A Lightweight 1-D Convolution Neural Network Model for Multi-class Classification of Heart Sounds

Puneet Kumar Jain
Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, Odisha- 769008
Email: jainp@nitrkl.ac.in

Ravi Raj Choudhary
Computer Science
Central University of Rajasthan
Rajasthan, India- 342011
Email: raviraj@curaj.ac.in

Mamta Rani Singh
Computer Science
Central University of Rajasthan
Rajasthan, India- 342011
Email: mamtarsingh@curaj.ac.in

***Abstract*—Analysis of heart sound signals provides ample features to diagnose cardiovascular diseases (CVDs) at an early stage. However, the role of cardiac auscultation is limited only to performing the preliminary screening. It is due to the subjectivity of nature in the diagnosis. Automatic analysis of heart sounds will address this issue as well as it will also reduce the burden of the already-stretched medical facility. This paper proposes a lightweight 1D-CNN model to analyse and classify heart sound signals into five categories. The CNN model is trained on the multi-resolution domain features obtained using the discrete wavelet transform (DWT). The signal is first pre-processed and then decomposed up to five levels using 'coif5' as the mother wavelet. The obtained detailed level and approximation level coefficients are applied to the 1-D CNN model. The proposed method yields 98.9% accuracy, 99.01% sensitivity, and 99.72% specificity, showing the proposed method's superiority on various methods proposed in the literature recently.**

***Index Terms*—Heart sound, Phonocardiography, Convolution neural network, Computer-aided diagnosis, Discrete wavelet transform.**

I. Introduction

Heart sound signals are generated due to the heart valves closing and opening, which occur during a cardiac cycle. Thus, the heart sound signal provides ample diagnostic features for various CVDs including valvular diseases, certain arrhythmias, and ventricular septal defect [1]. In general, medical fraternities uses a stethoscope to listen and analyse the heart sound for the preliminary screening of the health status of the heart as well as lungs. This process is called cardiac auscultation, which is very popular due to easy to operate and timeless process [2], [3]. However, cardiac auscultation is subjective in nature, and the diagnosis result largely depends on the experience and hearing ability of the physician [4]. Moreover, the low number of cardiac experts limits its availability to a remote location. Therefore, the need for automatic analysis of heart sound signals is eminent. To address this requirement, this paper presents a novel 1-D CNN model for automatic analysis of heart sounds. Moreover, classification of the signal with a specific diseases is a challange task due to the non-stationary nature of the signal and overlapping of murmur with fundamental heart sounds in time and frequency-domain both. The proposed work classify the signal into five categories.

In the literature, various methods have been proposed to classify heart sound based on classical machine learning techniques as well as deep learning techniques [1], [5]. In the classical machine learning approach, first, the features are extracted from the time, frequency, and time-frequency domain and then the extracted features are applied to train the model. N mei et al. [6] extracted the features from the time-frequency domain obtained using the wavelet scattering transform technique and then classified the signal as normal vs abnormal using a support vector machine (SVM). Q Mubarak et al. [7] extracted time-domain features including mean and standard deviation ratio between heart sound and cardiac cycle, kurtosis, fractal dimension, Hjorth parameter. These features combined with DWT coefficients are applied to the SVM model to identify the fundamental heart sounds, S1 and S2. In [8], SVM and K nearest neighbour (KNN) models were proposed with the one-dimensional binary patterns for multi-class classification. S K Gosh et al. [9] used the chirplet transform for the feature extraction and multi-class composite classifier for the identification of disease. Other classical machine learning techniques, including decision tree [10], random forest [11], and discriminant analysis [12] also have been used. Among these methods, SVM has produced promising results due to its capability to transform the feature domain to a non-linear sparse domain with the help of optimum non-linear kernel function [13]. However, the classical machine learning techniques significantly varies according to the selected features. Moreover, the classical methods have low generalisation capability than the recently emerged deep learning techniques.

