Phonocardiogram classification based on 1D CNN with pitch-shifting and signal uniformity techniques

Zafar Ahmad
School of Electrical Engineering and Computer Sciences (SEECS)
National University of Sciences and Technology (NUST) Islamabad, Pakistan
ahmadzafar577@gmail.com

Muhammad Salman Khan
Department of Electrical Engineering
College of Engineering, Qatar University
Doha, Qatar
salman@qu.edu.qa

Muhammad E. H. Chowdhury
Department of Electrical Engineering
College of Engineering, Qatar University
Doha, Qatar
mchowdhury@qu.edu.qa

Susu Zughaier
Department of Microbiology
Qatar University
Doha, Qatar
szughaier@qu.edu.qa

Wanis Hamad Ibrahim
Sr Consultant physician at Hamad General Hospital, Qatar
wibrahim@hamad.qa

***Abstract*— This study combines 1D CNN with advanced signal processing to enhance heart sound classification, presenting three key contributions. Initially, we used a pitch-shifting technique to expand the dataset by altering high-frequency components precisely, ensuring the preservation of vital information. Next, a signal normalization technique is deployed, equalizing signal lengths for uniform analysis across all samples. Utilizing 1D CNN and Mel-frequency cepstral coefficients (MFCCs) for feature extraction, our approach achieves notable classification accuracy, with results showing up to 99.57% accuracy, 99.80% specificity, 99.22% sensitivity, and a 99.22% F1 score. These developments not only advance the precision of heart sound classifications but also expand the potential for wider clinical applications, establishing a new benchmark in tele auscultation.**

Keywords—1D CNN, Pitch-shifting, Tele-Auscultation, Mel-frequency cepstral coefficients, heart sound classification

I. Introduction

Survey conducted by World Health Organization (WHO) shows that cardiovascular diseases (CVDs) is a global leading cause of death, enhanced diagnostic methods are vital. The increase in heart disease deaths, as reported by the WHO, highlights the urgent need for better detection techniques. The advent of artificial intelligence (AI), especially convolutional neural networks (CNNs), has revolutionized cardiac sound analysis, a critical but rarely used diagnostic tool [1]. This innovation, building on the foundational work of Laënnec in the early 1800s [2], enables more accurate diagnoses by identifying disorders through their unique sound characteristics, focusing on five heart murmurs: mitral stenosis, aortic stenosis, mitral regurgitation, mitral valve prolapses, and normal signals.

1) Mitral stenosis (MS): MS involves narrowing of the mitral valve between the left atrium and ventricle, limiting blood flow. This condition, similar to a pipeline constriction, results in increased pressure in the atrium and reduced flow to the ventricle, leading to lower cardiac output and potential pulmonary hypertension. Treatment is required to improve heart function.

2) Aortic stenosis (AS): AS refers to the narrowing of the aortic valve, impeding blood flow from the heart to the arteries and forcing the heart to exert more effort. This condition is similar to a blockage in a crucial pipeline supplying oxygen and nutrients to the body.

3) Mitral regurgitation (MR): MR occurs when the mitral valve doesn't close fully, causing blood to leak back into the left atrium from the ventricle. This results in decreased blood flow forward, leading to fatigue and breathlessness, often necessitating surgery to correct or replace the valve.

4) Mitral valve prolapse (MVP): when the mitral valve's flaps incorrectly bulge into the left atrium as the heart contracts, similar to a door that doesn't shut right. While it usually doesn't cause symptoms, it can sometimes result in chest discomfort, heart palpitations, and, in rare instances, lead to mitral regurgitation.

This paper presents the following distinctive contributions:

- Our key contribution is the implementation of a pitch-shifting technique aimed at expanding the dataset. This approach carefully targets frequency components above the signal's bandwidth (20-500 Hz) to ensure that the critical diagnostic elements of the heart sounds remain unaffected.
- To manage the variability in signal durations, we developed a specialized signal normalization technique, uniformly standardizing all recordings to a consistent 3-second length. This strategy carefully adjusts signals, maintaining temporal fidelity, to ensure consistency in analysis and model training without altering the signal's essential characteristics.
- We compare our approach to existing murmur detection and classification techniques found in the literature. Remarkably, our method outperforms these techniques, even when using the same database, feature set, and the same deep neural network 1D CNN.

