# Physiology-Aligned Hierarchical Network for Heart Murmur Detection

Wenjun Lin[1], Rui Guan[2,3], Ren Jun Soon[1], Dapeng Lei[2,3], and Chee Kong Chui[1]

[1] Department of Mechanical Engineering, National University of Singapore, Singapore
e0546044@u.nus.edu, mpecck@nus.edu.sg
[2] Department of Otorhinolaryngology, Qilu Hospital of Shandong University, China
[3] HC Key Laboratory of Otorhinolaryngology, Shandong University, China

**Abstract.** Heart murmur detection is critical for early diagnosis of cardiovascular conditions. However, existing automated methods typically operate at the recording level and lack interpretability regarding murmur timing and location, limiting their clinical utility. To address this gap, we propose the Physiology-Aligned Hierarchical Network (PAHNet), a novel hierarchical framework that explicitly models heart sounds across multiple physiological scales: cardiac phases, cycles, and auscultation sites. With a patient-level supervision strategy, PAHNet enables coarse inference of murmur-relevant phases and sites, improving interpretability. Experiments on the CirCor DigiScope data demonstrate that PAHNet achieves state-of-the-art performance in murmur detection while providing clinically interpretable insights without fine-grained annotations.

**Keywords:** Murmur detection · Phonocardiogram · Hierarchical model.

## 1 Introduction

Heart sounds are generated by the vibrations of cardiac structures during each cardiac cycle, with the first (S1) and second (S2) sounds marking the boundaries of systole and diastole [1]. Heart murmurs are abnormal heart sounds that can indicate underlying cardiovascular disease, making their timely and accurate detection essential for diagnosis [3]. Traditional auscultation requires expert interpretation of phonocardiogram (PCG) recordings, making it labor-intensive and susceptible to inter-observer variability [4]. These limitations motivate a growing interest in automated heart murmur detection.

Existing automated methods can be broadly categorized into machine learning (ML) and deep learning (DL) approaches. ML-based methods typically classify handcrafted features extracted from PCG recordings using algorithms such as k-nearest neighbors [5, 12], support vector machines [14, 17], or ensemble models [10, 13]. In contrast, DL-based methods operate directly on raw or minimally processed PCG signals using architectures such as convolutional neural networks [2, 6, 15] and recurrent neural networks [7, 18]. Although effective in identifying

the presence of murmurs, existing approaches do not provide information about their timing (cardiac phase) and location (auscultation site), limiting the interpretability and clinical utility of predictions. Furthermore, most existing methods focus on recording-level analysis instead of the patient-level diagnosis required in clinical practice. By relying solely on individual recordings, they are susceptible to noise and artifacts, affected by label-task mismatch, and unable to integrate information from multiple auscultation sites, which is essential for robust and clinically meaningful murmur detection.

To overcome these limitations, we propose a Physiology-Aligned Hierarchical Network (PAHNet) for heart murmur detection. PAHNet explicitly models heart sounds across multiple physiological scales, including cardiac phase, cycle, and auscultation site, enabling structured aggregation of information that reflects the underlying cardiac physiology. Trained with only patient-level supervision, PAHNet effectively leverages this weak supervision through two key components: a phase-aware feature fusion (PAFF) module that aggregates information across cardiac phases, and a site-aware prediction fusion (SAPF) module that integrates predictions from multiple auscultation sites. In addition, a patient-level supervision strategy is introduced to guide the learning of murmur-related cardiac phases and auscultation sites. Through attention-based fusion and the patient-level supervision strategy, PAHNet supports coarse inference of which cardiac phases and auscultation sites are most indicative of a murmur, providing clinically interpretable insights without requiring fine-grained annotations.

Experiments on the CirCor DigiScope dataset [8, 11], the largest publicly available heart sound data featuring multi-site PCG recording, demonstrate that PAHNet not only improves patient-level murmur detection performance but also provides clinically meaningful insights into the model's decision-making process by coarsely inferring murmur-relevant cardiac phases and auscultation sites.