Deep learning-based techniques extract the relevant features automatically and hence are very handy to classify the heart sounds [14]. Recently, S L Oh et al. [15] used the wavelet-based deep learning method called deep wavenet, and S B Shuvo [16] used a lightweight deep learning method called CardioXNet to classify the signals into five categories and achieved 98.2% accuracy. In [17], a combination of CNN and Bi-LSTM has been proposed, which produced 99.3% accuracy for five class classification. However, the method performed k-fold cross-validation instead of a separate train and test dataset. Moreover, the model does not exploit the multi-domain analysis due to which a complex model was developed. N Baghel et al. [18] proposed a CNN model and achieved 98.6% accuracy for five class classification. In [19], features were extracted from the variable-Q transform, hybrid constant-Q transform, and Mel-frequency cepstral coefficients to train the five-layer CNN model. However, most of the deep learning methods have been used to perform two-class (normal vs abnormal) classification [20], [21], and to identify the fundamental heart sound (FHS), S1 and S2, to segment the heart sound [22], [23].

In this paper, a lightweight 1-D CNN model for multi-class classification of heart sounds has been presented. The model is trained using the multi-resolution domain data obtained using the DWT. As per the literature survey, this is the first attempt to introduce the multi-resolution domain to train the CNN model. DWT decomposes the signal into approximation and various detailed level coefficients at different scales and hence provides a multi-resolution analysis of the data [24]. The multi-resolution analysis emphasises the time-frequency features of a signal. In the presence of pathology, the frequency range of FHS and extra sounds called murmur changes and, therefore, time-frequency features help to discriminate the pathological cases. The proposed model is applied on a publicly available dataset [25]. The dataset contains 1000 samples of five categories. In this dataset, we observe that the signal varies in length due to variations in the heartbeat. To overcome this issue, the signal is resized to equal length after recognising the onset and offset of the signal.

The rest of the paper is organised as follows. Section 2 presents each step of the proposed method in detail. Section 3 provides the results of the proposed method and its comparison with the state-of-art methods. At last, the conclusion and future works are discussed in Section 4.

II. The Proposed Method

The proposed method classifies the heart sound into five classes. It performs the classification in three steps as shown in Fig. 1. First, the signal is pre-processed to obtain an equal length and normalised signal. Then the signal is decomposed into five detailed levels and one approximation level coefficient. These coefficients are clubbed in a 1-D array to form the input for the CNN model. Finally, the CNN model is trained and tested on the dataset.

A. Pre-processing

The dataset used in this study is consists of five classes of heart sound signal with the sampling frequency of each signal is 8 kHz. Following are the data pre-processing steps applied on each signal:

Re-sampling: Since the frequency range of the FHS and various pathological sound lie below 500 Hz [5], the signal is down-sampled from 8 kHz to 1 kHz sampling frequency.

Normalization: To suppress the amplitude variation due to inter-class variation on the amplitude of the heart signal, the filtered signal is normalized as follow:

$$x_{norm}(n) = \frac{x\left(n\right)}{\max\left(|x|\right)} \tag{1}$$

Resizing: In the dataset, the length of the signal varies from 1.15 to 3.99 seconds. After the observation, it is found that each sample consists of approximately three cardiac cycles. However, due to variations in the heartbeat, the length of the signal varies. To address this issue, the signal was resized to an equal length (2800 samples) after recognising the onset and offset of the signal. The resizing of the signal is performed using the 'imresize' method of the Matlab®(version R2020, MathWorks, USA), which uses bicubic interpolation. The original signal and impact of the resizing is shown in Fig. 2.