This proposed study offers an in-depth analysis of our methods, results, and performance metrics in deep audio analysis for cardiac diagnostics, highlighting the fusion of AI with medical insight. We detail our methodology, review related literature, and discuss our experiments, findings, and network design, while acknowledging research limitations. Our work underscores its significance for cardiac diagnostics and patient care.

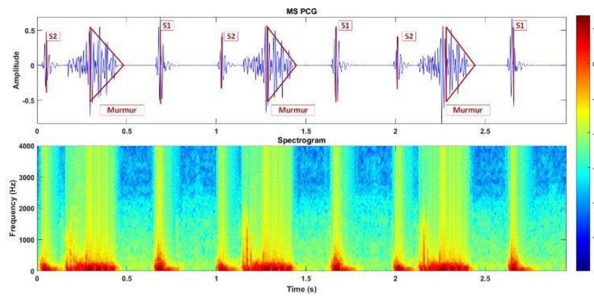

Fig. 1 (a). PCG and Spectrogram of a MS

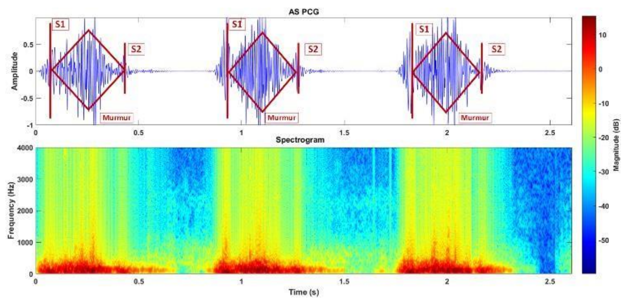

Fig. 1 (b). PCG and Spectrogram of AS

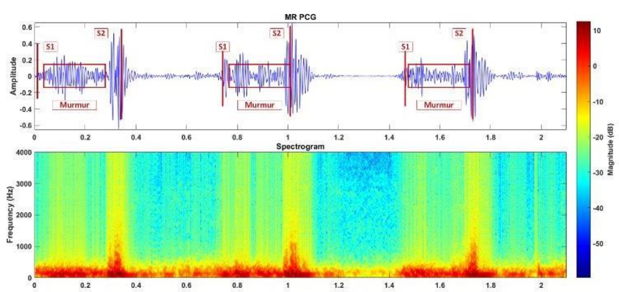

Fig. 1 (c). PCG and Spectrogram of MR

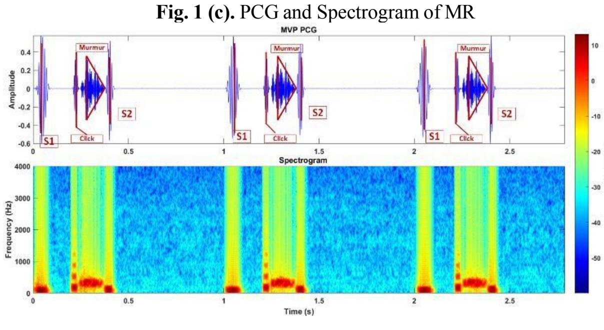

Fig. 1 (d). PCG and Spectrogram of MVP

II. Literature Review

There is rigorous research going on in parallel which highlights how deep learning, especially CNNs, and phonocardiogram (PCG) analysis work together to improve cardiovascular diagnosis. By revealing hidden patterns in

