Does Effective Heart Sound Segmentation Necessarily Require Deep Learning?

Salman Almuhammad Alali
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
salman.almuhammad-alali@univ-rennes.fr

Amar Kachenoura
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
amar.kachenoura@univ-rennes.fr

Lotfi Senhadji
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
lotfi.senhadji@univ-rennes.fr

Alfredo I. Hernandez
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
alfredo.hernandez@univ-rennes.fr

Laurent Albera
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
laurent.albera@univ-rennes.fr

Ahmad Karfoul
Univ. Rennes, Inserm, LTSI-UMR 1099
F-35000 Rennes, France
ahmad.karfoul@univ-rennes.fr

***Abstract*—Are complex Deep Learning (DL) models necessary for effective heart sound segmentation? This paper demonstrates that an effective denoising approach combined with a standard low-cost probabilistic Markov-based segmentation method provides performance that is comparable to DL models while maintaining reasonable numerical complexity. The proposed pipeline takes advantage of the pseudo-periodicity of heart events across cardiac cycles for an efficient denoising step. This preprocessing step is reformulated as a constrained low-rank matrix inference in the framework of a graph signal processing. Denoised heart sounds are then fed to the classical Logistic Regression Hidden Semi-Markov Model (LR-HSMM) for segmentation purposes. A comparative study of the proposed pipeline with one of the most effective DL models, namely CNN-LSTM, using the CirCor dataset containing both healthy and murmur PCG signals was conducted to emphasize our purpose. Obtained results in terms of both segmentation of the two primary heart sounds S1 and S2 and murmur classification of this two-step pipeline are quasi-equivalent to those obtained using the CNN-LSTM model, but with a relatively lower numerical complexity.**

***Index Terms*—Denoising, segmentation, classification, graph signal, Markov model, deep learning**

I. Introduction

Cardiovascular diseases (CVD) are the leading cause of death worldwide, accounting for 20.5 million deaths globally and 1.7 million in the European Union (EU) in 2021 [1], [2]. Beyond health impacts, CVD imposes a significant economic burden, costing the EU € 282 billion in 2021 [2], [3]. Cardiac auscultation offers a simple, cost-effective screening method for heart conditions, but accurate interpretation remains challenging due to the faint nature and low-frequency content of heart sounds. This has driven research toward automated PhonoCardioGram (PCG) analysis [4]. A typical PCG signal contains two primary heart sounds: S1 (corresponding to atrioventricular valve closure) marks the start of systole, while S2 (resulting from semilunar valve closure) is related to the beginning of diastole [5]. Heart sound segmentation, the process of accurately identifying the on/off of these sounds, forms the foundation of PCG analysis. The precision of this segmentation directly determines the quality of diagnostic information extracted from heart sounds, which is essential for both manual clinical assessments by healthcare providers and automated diagnostic algorithms.

To meet this requirement, multiple segmentation techniques have been developed. Initially, the methods were energy-based, envelope-based, or loudness-based [6]. Subsequently, probabilistic models like Hidden Markov Models (HMM) emerged, which were successful in modeling heart sound transition statuses [7], followed by probabilistic models supported by advanced machine learning (ML) approaches such as LR-HSMM, which better-modeled heart sound durations and intervals [5]. More recently, DL techniques have achieved state-of-the-art segmentation performance through various architectures including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs), and attention-based models [6] A key advantage of DL approaches is their ability to work directly with raw data and perform inherent denoising and feature extraction through their hierarchical layer structure. However, despite their high accuracy, DL methods often suffer from increased computational complexity and reduced interpretability. This raises an important question: do we always need high-complexity DL models to achieve high segmentation accuracy? Or can an effective denoising of PCG make simpler approaches comparable to DL performance?