In conclusion, our contributions are summarized as follows:

1. We propose the Physiology-Aligned Hierarchical Network (PAHNet) for heart murmur detection, a framework that explicitly models heart sounds at the level of cardiac phase, cycle, and auscultation site, enabling structured information aggregation across multiple physiological scales.
2. We introduce a patient-level supervision strategy that allows coarse inference of murmur-relevant cardiac phases and auscultation sites, improving the interpretability of patient-level predictions.
3. We demonstrate improved heart murmur detection performance on the CirCor DigiScope data and show that our model provides clinically interpretable insights by identifying murmur-relevant cardiac phases and auscultation sites.

## 2 Methodology

The proposed Physiology-Aligned Hierarchical Network (PAHNet) is illustrated in Fig. 1. PAHNet follows the physiological hierarchy of cardiac auscultation through four sequential stages: phase-level feature encoding, phase-aware feature fusion within each cardiac cycle, cycle-level temporal modeling using a Temporal

Convolutional Network (TCN), and site-aware fusion for patient-level prediction. Heart sound representations are learned and aggregated bottom-up across cardiac phases, cycles, auscultation sites, and the patient level, forming an inference pipeline that is consistent with clinical auscultation principles.

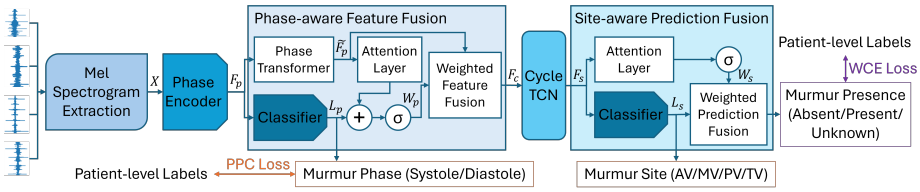


**Fig. 1.** Illustration of Physiology-Aligned Hierarchical Network (PAHNet).

### 2.1 Phase-level Feature Extraction

To support murmur analysis at different physiological levels, PCG recordings acquired from multiple auscultation sites are segmented at the cycle level using physiological boundary information. Each heart cycle is decomposed into four physiological components: S1, systole, S2, and diastole. For each phase, log-scaled Mel-spectrograms are computed to capture discriminative time-frequency characteristics. To ensure uniformity across samples, temporal dimensions are standardized via zero-padding or truncation. Consequently, the resulting hierarchical representation for each patient is formulated as $X \in \mathbb{R}^{S\times C\times P\times F\times T}$, where $S$ denotes the number of auscultation sites, $C$ represents the number of cardiac cycles per site, $P = 4$ corresponds to the cardiac phases within each cycle, and $F$ and $T$ denote the Mel frequency bins and time frames, respectively. Subsequently, the representation is fed into a ResNet18-based phase encoder to generate phase features $F_p \in \mathbb{R}^{S\times C\times P\times D}$, where $D$ indicates feature dimension.

### 2.2 Phase-aware Feature Fusion (PAFF)

A key challenge after feature extraction is aggregating these phase-specific features into a unified cycle-level representation. Simple averaging treats all phases equally, overlooking the fact that murmurs typically manifest in specific phases (systole and diastole). While attention mechanisms offer a solution, standard approaches relying on feature saliency alone are often unstable during early training or biased toward high-energy but clinically irrelevant sounds.

To address this issue, we propose a phase-aware feature fusion (PAFF) module that employs a dual-guided phase attention mechanism. This mechanism computes robust phase weights by integrating two complementary guides: a classification guide and a feature saliency guide. The classification guide, denoted as $G_{cls}$, is produced by a phase-level classifier that functions as a content-based

detector for identifying murmur-like phases. Specifically, the classifier predicts phase-level murmur logits $L_p \in \mathbb{R}^{S\times C\times P\times K}$, where $K$ denotes the number of classification categories. $G_{cls}$ is then extracted as the logit corresponding to the murmur present category, serving as a direct indicator of each phase's likelihood of containing a murmur. However, relying solely on this guide is insufficient, as phase-independent predictions are sensitive to local noise and fail to account for each phase's contextual role within the cardiac cycle.