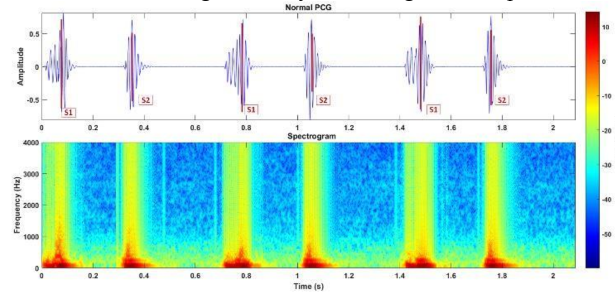

Fig. 1 (e). PCG and Spectrogram of a Normal Signal

PCG data, deep learning enhances the ability to diagnose diseases and evaluate cardiac health. Notable studies that highlight AI's ability for complex heart sound processing include Phan et al.'s [4] classification of cardiac diseases using CNNs and A. Mukherjee [3][4] automated heart illness diagnosis. Advancements include personalized PCG analysis [6] and merging MFCCs and CNNs [5] improves diagnostic precision even more. Yaseen et al.'s [11] and other researchers' [12][16] investigations into feature extraction, the effect of signal duration on classification, and creative training techniques highlight the effectiveness of deep learning in classifying heart sounds. Future work should enhance the accuracy and dependability of PCG analysis, highlighting deep learning's revolutionary potential in cardiac diagnostics, even though there are still problems with model interpretability and data scarcity.

III. Methodology

As shown in Fig. 2, our methodology includes dataset preparation, preprocessing, and data augmentation. The dataset, sourced from [11], is relatively small for deep learning needs. Therefore, it was crucial to expand the dataset's size while preserving signal quality, as any changes in quality could significantly impact model performance. This approach guided our efforts to balance data quantity and quality for optimal deep learning results.

A. Dataset

For this research, we used the dataset from [11], which offers five classes of cardiovascular diseases. This includes MS, MR, AS, MVP and normal heart signals. Each class in the dataset has 200 samples, and it's quality is quite good in terms of signal to noise ratio (SNR). However, it is important to note that the samples vary in duration: three of the classes have recordings lasting 3 seconds, while the remaining two span 2 seconds. These signals are recorded at a sampling frequency of 8000 Hz, stand out due to their minimal noise interference. The low noise level is crucial for heart signal analysis because these signals, primarily operating within a frequency bandwidth of 20 to 500 Hz, are delicate and contain vital biological insights. In literature, this bandwidth varies, some researchers have used 650 Hz [17], [18] have used their filters having cut-off frequencies from 20-400 Hz, so fundamentally most of the biological information circulates up to 500 Hz [11]. In essence, despite its limited size, the dataset's high quality, reliability, and precision make it ideal for exploring cardiac conditions in detail.