This paper addresses these questions by introducing a two-step pipeline that combines graph-based denoising [8] with LR-HSMM [5], achieving performance comparable to DL-based CNN-LSTM [6] models while maintaining less complexity and better interpretability. CNN-LSTM has been selected here as a comparative method, due to its effectiveness in capturing both local and temporal dependencies, providing a reasonable balance between performance and complexity for segmentation tasks. The proposed denoising method leverages heart events of pseudo-periodicity by formulating the problem as inferring a low-rank matrix of denoised cardiac cycles using graph

This work was supported in part by the ANR-DiGS project Project-ANR-18-CE19-0008, in part by the ARED program of Région Bretagne DeMuG project 2497, and in part by the PEPR Santé Numérique; Projet DIIP-HEART: ANR-22-PESN-0018.

smoothness constraints. We evaluated segmentation performance by comparing LR-HSMM and CNN-LSTM models on both raw and denoised PCG CirCor datasets, followed by classification of heart sounds as healthy or pathological based on features extracted from segmented cardiac cycles. Our findings with the proposed two-step pipeline show that well-designed denoising procedure significantly improves LR-HSMM performance without removing diagnostically important low-amplitude events such as murmurs, resulting in improved segmentation accuracy and more reliable classification.

II. Notations and definitions

This paper uses italic lowercase for scalars (a), bold lowercase for vectors ($\mathbf{a}$), and bold uppercase for matrices ($\mathbf{A}$). The (i,j)-th entry and the n-th column vector of $\mathbf{A}$ are denoted by A_{ij} and $\mathbf{a}_n$, respectively. For a given matrix $\mathbf{A}$, its Frobinus norm, trace and transpose are denoted, respectively, by $\|\mathbf{A}\|_F, \mathrm{Tr}(\mathbf{A})$ and $\mathbf{A}^\mathsf{T}$. In addition, its L21-norm is given by $\|\mathbf{A}\|_{2,1} = \mathrm{Tr}(\mathbf{A}^\mathsf{T}\mathbf{\Gamma}\mathbf{A})$ where $\mathbf{\Gamma}$ is diagonal with $\Gamma_{ii} = 1/\sqrt{\sum_{j=1}^{J} A_{ij}^2}$; $\mathbf{I}_N$ represents the identity matrix of size $(N \times N)$. An undirected graph $\mathcal{G}(\mathbb{V}, \mathbb{E}, \mathbf{A})$ is defined by a set of N nodes $\mathbb{V}$ with cardianlity $|\mathbb{V}| = N$, a set of edges $\mathbb{E}$ and a symmetric adjacency matrix $\mathbf{A}$ where $A_{ij} = 1$ if the i-th and j-th nodes are connected and 0 otherwise. Non-zero entries represent similarity between nodes. The graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is symmetric positive semi-definite, with the diagonal degree matrix $\mathbf{D}$ where $D_{ii} = \sum_{j=1}^{N} A_{ij}$ (the number of connections at the i-th node). A graph signal is a function $f : \mathbb{V} \to \mathbb{R}^{\mathbb{N}}$ that assigns a real value to each node. The smoothness of a graph signal is measured by the graph total variation $\Delta_{\mathbf{L}}(\mathbf{x}) = \sum_{(i,j)\in\mathbb{E}} A_{i,j}(x_i - x_j)^2 = \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x} = \|\mathbf{L}^{1/2}\mathbf{x}\|_2^2$. A low value of $\Delta_{\mathbf{L}}(\mathbf{x})$ indicates a low graph signal variation between the connected nodes (i.e., a smooth graph signal), while a high value refers to a non-smooth graph signal.

III. Dataset

This study exploits the publicly available CirCor DigiScope dataset [9], a comprehensive collection of heart sound recordings specifically designed for cardiac auscultation research. The dataset comprises 5,282 PCG recordings from 1,568 subjects, with 3,163 recordings from 942 subjects (totaling 10.44 hours) being publicly accessible. The PCG signals were recorded at a sampling rate of 4 kHz from the four standard cardiac auscultation locations (aortic, pulmonary, tricuspid, and mitral), about 4 recordings per subject. The annotation file contains precise temporal markers that indicate the beginning and ending time instances of fundamental heart sounds S1 and S2, effectively delineating the systolic and diastolic intervals. For classification purposes, the dataset includes labels for murmur-free subjects (695) and those with murmurs (179), as well as unknown diagnoses (68) with detailed characterization of murmur properties. In this paper, we target a binary classification (murmur-free/murmur) task Thus, the recordings with unknown diagnoses will not be considered hereafter. This labeling was performed by cardiac physiologists who independently assessed each recording without relying on preexisting segmentation data. The real-world clinical conditions of the dataset, which contain various ambient and procedural noises, make it particularly suitable for evaluating robust segmentation and classification algorithms for heart sound analysis.