The feature saliency guide overcomes this problem by first capturing contextual dependencies across phases via a Transformer encoder, yielding contextualized features $\tilde{F}_p = \text{TransformerEncoder}(F_p)$. A linear attention layer then transforms these features into a saliency guide $G_{sal} = \text{FC}_{sal}(\tilde{F}_p) \in \mathbb{R}^{S\times C\times P}$, which quantifies the distinctiveness of each phase within its temporal context. The two guides are fused via element-wise addition and normalized using a Sigmoid function to produce the final attention weights: $W_p = \sigma(G_{sal} + G_{cls}) \in \mathbb{R}^{S\times C\times P}$. This additive fusion ensures that high attention weights are assigned only when a phase is both acoustically distinctive and predicted to contain a murmur.

Finally, the cycle-level feature representation $F_c$ is obtained by performing a weighted sum over the phase dimension: $F_c = \sum_{p=1}^{P}(\tilde{F}_p \odot W_p) \in \mathbb{R}^{S\times C\times D}$, where $\odot$ denotes element-wise multiplication. This aggregation preserves diagnostic information from critical phases while suppressing noise from irrelevant segments. Since murmurs often span multiple consecutive cycles, we further employ a TCN to model inter-cycle dependencies and get the site-level feature by average pooling: $F_s = \text{AveragePool}(\text{TCN}(F_c)) \in \mathbb{R}^{S\times D}$.

### 2.3 Site-aware Prediction Fusion (SAPF)

While $F_s$ captures temporal dynamics within a specific auscultation site, clinical diagnosis typically integrates evidence from multiple auscultation sites. Simple aggregation strategies, such as averaging or max-pooling across sites, assume equal clinical value or rely on the single strongest signal, potentially attenuating critical features or overfitting to noisy recordings from specific sites. Therefore, we propose a site-aware prediction fusion (SAPF) module that adaptively weights the contribution of each auscultation site. This mechanism learns to prioritize sites with high confidence while suppressing irrelevant or noisy sites, thereby robustly aggregating multi-site evidence into a patient-level prediction.

In this module, each site is first processed independently to generate site-specific classification logits: $L_s = \text{Classifier}(\text{LayerNorm}(F_s)) \in \mathbb{R}^{S\times K}$, which reflect the preliminary diagnostic assessment based on features from each site. To determine the reliability of the prediction of each site, an attention score is computed based on the site's feature: $W_s = \sigma(\text{FC}(F_s)) \in \mathbb{R}^{S}$. Patient-level prediction logits are computed as the weighted sum of site-specific logits, guided by the learned attention weights: $L_{patient} = \sum_{s=1}^{S}(L_s \cdot W_s) \in \mathbb{R}^{K}$. This fusion strategy effectively implements a confidence-weighted voting scheme, where sites with higher learned importance contribute more significantly to the final diagnosis.

### 2.4 Patient-level Supervision Strategy

To effectively optimize PAHNet's hierarchical structure, we employ a composite loss function consisting of a patient-level classification loss and a phase-level physiology-constrained (PPC) loss. An explicit site-level loss is omitted because site logits are directly aggregated into the patient prediction, ensuring adequate gradient flow. In contrast, phase-level features are deeper in the hierarchy and require explicit supervision to ensure physiological consistency.

The patient-level classification loss, denoted as $\mathcal{L}_{patient}$, is implemented as a weighted cross-entropy (WCE) loss to handle class imbalance. The PPC loss adopts a dual-strategy formulation designed to jointly encourage the detection of murmur evidence in clinically relevant phases while suppressing activations elsewhere. For patients with murmurs (positive samples), pathological acoustic patterns are expected to manifest primarily during systole or diastole, rather than during the S1 or S2 heart sounds. Leveraging this clinical prior, we apply max-pooling over the murmur-class logits at systole and diastole phases, taking the maximum value as the representative score $L_{max}^{b,s,c,k}$. The positive loss is computed by applying binary cross-entropy to the max-pooled logits against the murmur-present target $y_{present}$, which is broadcast from patient-level labels:

$$\mathcal{L}_{pos} = \frac{1}{N_{pos}} \sum_{b \in \mathcal{P}} \sum_{s,c} \mathrm{BCE}(L_{\max}^{b,s,c,k}, y_{present}) \tag{1}$$

where $N_{pos}$ represents the number of instances in positive sample set $\mathcal{P}$. This formulation encourages the model to produce at least one high-confidence murmur prediction within the systolic or diastolic phases of each valid cycle.