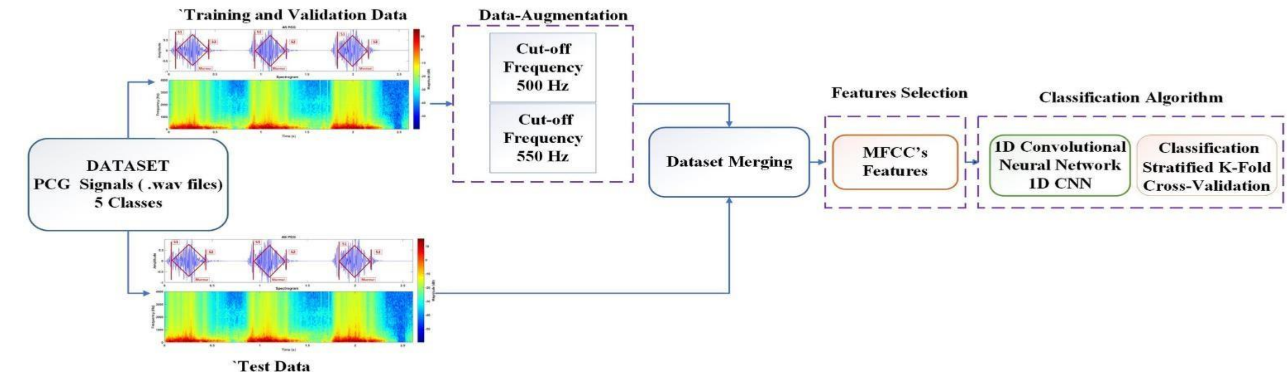

Fig. 2. Our proposed methodology

B. Data Preprocessing

The dataset utilized in this study comprises clear PCG signals. However, a notable variation exists in the signal durations across the five distinct classes, with three classes exhibiting a signal duration of 3 seconds and the remaining two classes featuring a duration of 2 seconds. This temporal inconsistency in signal lengths can pose challenges when employing deep learning models for classification tasks. To address this issue, a preprocessing technique was applied, as

Table. 1. Dataset details.

Classes	No of Samples	Duration of each signal (sec)	Sampling Frequency (kHz)
Mitral Regurgitation	200	2	8
Mitral Stenosis	200	2	8
Aortic Stenosis	200	3	8
Mitral Valve prolapse	200	3	8
Normal	200	3	8

illustrated in Table.1, which provides an overview of the dataset.

C. Signal Duration Standardization Procedure

To achieve uniform signal durations and enhance data preprocessing, we utilized MATLAB's flexible capabilities. Our methodology incorporated a comprehensive procedure driven by best practices in audio signal processing. Firstly, a predefined sampling rate of 8000 Hz was selected, allowing for precise signal characterization. Subsequently, the actual duration of each signal was assessed, considering the signal [9] length in samples and the established sampling rate. If a given signal fell short of the desired 3-second duration, an adaptive extension process was initiated. This involved replicating the signal as required using the 'repmat' function, thereby maintaining temporal coherence while effectively reaching the desired duration. Conversely, if a signal exceeded the target duration, a truncation mechanism was applied, preserving the signal's integrity by selectively retaining the specified number of samples. The result of this careful procedure was a collection of audio signals, each uniformly standardized to the 3-second temporal dimension. This standardized dataset helps avoid potential issues during model training and classification such as overfitting and underfitting, ensuring our audio signal processing is reliable and accurate.

D. Data Augmentation (DA)

Given the relatively small size of the dataset, particularly in the context of multi-class classifications and DA techniques were employed to reduce concerns regarding data scarcity. There are various DA techniques, such as pitch-shifting, time-stretching, and time compression. We used pitch-shifting technique to generate synthetic data, thereby enhancing the model's capacity to generalize and improve classification performance.

E. Pitch-Shifting (PS)

Many pitch-shifting methods have been explored in the literature [8], but our approach specifically targets only the high-frequency components of audio signals. This selective pitch-shifting method primarily affects the high- frequency segments related to the heart's first (S1) and second (S2) sounds, which are marked by their significant amplitude and frequency spikes. Unlike general pitch-shifting techniques that may broadly modify the audio signal, our method introduces fine adjustments to these high- frequency components, thereby enhancing the diversity of our dataset without compromising the signal's diagnostic integrity. Specifically, we apply pitch-shifting to frequencies above 500 Hz, carefully avoiding interference with the vital information present within the lower frequencies, typically between 20 to 500 Hz [11], crucial for PCG signal analysis. This approach ensures that while we augment the dataset for improved model training, the essential diagnostic features of heart sounds remain intact and unaltered. Mathematically, the process is shown as under:

Given an input discrete audio signal $x[n]$ sampled at a rate fs, the pitch-shifting process can be described as follows*:*

1. Compute the Discrete Fourier Transform (DFT) of the input audio signal:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j2\pi kn/N}$$