IV. Proposed approach

Our pipeline follows a structured sequence: first, we denoise PCG signals using our graph-based denoising method, recently proposed in [8]. Next, we use the denoised signals to train and test the LR-HSMM segmentation model [5]. Since the CirCor dataset contains raw PCG recordings (CirCor-Raw), we apply graph-based denoising to produce a cleaner version (CirCor-Denoised). Then, in the proposed two-step pipeline, we train the LR-HSMM model with this denoised dataset and test it to assess the impact of denoising on segmentation performance and the subsequent classification of segmented recordings.

A. Graph-based denoising

Let $y(t)$ be a noisy observation of a clean PCG signal $\psi(t)$ acquired at time t, such that:

$$y(t) = \psi(t) + b(t) \tag{1}$$

where $b(t)$ stands for an additive noise. Given the pseudo-periodicity of the target heart sounds across cardiac cycles, a PCG signal with N cycles can be represented in the following matrix form:

$$\mathbf{Y} = \mathbf{\Psi} + \mathbf{B} \tag{2}$$

where $\mathbf{Y}, \mathbf{\Psi}, \mathbf{B} \in \mathbb{R}^{N\times T}$, with T denoting the cycle duration. In fact, each recording is divided into individual cardiac cycles, with the start of each cycle defined as 0.05 seconds before the S1 label and the end set to 0.05 seconds before the onset of the following S1, accommodating Heart Rate Variability (HRV). Shorter cycles are zero-padded to length T. The consistent shape of S1 and S2 events across clean cycles allows each column of $\mathbf{\Psi}$ to be interpreted as a smooth signal on a graph (see Figure 1). Consequently, the clean matrix $\mathbf{\Psi}$ has a lower rank than noisy matrix $\mathbf{Y}$.

Following our previous work [8], we formulate the denoising problem as recovering a clean, low-rank matrix $\mathbf{\Psi}$ from the noisy observed matrix $\mathbf{Y}$, subject to the smoothness on graph constraint:

$$\arg\min_{\mathbf{\Psi}} \|\mathbf{Y} - \mathbf{\Psi}\|_F^2 + \lambda_1 \|\mathbf{\Psi}\|_{2,1} + \lambda_2 \left\|\mathbf{L}^{1/2}\mathbf{\Psi}\right\|_F^2 \tag{3}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ are penalty parameters. The $L_{2,1}$-norm enforces the low-rank structure [10], while $\left\|\mathbf{L}^{1/2}\mathbf{\Psi}\right\|_F^2$ promotes the smoothness on graph assumption. We construct the graph using normalized cross-correlation coefficients between cardiac cycles, forming the correlation matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$. The adjacency matrix $\mathbf{A}$ connects nodes with cross-correlation values exceeding 0.6 [11]. The Laplacian matrix $\mathbf{L}$ is derived as noted in Section II. Using the $L_{2,1}$-norm definition, the denoising problem can be rewritten as follows:

$$\arg\min_{\mathbf{\Psi}} \|\mathbf{Y} - \mathbf{\Psi}\|_F^2 + \lambda_1 \mathrm{Tr}(\mathbf{\Psi}^\mathsf{T}\mathbf{\Gamma}\mathbf{\Psi}) + \lambda_2 \mathrm{Tr}(\mathbf{\Psi}^\mathsf{T}\mathbf{L}\mathbf{\Psi}) \tag{4}$$