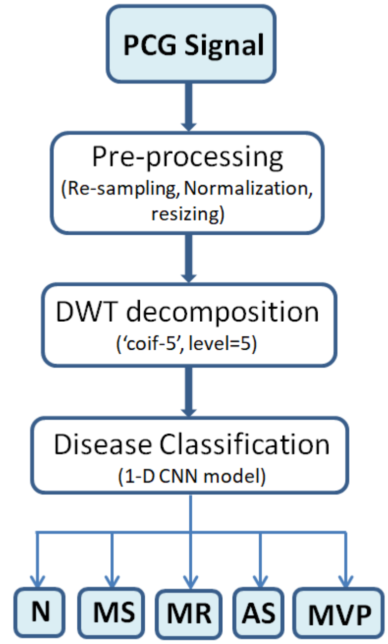

Fig. 1. Block diagram of the proposed method

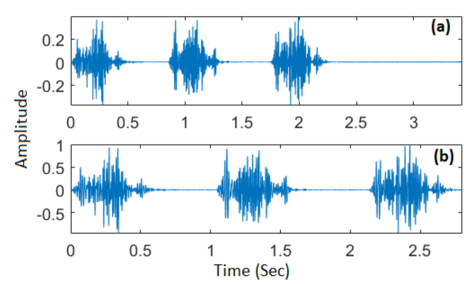

Fig. 2. Pre-processing of the signal (a) original signal, (b) resized and normalized signal

B. DWT decomposition

Mallat provided a fast approach to perform the DWT and called it a sub-band coding algorithm [26]. In this approach, the signal is convolved with two filters, low-pass (H) and high-pass (G), called analysis filters to produce the approximation (A) and detailed (D) level coefficients, respectively. For the first level ($j = 1$), the signal $x(n)$ itself will be convolved with both filters. Then the next level coefficients are obtained by applying the filters on the down-sampled approximation coefficients obtained from the previous level, as shown in Fig. 3. The detailed and approximation coefficients at particular level (j) can be obtained as follows:

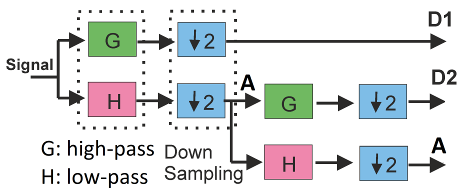

Fig. 3. Decomposition steps according to Mallat's subband coding algorithm

$$A_j(k) = \sum_n A_{j-1}(n)H(n-2k) \tag{2}$$

and

$$D_j(k) = \sum_n A_{j-1}(n)G(n-2k) \tag{3}$$

In the proposed method, the heart sound signal is decomposed up to 5 levels using 'coif5' as mother wavelet. 'coif5' wavelet is selected due to its good analytical performance for the heart sound signal [27]. Since the sampling frequency of the signal is set to 1 kHz, the frequency band of the Detail-1, Detail-2, Detail-3, Detail-4, Detail-5, and approximation level will be 250-500, 125-250, 62.5-125, 31.25-62.5, 15.56-31.25, and 0-15.56 Hz, approximately [28]. The obtained five detailed level coefficients and the approximation level signal for the input signal is shown in Fig. 4.

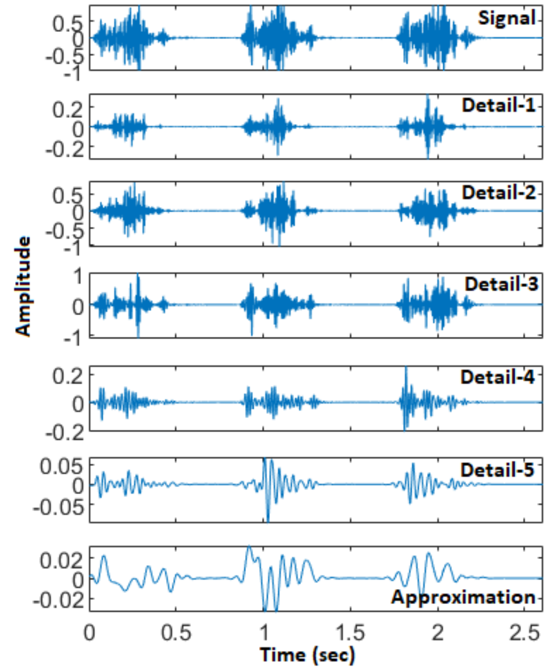

Fig. 4. Obtained detailed and approximation level signal using DWT decomposition

TABLE I
THE ARCHITECTURE OF PROPOSED CNN MODEL

Operation Layer	Size of filter	Number of filters
Input	-	-
Conv1	[1, 5]	16
Pool1	[2, 16]	1
Conv2	[1, 5]	8
Pool2	[2, 8]	1
FC	[1, 5]	1
Softmax	1	-

C. Disease classification: 1-D CNN model

The obtained detailed and approximation level coefficients using DWT are arranged in 1-D array which results in an array of length 2942. This array was feed to train the 1-D CNN model for which the architecture is described as follow.