For patients without murmurs (negative samples), all phases should be suppressed, as none is expected to exhibit murmur-like characteristics. For positive patients, suppression is applied only to the S1 and S2 phases to prevent the model from attending to these high-energy but clinically irrelevant segments. This suppression loss $\mathcal{L}_{neg}$ is defined as:

$$\mathcal{L}_{neg} = \frac{1}{|\Omega_{neg}|} \sum_{(b,s,c,p) \in \Omega_{neg}} \mathrm{BCE}(L_p, y_{absent}) \tag{2}$$

where $\Omega_{\text{neg}} = \{(b,s,c,p) \mid b \notin \mathcal{P}\} \cup \{(b,s,c,p) \mid b \in \mathcal{P},\ p \in \{\text{S1}, \text{S2}\}\}$, and $y_{absent}$ is the target vector for murmur absence.

The final PPC loss is computed as the weighted sum of $\mathcal{L}_{pos}$ and $\mathcal{L}_{neg}$. The overall loss is the weighted sum of PPC loss and patient-level classification loss: $\mathcal{L} = \lambda_{patient}\mathcal{L}_{patient} + \lambda_{pos}\mathcal{L}_{pos} + \lambda_{neg}\mathcal{L}_{neg}$, where the $\lambda$ terms are weights.

## 3 Experiments

### 3.1 Experimental Settings

**Dataset** We evaluate our method on the CirCor DigiScope dataset [8, 11], currently the largest publicly available heart sound dataset. The training set, which

is the only released portion, comprises 3,163 PCG recordings from 942 patients, with each patient contributing one or more recordings from different auscultation sites. The dataset provides patient-level murmur presence labels (absent, present, or unknown) and temporal segmentation information (S1, systole, S2, and diastole) for each PCG recording. Additionally, for each patient, the training set includes phase-level annotations (systole vs. diastole) and location-based information, including the sites where murmurs are present and the site where the murmur is most audible. We conduct 5-fold cross-validation experiments based on the training data, as it is the only portion released and contains the necessary annotations for phase-level and site-level evaluation. To prevent data leakage, the folds are generated using a patient-level split, ensuring that all recordings from a given patient are assigned exclusively to either the training or validation set within each fold.

**Evaluation Metrics** The models are evaluated using multiple metrics: the weighted accuracy introduced in the dataset benchmark [11], along with standard classification metrics including Accuracy, Precision, Recall, F1-score, area under the ROC curve (AUC), and area under the precision-recall curve (AUPRC). Given the multi-class classification setting, Precision, Recall, F1-score, AUC, and AUPRC are computed using macro-averaging. To evaluate murmur site localization performance, we also compute the Hit rate, defined as the proportion of actual positive recordings that are correctly predicted as positive.

**Implementation Details** To handle recordings of varying lengths, we cap the maximum number of cycles per recording at 16 (randomized in training, first 16 in testing). Zero-padding and binary masking are applied to missing cycles or sites. For spectrogram extraction, we use a sample rate of 4000 Hz, an FFT window size of 256, a hop length of 32, and 64 Mel-frequency bins. For the loss function, we set weighting factors to $\lambda_{patient}$=1, $\lambda_{pos}$=0.2, and $\lambda_{neg}$=0.1. The weighted cross-entropy loss employs class weights of 1.5, 5.5, and 20.5 for the absent, present, and unknown classes, respectively. The model is trained for 30 epochs with a batch size of 4. The initial learning rate is $5 \times 10^{-5}$ and reduced to $5 \times 10^{-6}$ at epoch 25. We use AdamW optimizer with a weight decay of $10^{-4}$.

### 3.2 Method Comparison

We compare the proposed PAHNet against several open-source state-of-the-art methods based on the CirCor DigiScope data under 5-fold cross-validation. The selected methods include the leading entries from the PhysioNet Challenge [11]: HearHeart [6] (1st place) and CUED_Acoustics [7] (2nd place), alongside other participants (Simulab [15]; Murmur Mia [16]) and a more recent Wav2Vec 2.0-based [9] model. Quantitative results are shown in Table 1.