Assuming at this stage the full independence between $\mathbf{\Psi}$ and $\mathbf{\Gamma}$, the update rule of $\mathbf{\Psi}$ is then given by:

$$\mathbf{\Psi} = \left(\mathbf{I}_N + \lambda_1 \mathbf{\Gamma} + \lambda_2 \mathbf{L}\right)^{-1} \mathbf{Y} \tag{5}$$

Now, by recalling that, according to the definition of the $L_{2,1}$-norm, both $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ are linked, the diagonal matrix $\mathbf{\Gamma}$ is computed at each iteration as follows:

$$\Gamma_{i,i} = \frac{1}{\sqrt{\sum_{j=1}^{J} \Psi_{i,j}^2}} \tag{6}$$

The algorithm alternates between updating $\mathbf{\Psi}$ and $\mathbf{\Gamma}$ until convergence. Finally, we reconstruct the denoised PCG signal using the denoised cardiac cycles by preserving the original cycle lengths (removing the zero padding) and applying a spline filter to smooth potential gaps between limbs of cycles, creating the CirCor-Denoised dataset for subsequent LR-HSMM training. Figure 1 illustrates the denoising process.

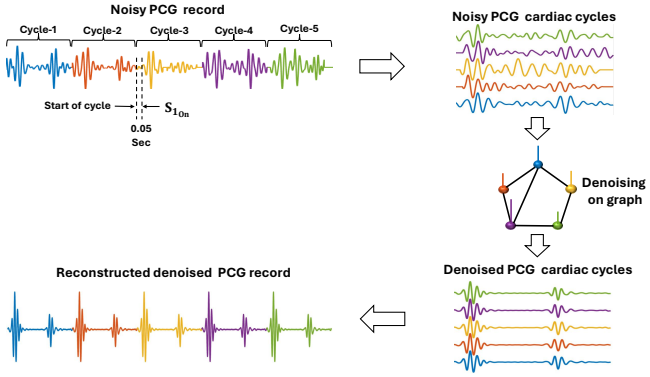

Fig. 1. Steps of denoising a PCG signal on graph.

B. Heart Sound Segmentation

The CirCor-Denoised dataset, generated using the graph-based denoising method, serves as input for heart sound segmentation using the LR-HSMM. The LR-HSMM proposed by [5], is defined by the tuple $\lambda = (\mathbf{S}, \mathbf{E}, \pi, p)$ where $\mathbf{S} = \{S_{ij}\}$ is the state transition matrix with S_{ij} representing the probability of transitioning from state q_i to state q_j; $\mathbf{E} = \{e_j(\mathbf{o}_t)\}$ are emission probabilities enhanced by logistic regression, where $e_j(\mathbf{o}_t)$ represents the probability of being in state q_j given observation $\mathbf{o}_t$; $\pi = \{\pi_i\}$ is the initial state distribution with π_i as the probability of starting in state q_i; and $p = \{p_i(d)\}$ represents explicit state duration probabilities, where $p_i(d)$ is the probability of remaining in state q_i for duration d. For heart sound segmentation, the model employs four states: q_1: S1, q_2: Systole, q_3: S2, and q_4: Diastole. Before feature extraction, PCG signals undergo Butterworth filtering with cutoff frequencies of 20 Hz and 400 Hz, followed by z-score normalization. We extract five key features from the preprocessed PCG signal: (i) **Hilbert envelope (Hilb)**, capturing instantaneous amplitude: $\text{Env}_{\text{Hilb}}(t) = |\psi(t) + j\mathcal{H}\{\psi(t)\}| = \sqrt{\psi(t)^2 + \mathcal{H}\{\psi(t)\}^2}$ where $\mathcal{H}\{\psi(t)\}$ is the Hilbert transform of $\{\psi(t)\}$; (ii) **Homomorphic envelope (Homo)**: $\text{Env}_{\text{Homo}}(t) = e^{\text{LPF}(\ln(|\psi(t)+j\mathcal{H}\{\psi(t)\}|))}$ where LPF represents a low-pass filter; (iii) **Wavelet envelope** using DWT with 'rbio3.9' wavelet at level 3: $\text{Env}_{\text{wavelet}}(t) = |\text{DWT}[\psi(t)]|$; (iv) **Power Spectral Density (PSD) envelope**: calculated using short-time Fourier transform (STFT) with Hamming windows of 0.05s width and 50% overlap, focusing on the 40-60 Hz frequency band where most S1 and S2 energy is concentrated (see [5]); and (v) **Shannon (Shan) envelope**: $\text{Env}_{\text{Shan}}(t) = -\psi(t)^2 \log(\psi(t)^2)$, which enhances segmentation performance. These features are combined into an observation vector, $\mathbf{o}_t = [\text{Env}_{\text{Homo}}(t), \text{Env}_{\text{Hilb}}(t), \text{Env}_{\text{wavelet}}(t), \text{Env}_{\text{PSD}}(t), \text{Env}_{\text{Shan}}(t)]^T$. The LR-HSMM algorithm trains a logistic regression model to estimate emission probabilities for each heart sound state, while also estimating transition probabilities between states (see [5]). In this study, we use a nonspecific patient strategy with 10-fold cross-validation with subject-stratified folds, using 90% of the data for training and 10% for evaluation.