As shown in the Table I, the proposed CNN model is consist of 5 layers, 1 input layer, 2 convolution and pooling layers, 1 fully connected (FC) layer and 1 output layer (softmax). In each convolution and pooling layer padding is used to produce the output of same size as input. Number of nurons at the input layer is 2942, which is equal to the number of coefficients obtained using the DWT decomposition. The proposed model was trained for 50 epochs and nine iterations in each epoch, resulting in 450 total iterations with a learning rate of 0.01. A mini-batch (batch size:64) stochastic gradient descent with momentum is used to optimise the model parameters.

III. Results and Discussion

The experiments were performed on a publicly available dataset consisting of 1000 samples, 200 samples each for five categories, including the aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS), mitral valve prolapse (MVP), and normal (N) [25]. The sampling frequency of each sample is set to 1 kHz and a constant length of 2800 samples. The complete dataset was randomly split into train (70%) and test (30%) datasets. All the experiments are conducted using the Matlab® (version R2020, MathWorks USA) software on a desktop computer equipped with a Core-i9(10 cores) 64-bit processor and 32-GB RAM.

For the quantitative performance analysis, sensitivity, specificity, precision, recall, and F-score metrics have been used [29]. Moreover, the confusion matrix is also calculated to show the specific number of classification of each input class to the output class. Fig. 5 shows the confusion matrix obtained using the proposed model on the test dataset. The figure shows that the proposed model efficiently classifies all the categories.

Table II shows the sensitivity, specificity, precision, recall, and F-score for the proposed model. All five classes are classified with an F-score of above 98.18%. For the classes MR and N, the F-score is more than 99%. In addition to F-

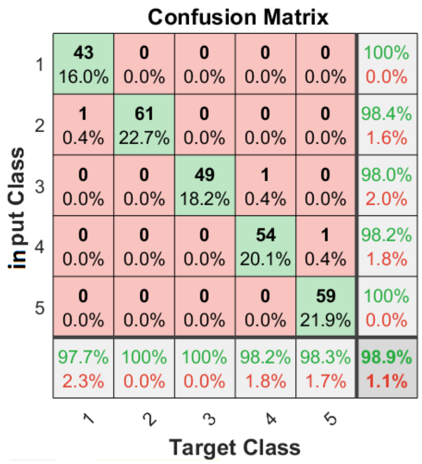

Fig. 5. Confusion matrix obtained using the proposed method

score, all four metrics are higher than 98% except the precision of As class which is 97.73%. It can also be observed that for all five categories, a high sensitivity (>98%) with high specificity(>99%) have been achieved.

TABLE II
OBTAINED PERFORMANCE EVALUATION METRICS USING THE PROPOSED METHOD

Diseases Class	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-Score (%)
AS	1.0000	0.9956	0.9773	1.0000	0.9885
MR	0.9839	1.0000	1.0000	0.9839	0.9919
MS	0.9800	1.0000	1.0000	0.9800	0.9899
MVP	0.9818	0.9953	0.9818	0.9818	0.9818
N	1.0000	0.9952	0.9833	1.0000	0.9916

Fig. 6 shows the accuracy and loss with respect to the number of epochs during the training and validation. It is apparent that validation accuracy has increased rapidly from epoch 1 to 10 and then gradual improvement till the last epoch. Similarly, the loss reduces drastically from epoch 1 to 10. The figure also clearly shows no over-fitting in the trained model since the gap between accuracy and loss during the training and validation is marginal.

