, 2018.