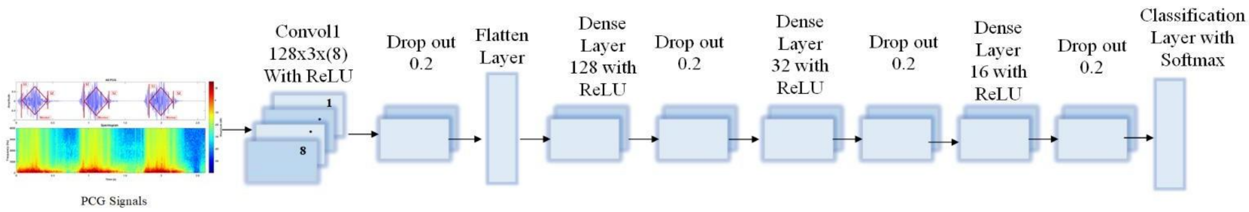

Fig. 3. Our proposed network architecture

2. Create a frequency vector f corresponding to the DFT components:

$$f[k] = \frac{f_s \cdot k}{N}$$

3. Identify the indices k of the frequency components that are above the cutoff frequency f_{cutoff}:

$$k_{\text{high}} = \{k \mid f[k] > f_{\text{cutoff}}\}$$

4. Apply a pitch-shifting operation to the high-frequency components. For example, doubling the frequency:

$$X_{\text{altered}}[k] = \begin{cases} X[k] \cdot 2 & \text{if } k \in k_{\text{high}} \\ X[k] & \text{otherwise} \end{cases}$$

5. Compute the Inverse Discrete Fourier Transform (IDFT) to obtain the altered audio signal:

$$x_{\text{altered}}[n] = \frac{1}{N} \sum_{k=0}^{N-1} X_{\text{altered}}[k] \cdot e^{j2\pi kn/N}$$

Where: $-N$ is the length of the audio signal. - $x_{\text{altered}}[n]$ is the altered audio signal. - $X[k]$ and $X_{\text{altered}}[k]$ are the DFT coefficients of the original and altered audio signals, respectively. - $f[k]$ is the frequency corresponding to the DFT coefficient $X[k]$. - k_{high} represents the set of indices of high-frequency components. - f_{cutoff} is the cutoff frequency above which pitch-shifting is applied.

IV. NEURAL NETWORKS ARCHITECTURE

In this study, we introduce a 1D CNN architecture optimized for heart sound classification using MFCCs. The model includes an input layer for MFCC data, followed by eight convolutional layers with 128 filters and 3x3 kernels for feature extraction, integrated with ReLU activation, and dropout layers to prevent overfitting. A flattening layer transitions to three dense layers with ReLU activation and dropout, leading to a Softmax output layer for class probability assignment. This architecture efficiently learns from MFCCs, achieving precise classification. Our exploration of model parameters showed that increasing the number of convolutional layers, as illustrated in Table 2, significantly enhanced performance, as evidenced by our results. Despite extensive fine- tuning of batch size, epochs, learning rate, and layer configuration, the core architecture remained robust, indicating an optimally designed model. Through trial-and-error, as detailed in our findings, we established an effective and highly optimized architecture. Our study used Stratified-KFold cross- validation to rigorously evaluate our deep learning models for heart sound classification, accommodating the dataset's unique characteristics.

Table.2. Increasing convolutional layers significantly enhanced our network's performance.

Number of Convolutional Layers	1	2	3	4	5	6	5	8	10
Accuracy									
Training	94	94.32	96.6	94.9	95.2	95.04	94.1	96.5	95.26
Validation	98.6	99.54	99.54	99.62	99.8	99.79	99.87	99.92	99.75
Test	98.41	99.14	99.1	99.27	99.36	99.46	99.35	99.57	99.34
Test Sensitivity	98.24	98.73	98.63	98.92	98.93	99.12	98.83	99.22	98.93
Test Specificity	99.56	99.68	99.66	99.73	99.73	99.78	99.7	99.8	99.73
Test F1 Score	98.24	98.73	98.63	98.93	98.93	99.12	98.83	99.22	98.93

V. EXPERIMENTATION AND RESULTS

In this section, we detail the significant outcomes of our research, demonstrating our method's effectiveness in classifying five heart disease classes. Highlighted in Figure 4, our experimental approach focused on optimizing model parameters for a highly effective architecture. Unlike traditional PCG signal classification methods summarized in Table 3, our strategy achieved superior performance. Initially, using a 1D CNN without features, we reached a 96.8% accuracy, 95% sensitivity, 98.75% specificity, and a 95% F1 Score. Incorporating MFCC features further enhanced our results to a 99.57% accuracy, 99.21% sensitivity, 99.80% specificity, and a 99.21% F1 Score, exceeding previous work as shown in Table 3 and validating our methodology's advanced PCG classification capability.