V. Evaluation Criteria

We assess our pipeline performance using three criteria: segmentation quality, classification accuracy, and computational complexity.

A. Segmentation Quality

We quantitatively evaluate the segmentation quality using four classical metrics: Sensitivity (Se), Specificity (Sp), F1-Score, and Accuracy (Acc). These metrics are defined as: $\text{Se} = \frac{\text{TP}}{\text{TP+FN}}$, $\text{Sp} = \frac{\text{TN}}{\text{TN+FP}}$, $\text{F1-Score} = 2 \times \frac{\text{P}^+ \times \text{Se}}{\text{P}^+ + \text{Se}}$, $\text{Acc} = \frac{\text{TP+TN}}{\text{TP + TN + FP + FN}}$; here, $\text{P}^+ = \frac{\text{TP}}{\text{TP+FP}}$ is the precision, TP represents time samples correctly identified as belonging to the target events (S1, Systole, S2, or Diastole); TN refers to samples correctly identified as not belonging to a target event; FP and FN represent incorrect classifications.

B. Classification Performance

To further evaluate our approach, we assess how improved signal clarity and segmentation affect pathology classification (healthy vs. murmur). After segmentation, we extract features from systolic and diastolic regions as proposed by Kumar et al. [12]. The extracted features span four domains: (i) **Temporal domain (4 features)**: loudness (signal intensity), jitters (irregularity in periodicity), transition ratio (crescendo/decrescendo patterns), and zero-crossing rate (frequency characteristics); (ii) **Frequency domain (9 features)**: spectral power in four frequency bands (0-100Hz, 100-200Hz, 200-300Hz, 300-400Hz), spectral flux (temporal changes in spectral content), spectral centroid (center of gravity of the spectrum), spectral skewness and kurtosis (shape of spectral distribution), and fundamental frequency (lowest frequency peak); (iii) **Statistical domain (2 features)**: skewness and kurtosis of time-domain samples; and (iv) **Nonlinear dynamics (1 feature)**: largest Lyapunov exponent (chaos measurement), which quantifies complexity with higher values indicating pathological conditions (see [12] for more details). In total, 16 features are extracted from

each of the systolic and diastolic segments, resulting in a combined feature vector of 32 dimensions for classification. Following comparative tests of multiple classifiers (Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Adaptive Boosting (AdaBoost), and Naive Bayes (NB)), the weighted KNN yielded the best performance and was selected for our evaluation. We use the same four metrics (Se, Sp, F1-Score, Acc) to assess classification performance, where TP represents correctly classified murmur cases, and TN represents correctly classified healthy subjects.

C. Computational Complexity

We evaluate computational complexity by analyzing numerical complexity using Big-O notation [13], and practical implementation aspects including the number of operations per timestep, memory usage, inference time, and the number of learnable parameters. This analysis aims to show the practical advantages of the simpler proposed two-step pipeline compared to DL approaches, particularly in resource-constrained settings.