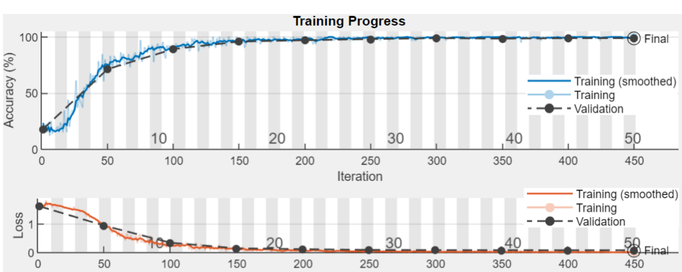

Fig. 6. Accuracy and loss with respect to the number of epochs during the training and validation

Comparison with other methods: Table III shows the sensitivity, specificity, and overall accuracy(OAccuracy) for the proposed method and recently proposed methods in the literature. The highest accuracy (98.9%) is achieved using the proposed method. Moreover, the sensitivity and specificity are significantly higher for the proposed method. It shows the superior performance of the proposed method over other methods. There are two significant reasons for the outstanding results. First, the resizing of the signal overcome the variation in the frequency of components due to variation in heartbeats. Second, the inclusion of multi-resolution analysis emphasized the time-frequency features of various pathological cases.

TABLE III
OBTAINED PERFORMANCE EVALUATION METRICS USING THE PROPOSED METHOD AND THE VARIOUS METHOD PROPOSED IN THE LITERATURE RECENTLY.

Authors, year	Subject type	Sensitivity	Specificity	OAccuracy
Yaseen et al. 2018 [30]	AS	99.00	98.25	97.6
	MR	94.00	99.88	
	MS	97.50	99.50	
	MVP	99.00	99.75	
	N	98.50	99.62	
S.K. Ghosh et al. 2019 [11]	AS	96.77 (Accuracy)		95.13
	MR	90.55 (accuracy)		
	MS	89.77 (accuracy)		
	N	98.55 (accuracy)		
Shu Lih Oh et al. 2020 [15]	AS	94.50	98.50	97.0
	MR	89.00	97.87	
	MS	96.50	98.12	
	MVP	88.50	96.87	
	N	94.00	99.25	
S.K. Ghosh et al. 2020 [9]	AS	99.66	99.04	98.54
	MR	96.33	99.49	
	MS	98.83	99.26	
	N	98.49	99.94	
Proposed method	AS	100.0	99.56	98.90
	MR	98.39	100.0	
	MS	98.00	100.0	
	MVP	98.18	99.53	
	N	100.0	99.52	

IV. CONCLUSIONS AND FUTURE WORKS

This paper proposes a lightweight 1-D CNN model to classify heart sound into five categories, AS, MR, MS, MVP, and N. The CNN model is trained using the multi-resolution analysis obtained using the DWT. The proposed method achieved an overall accuracy of 98.9% with high sensitivity and specificity. It shows that the proposed method can effectively classify heart sounds. Two reasons are observed for the efficacy of the proposed method. First, the resizing of the signal overcome the variation in the frequency of components due to variation in heartbeats. Second, the inclusion of multi-resolution analysis emphasized the time-frequency features of various pathological cases. A system equipped with the proposed method will be helpful to perform computer-aided diagnosis by the user without the intervention of a medical expert. Thus, it will reduce the burden of already stretched medical facilities. Moreover, with the help of such a system, users can perform frequent check-ups, leading to early-stage diagnosis of the disease. The proposed model can be improvised further in manifolds. First, the inclusion of the noisy signal is crucial to establish the applicability of the system in daily-life scenario. Second, a large number of pathological cases have to be incorporated. Moreover, the large size of dataset will improve the generalization capability of the model. Third, various other translation invariant time-frequency transformation techniques has to be explored to emphasize the features.

REFERENCES

[1] P. K. Jain and A. K. Tiwari, "Heart monitoring systems—a review," *Computers in Biology and Medicine*, vol. 54, no. 0, pp. 1–13, 2014.

[2] S. Leng, R. Tan, K. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *BioMedical Engineering OnLine*, vol. 14, no. 1, 2015.

[3] P. K. Jain and A. K. Tiwari, "A robust algorithm for segmentation of phonocardiography signal using tunable quality wavelet transform," *Journal of Medical and Biological Engineering*, vol. 38, no. 3, pp. 396–410, 2017.

[4] T. R. Reed, N. E. Reed, and P. Fritzson, "Heart sound analysis for symptom detection and computer-aided diagnosis," *Simulation Modelling Practice and Theory*, vol. 12, no. 2, pp. 129 – 146, 2004, advances in modelling and simulation in biology and medicine.

[5] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds–a systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.