A. Performance Metrics

Our model evaluation demonstrated outstanding performance across key metrics, as detailed in Table 4. It achieved an impressive accuracy of 99.57%, a specificity of 99.80%, effectively minimizing false positives, and a sensitivity of 99.22%, showcasing its ability to accurately detect the target heart diseases. With an F1 score of 99.22%, the model balanced precision and recall exceptionally well, confirming its precision and reliability in classifying five heart disease categories, thus demonstrating its applicability in clinical settings.

VI. DISCUSSION AND FUTURE DIRECTIONS

This study introduces a deep learning approach for heart sound classification, significantly enhancing accuracy with sophisticated preprocessing and a complex CNN architecture. Despite focusing on five heart disease classes and facing limitations like potential overfitting due to limited data, our work underscores the potential for broader applications specifically tele-auscultation. Future efforts will aim to expand our dataset and scope, addressing a wider spectrum of heart diseases to improve the detection accuracy, patient care in cardiology and tele-auscultation using deep learning models.

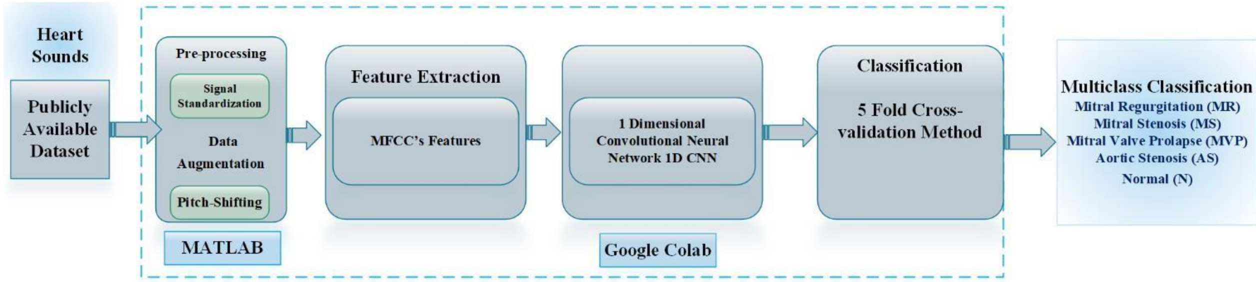

Fig. 4. Experimental setup of our proposed study.

Table.3. Relevant work

	Classifiers	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
[10]	Centroid Displacement based KNN	• MFCCS • DWT • MFCCsZ+DWT	97.4	81.98 92.00 97.60	93.50 97.90 98.80	99.20
[10]	DNN	• MFCCS • DWT • MFCCsZ+DWT	92.10	86.80 91.60 94.50	95.10 97.40 98.20	98.30
[10]	SVM	• MFCCS • DWT • MFCCsZ+DWT	97.90	87.30 92.30 98.20	96.60 98.40 99.40	99.70
[11]	BiLSTM		92.64	95.14	84.77	89.95
[12]	1D CNN	MFCCs	87.11	82.41	91.80	
[13]	1D CNN	MFCCs	95.30	95.80	96.50	95.55
[14]	1D CNN	MFCCs	81.50	84.50	78.50	
[15]	1D CNN	88	77.00	91.00	---	---
Our proposed method	**1D CNN**	No features extraction method was used	**96.80**	**95.00**	**98.75**	**95.00**
Our proposed method	**1D CNN**	**MFCCs**	**99.57**	**99.22**	**99.80**	**99.22**

ACKNOWLEDGMENT

This work is supported by Qatar University QUHI-CENG-23/24-216. The findings achieved herein are solely the responsibility of the authors.