As shown in Table 1, PAHNet achieves state-of-the-art performance in four imbalance-robust metrics, ranking first in Precision (0.734), F1-score (0.672), AUC (0.880), and AUPRC (0.713), while securing the second-highest Recall

**Table 1.** Comparison of murmur detection methods under 5-fold cross-validation

| Methods | Weighted accuracy | Accuracy | Precision | Recall | F1 | AUC | AUPRC |
|---|---|---|---|---|---|---|---|
| Simulab [15] | 0.722± 0.037 | 0.728± 0.011 | 0.461± 0.118 | 0.368± 0.032 | 0.338± 0.039 | 0.513± 0.048 | 0.380± 0.036 |
| Murmur Mia [16] | 0.729± 0.030 | 0.621± 0.051 | 0.449± 0.066 | 0.482± 0.035 | 0.423± 0.043 | 0.724± 0.037 | 0.549± 0.036 |
| CUED_Acoustics [7] | *0.878*± 0.016 | *0.847*± 0.013 | 0.595± 0.012 | 0.531± 0.025 | 0.546± 0.020 | 0.852± 0.039 | 0.658± 0.037 |
| HearHeart [6] | **0.903**± 0.019 | **0.860**± 0.020 | 0.611± 0.009 | 0.549± 0.033 | 0.565± 0.028 | *0.870*± 0.021 | *0.691*± 0.034 |
| Wav2Vec 2.0 [9] | 0.753± 0.060 | 0.817± 0.039 | *0.670*± 0.045 | **0.661**± 0.040 | *0.654*± 0.026 | 0.758± 0.027 | 0.549± 0.030 |
| PAHNet (Ours) | 0.814± 0.034 | 0.840± 0.022 | **0.734**± 0.062 | *0.643*± 0.036 | **0.672**± 0.032 | **0.880**± 0.019 | **0.713**± 0.032 |

(0.643). Although PAHNet ranks third in Accuracy (0.840), this trade-off reflects its strength in minimizing false positives without severely compromising sensitivity. Specifically, PAHNet improves Precision over HearHeart (0.734 vs. 0.611) with only a 2% decrease in Accuracy, while maintaining Recall within 1.8% of the top method (Wav2Vec 2.0). This balance is particularly valuable in clinical screening scenarios, where false positives can lead to unnecessary follow-up tests and increased healthcare costs, while false negatives risk missing pathological cases. The superior F1-score and AUPRC confirm PAHNet achieves a more favorable precision-recall trade-off, making it robust to class imbalance. Moreover, PAHNet exhibits the lowest standard deviation in AUC (0.019) among top models, indicating consistent performance and enhanced generalizability.

### 3.3 Ablation Study

Ablation studies are conducted to validate the effectiveness of the three key components: the PPC loss, the PAFF module, and the SAPF module. The results are listed in Table 2, where seven configurations are evaluated across four tasks: murmur detection, murmur localization, most audible auscultation site prediction, and phase-level classification. For murmur localization, a positive site prediction is defined as a site-level prediction with a confidence score greater than 0.5 when the patient-level prediction indicates murmur presence. Given that only one patient sample exhibits a diastolic heart murmur, the phase-level classification task is simplified to binary detection of systolic murmurs. In Table 2, "Phase feature average" represents replacing the PAFF module with direct averaging of the phase feature $F_p$. "Phase-level classifier" indicates whether the phase-level classifier is retained. "Site feature average" denotes replacing the SAPF module with averaging of the site feature $F_s$ before feeding it into the classifier for patient-level murmur prediction. "Site logits maximum" represents keeping the site-level classifier and selecting maximum logits as patient-level murmur predictions.