[6] N. Mei, H. Wang, Y. Zhang, F. Liu, X. Jiang, and S. Wei, "Classification of heart sounds based on quality assessment and wavelet scattering transform," *Computers in Biology and Medicine*, vol. 137, p. 104814, 2021.

[7] Q. ul Ain Mubarak, M. U. Akram, A. Shaukat, F. Hussain, S. G. Khawaja, and W. H. Butt, "Analysis of pcg signals using quality assessment and homomorphic filters for localization and classification of heart sounds," *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 143–157, 2018.

[8] M. Ali Kobat and S. Dogan, "Novel three kernelled binary pattern feature extractor based automated pcg sound classification method," *Applied Acoustics*, vol. 179, p. 108040, 2021.

[9] S. K. Ghosh, R. Ponnalagu, R. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with pcg signals," *Computers in Biology and Medicine*, vol. 118, p. 103632, 2020.

[10] S.A. Pavlopoulos, A.C. Stasis, and E.N. Loukis, "A decision tree – based method for the differential diagnosis of aortic stenosis from mitral regurgitation using heart sounds," *BioMedical Engineering OnLine*, vol. 21, no. 3, 2004.

[11] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the pcg signal using time-frequency magnitude and phase features," *IEEE Sensors Letters*, vol. 3, no. 12, pp. 1–4, 2019.

[12] G. Amit, N. Gavriely, and N. Intrator, "Cluster analysis and classification of heart sounds," *Biomedical Signal Processing and Control*, vol. 4, no. 1, pp. 26–36, 2009.

[13] O. El Badlaoui, A. Benba, and A. Hammouch, "Novel pcg analysis method for discriminating between abnormal and normal heart sounds," *IRBM*, vol. 41, no. 4, pp. 223–228, 2020.

[14] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection," *Computers in Biology and Medicine*, vol. 100, pp. 132–143, 2018.

[15] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, and U. Rajendra Acharya, "Classification of heart sound signals using a novel deep wavenet model," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105604, 2020.

[16] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36 955–36 967, 2021.

[17] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105940, 2021.

[18] N. Baghel, M. K. Dutta, and R. Burget, "Automatic diagnosis of multiple cardiac diseases from pcg signals using convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105750, 2020.

[19] S. Tiwari, A. Jain, A. K. Sharma, and K. Mohamad Almustafa, "Phonocardiogram signal based multi-class cardiac diagnostic decision support system," *IEEE Access*, vol. 9, pp. 110 710–110 722, 2021.

[20] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9393–9400, 2018.

[21] O. Faust, M. Kareem, A. Shenfield, A. Ali, and U. R. Acharya, "Validating the robustness of an internet of things based atrial fibrillation detection system," *Pattern Recognition Letters*, vol. 133, pp. 55–61, 2020.

[22] M. Mishra, H. Menon, and A. Mukherjee, "Characterization of s_1 and s_2 heart sounds using stacked autoencoder and convolutional neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 9, pp. 3211–3220, 2019.

[23] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2435–2445, 2019.

[24] P. K. Jain and A. K. Tiwari, "An adaptive thresholding method for the wavelet based denoising of phonocardiogram signal," *Biomedical Signal Processing and Control*, vol. 38, pp. 388–399, 2017.

[25] Y. Khan, "Classification-of-heart-sound-signal-using-multiple-features," https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features-, 2018, [Online; accessed 11-March-2021].

[26] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, Jul 1989.

[27] S. kaur, G. Kaur, and D. Singh, "Comparative analysis of haar and coiflet wavelets using discrete wavelet transform in digital image compression," *International Journal of Engineering Research and Applications*, vol. 3, no. 3, pp. 669–673, May 2013.

[28] P. K. Jain and A. K. Tiwari, "An adaptive method for shrinking of wavelet coefficients for phonocardiogram denoising," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 1–5.

[29] T. Kautz, B. M. Eskofier, and C. F. Pasluosta, "Generic performance measure for multiclass-classifiers," *Pattern Recognition*, vol. 68, pp. 111–125, 2017.

[30] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Applied Sciences*, vol. 8, no. 12, 2018.