REFERENCES

[1] B. Ahmad, F. A. Khan, K. N. Khan, and M. S. Khan, "Automatic Classification of Heart Sounds Using Long Short-Term Memory," *2021 15th Int. Conf. Open Source Syst. Technol. ICOSST 2021 - Proc.*, pp. 1–6, 2021, doi: 10.1109/ICOSST53930.2021.9683975.

[2] K. N. Khan *et al.*, "Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning," *Physiol. Meas.*, vol. 42, no. 9, 2021, doi: 10.1088/1361-6579/ac1d59.

[3] A. Mukherjee, "A Novel U-Net Architecture for Denoising of Real-world Noise Corrupted Phonocardiogram Signal".

[4] B. Xiao *et al.*, "Follow the Sound of Children's Heart: A Deep-Learning-Based Computer-Aided Pediatric CHDs Diagnosis System," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1994–2004, 2020, doi: 10.1109/JIOT.2019.2961132.

[5] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65-69.

[6] Phan, D., Bonnet, M., Hervé, J., Espinosa, M. A., Carrault, G., & Abächerli, R. (2017). A deep learning approach to cardiac abnormality detection. In Computing in Cardiology Conference (CinC) (pp. 1-4).

[7] Li, Y., Shen, Y., Zhai, Z., & Hu, S. (2019). Improved classification of heart murmur based on deep convolutional neural network with mel-frequency cepstral coefficient. Applied Sciences, 9(23), 5142.

[8] C. Gonzalez-Rodriguez, M. A. Alonso-Arevalo, and E. Garcia-Canseco, "Robust Denoising of Phonocardiogram Signals Using Time-Frequency Analysis and U-Nets," *IEEE Access*, vol. 11, no. April, pp. 52466–52479, 2023, doi: 10.1109/ACCESS.2023.3280453.

[9] Y. R. Chien, K. C. Hsu, and H. W. Tsao, "Phonocardiography signals compression with deep convolutional autoencoder for telecare applications," *Appl. Sci.*, vol. 10, no. 17, 2020, doi: 10.3390/app10175842.

[10] Yang, Y., Lo, B. P. L., & Chen, Z. (2019). A hybrid deep learning approach for heart sound classification. Biomedical Signal Processing and Control, 52, 191-198.

[11] Yaseen, G. Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, 2018, doi: 10.3390/app8122344.

[12] X. Bao, Y. Xu, and E. N. Kamavuako, "The Effect of Signal Duration on the Classification of Heart Sounds: A Deep Learning Approach," *Sensors*, vol. 22, no. 6, 2022, doi: 10.3390/s22062261.

[13] N. Rashid *et al.*, "Heart Abnormality Detection from Heart Sound Signals using MFCC Feature and Dual Stream Attention Based Network," 2022, [Online]. Available: http://arxiv.org/abs/2211.09751

[14] K. N. Khan *et al.*, "Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning," *Physiol. Meas.*, vol. 42, no. 9, 2021, doi: 10.1088/1361-6579/ac1d59.

[15] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection," *Comput. Biol. Med.*, vol. 100, no. June, pp. 132–143, 2018, doi: 10.1016/j.compbiomed.2018.06.026.

[16] T. Li, Y. Yin, K. Ma, S. Zhang, and M. Liu, "Lightweight end-to-end neural network model for automatic heart sound classification," *Inf.*, vol. 12, no. 2, pp. 1–11, 2021, doi: 10.3390/info12020054.

[17] Z. Ahmad *et al.*, "Automatic Detection of Paediatric Congenital Heart Diseases from Phonocardiogram Signals," *3rd IEEE Int. Conf. Artif. Intell. ICAI 2023*, no. February, pp. 188–195, 2023, doi: 10.1109/ICAI58407.2023.10136668.

[18] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Networks*, vol. 130, pp. 22–32, 2020, doi: 10.1016/j.neunet.2020.06.015.