**Table 2.** Ablation studies under 5-fold cross-validation

| | | config 1 | config 2 | config 3 | config 4 | config 5 | config 6 | config 7 |
|---|---|---|---|---|---|---|---|---|
| PPC loss | | ✓ | | ✓ | | ✓ | ✓ | |
| PAFF module | | ✓ | ✓ | | | ✓ | ✓ | |
| Phase feature average | | | | ✓ | ✓ | | | ✓ |
| Phase-level classifier | | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| SAPF module | | ✓ | ✓ | ✓ | ✓ | | | |
| Site feature average | | | | | | ✓ | | ✓ |
| Site logits maximum | | | | | | | ✓ | |
| Murmur detection | Weighted accuracy | **0.814** | 0.777 | 0.763 | *0.811* | 0.773 | 0.772 | 0.775 |
| | Accuracy | **0.840** | 0.829 | 0.813 | *0.830* | 0.822 | 0.818 | 0.813 |
| | Precision | **0.734** | 0.730 | 0.697 | *0.731* | 0.724 | 0.697 | 0.700 |
| | Recall | 0.643 | 0.635 | 0.675 | 0.596 | **0.696** | *0.685* | 0.631 |
| | F1 | 0.672 | 0.647 | 0.666 | 0.633 | **0.687** | *0.678* | 0.644 |
| | AUC | **0.880** | 0.861 | 0.865 | 0.849 | *0.870* | 0.856 | 0.837 |
| | AUPRC | **0.713** | 0.697 | 0.687 | 0.672 | *0.710* | 0.686 | 0.664 |
| Murmur localization | Accuracy | **0.933** | 0.906 | 0.915 | 0.922 | / | *0.924* | / |
| | Precision | *0.802* | 0.670 | 0.695 | 0.777 | / | **0.830** | / |
| | Recall | 0.668 | *0.670* | **0.676** | 0.617 | / | 0.546 | / |
| | F1 | **0.721** | 0.648 | 0.670 | *0.675* | / | 0.651 | / |
| | AUC | **0.855** | *0.854* | 0.847 | 0.822 | / | 0.848 | / |
| | AUPRC | **0.702** | 0.686 | *0.696* | 0.667 | / | 0.679 | / |
| | Hit rate | *0.667* | 0.662 | **0.674** | 0.612 | / | 0.543 | / |
| Most audible auscultation site prediction | Accuracy | *0.849* | 0.824 | 0.840 | 0.845 | / | **0.852** | / |
| | Precision | *0.521* | 0.468 | 0.514 | 0.487 | / | **0.548** | / |
| | Recall | *0.472* | 0.456 | **0.482** | 0.454 | / | 0.456 | / |
| | F1 | **0.478** | 0.446 | 0.466 | 0.438 | / | *0.476* | / |
| | AUC | 0.796 | **0.810** | *0.804* | 0.775 | / | 0.787 | / |
| | AUPRC | 0.439 | **0.451** | *0.448* | 0.431 | / | 0.435 | / |
| Phase-level classification | Accuracy | **0.909** | *0.851* | 0.841 | / | 0.840 | 0.848 | / |
| | Precision | **0.892** | 0.737 | 0.839 | / | *0.885* | 0.803 | / |
| | Recall | **0.796** | *0.668* | 0.593 | / | 0.582 | 0.606 | / |
| | F1 | **0.830** | *0.680* | 0.603 | / | 0.595 | 0.602 | / |
| | AUC | **0.800** | 0.434 | *0.629* | / | 0.590 | 0.485 | / |
| | AUPRC | **0.607** | 0.214 | 0.321 | / | *0.342* | 0.317 | / |

As shown in Table 2, the proposed PAHNet (config 1) enables coarse inference of murmur-relevant phases and sites while outperforming all variants across all tasks. Removing the PPC loss (config 2) notably degrades both patient-level and phase-level performance, confirming its role in physiological alignment. Replacing the PAFF module with averaging (config 3 and 4) reduces performance, validating the module's effectiveness. Replacing the SAPF module with feature averaging (config 5) loses spatial relationships, making localization metrics unavailable. Using maximum logits instead (config 6) improves localization Precision but substantially reduces Recall, indicating that the attention mechanism

better balances the precision-recall trade-off by adaptively weighting murmur relevant sites. Replacing both modules with averaging (config 7) further degrades performance, confirming the value of PAHNet's hierarchical design.

## 4 Conclusion

This paper presents PAHNet, a physiology-aligned hierarchical network for heart murmur detection that addresses key limitations of existing methods in interpretability and multi-level information integration. By explicitly modeling the physiological hierarchy of cardiac auscultation, PAHNet achieves superior murmur detection performance while providing clinically meaningful insights through coarse inference of murmur phases and sites. Experiments on the CirCor DigiScope data demonstrate that PAHNet outperforms state-of-the-art methods while providing detailed phase and site information for detected murmurs.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Chakrabarti, T., Saha, S., Roy, S., Chel, I.: Phonocardiogram signal analysis-practices, trends and challenges: A critical review. In: 2015 international conference and workshop on computing and communication (IEMCON). pp. 1–4. IEEE (2015)
2. Das, S., Dandapat, S.: Heart murmur severity stages classification using multikernel residual cnn. IEEE Sensors Journal **24**(8), 13019–13027 (2024)
3. Hanna, I.R., Silverman, M.E.: A history of cardiac auscultation and some of its contributors. The American journal of cardiology **90**(3), 259–267 (2002)
4. Jiang, Z., Choi, S.: A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. Expert Systems with Applications **31**(2), 286–298 (2006)
5. Juniati, D., Khotimah, C., Wardani, D., Budayasa, K.: Fractal dimension to classify the heart sound recordings with knn and fuzzy c-mean clustering methods. In: Journal of Physics: Conference Series. vol. 953, p. 012202. IOP Publishing (2018)
6. Lu, H., Yip, J.B., Steigleder, T., Grießhammer, S., Heckel, M., Jami, N.V.S.J., Eskofier, B., Ostgathe, C., Koelpin, A.: A lightweight robust approach for automatic heart murmurs and clinical outcomes classification from phonocardiogram recordings. In: 2022 Computing in Cardiology (CinC). vol. 498, pp. 1–4. IEEE (2022)
7. McDonald, A., Gales, M.J., Agarwal, A.: Detection of heart murmurs in phonocardiograms with parallel hidden semi-markov models. In: 2022 Computing in Cardiology (CinC). vol. 498, pp. 1–4. IEEE (2022)
8. Oliveira, J., Renna, F., Costa, P.D., Nogueira, M., Oliveira, C., Ferreira, C., Jorge, A., Mattos, S., Hatem, T., Tavares, T., et al.: The circor digiscope dataset: from murmur detection to murmur classification. IEEE journal of biomedical and health informatics **26**(6), 2524–2535 (2021)
9. Panah, D.S., Hines, A., McKeever, S.: Exploring wav2vec 2.0 model for heart murmur detection. In: 2023 31st European Signal Processing Conference (EUSIPCO). pp. 1010–1014. IEEE (2023)

10. Rath, A., Mishra, D., Panda, G., Pal, M.: Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal. Biomedical Signal Processing and Control **76**, 103730 (2022)
11. Reyna, M.A., Kiarashi, Y., Elola, A., Oliveira, J., Renna, F., Gu, A., Alday, E.A.P., Sadr, N., Sharma, A., Mattos, S., et al.: Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022. In: 2022 Computing in Cardiology (CinC). vol. 498, pp. 1–4. IEEE (2022)
12. Riaz, U., Aziz, S., Umar Khan, M., Zaidi, S.A.A., Ukasha, M., Rashid, A.: A novel embedded system design for the detection and classification of cardiac disorders. Computational Intelligence **37**(4), 1844–1864 (2021)
13. Sabouri, Z., Ghadimi, A., Kiani-Sarkaleh, A., Khoshhal Roudposhti, K.: Effective features in the diagnosis of cardiovascular diseases through phonocardiogram. Multidimensional Systems and Signal Processing **34**(3), 595–632 (2023)
14. Shervegar, V.M.: Heart sound classification using wavelet scattering transform and support vector machine. Intelligent Data Analysis **27**(1_suppl), 47–63 (2023)
15. Singstad, B.J., Gitau, A.M., Johnsen, M.K., Ravn, J., Bongo, L.A., Schirmer, H.: Phonocardiogram classification using 1-dimensional inception time convolutional neural networks. In: 2022 Computing in Cardiology (CinC). vol. 498, pp. 1–4. IEEE (2022)
16. Summerton, S., Wood, D., Murphy, D., Redfern, O., Benatan, M., Kaisti, M., Wong, D.C.: Two-stage classification for detecting murmurs from phonocardiograms using deep and expert features. In: 2022 Computing in Cardiology (CinC). vol. 498, pp. 1–4. IEEE (2022)
17. Tang, H., Dai, Z., Jiang, Y., Li, T., Liu, C.: Pcg classification using multidomain features and svm classifier. BioMed Research International **2018**(1), 4205027 (2018)
18. Wu, T., Huang, Z., Li, S., Zhao, Q., Pan, F.: Heart murmur quality detection using deep neural networks with attention mechanism. Applied Sciences **14**(15), 6825 (2024)