VI. Results and Discussion

In this section, we evaluate the impact of the graph-based denoising method on segmentation quality and classification accuracy on the CirCor dataset.

A. Influence of denoising on segmentation quality

We compare the performance of LR-HSMM and a CNN-LSTM model on the raw and denoised versions of CirCor dataset. The used CNN-LSTM model consists of three 1D convolutional layers (with 16, 32, and 64 filters respectively, each with a filter size of 3), tanh activation function, layer normalization, followed by two bi-directional LSTM layers (128 hidden units each), a fully connected layer with 4 units, and a softmax output layer. This architecture directly processes raw PCG signals, where the CNN layers inherently function as feature extractors. Both models were trained on raw and denoised versions of the CirCor dataset using a non-patient-specific strategy with 10-fold cross-validation, employing subject-stratified folds. For the CNN-LSTM model, each training set was further divided with 90% for training and 10% for validation. The CNN-LSTM model processed 4-second PCG signal windows in batches during training due to memory limitations, while the LR-HSMM model processed complete recordings. The models were tested on both raw and denoised data.Table I demonstrates the impact of denoising on segmentation performance. When applied to raw PCG signals (CirCor-Raw), the LR-HSMM model shows moderate performance (Acc: 0.830, Se: 0.812, Sp: 0.940, F1-score: 0.819), significantly lower than the CNN-LSTM model (Acc: 0.931, Se: 0.929, Sp: 0.977, F1-score: 0.923). This performance gap highlights the limitation of LR-HSMM when processing noisy signals. However, after applying the graph-based denoising (CirCor-Denoised), with the proposed two-step pipeline, the LR-HSMM model's performance improved significantly (Acc: 0.916, Se: 0.895, Sp: 0.969, F1-score: 0.898), and is quasi-equivalent to those of CNN-LSTM (Acc: 0.937, Se: 0.926, Sp: 0.978, F1-score: 0.924). Interestingly, the denoising did not lead to a notable improvement in segmentation with DL, as the deep neural networks can perform some type of inherent denoising. Figure 2 visually confirms these findings. In the left column, the LR-HSMM segmentation (red line) of the raw murmur PCG signal shows significant misalignment with expert annotations (black dashed line), while after denoising, the alignment is improved. For murmur-free PCG recordings (right column), denoising similarly enables the LR-HSMM model to achieve nearly perfect alignment with expert annotations.

TABLE I
Average segmentation performance across 10 folds

Dataset	Model	Acc	Se	Sp	F1-Score
Raw	LR-HSMM	0.830	0.812	0.940	0.819
	CNN-LSTM	0.931	0.929	0.977	0.923
Denoised	LR-HSMM	0.916	0.895	0.969	0.898
	CNN-LSTM	0.937	0.926	0.978	0.924

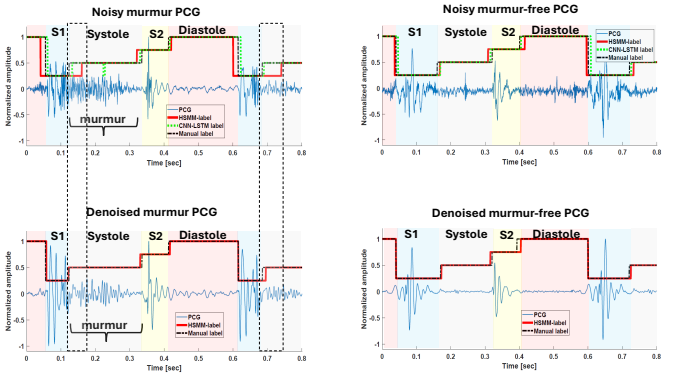

Fig. 2. Comparison of PCG segmentation before and after denoising. Top row shows raw signals and bottom row shows denoised signals using graph-based method. Left column shows PCG with murmur, right column shows murmur-free PCG, and black dashed lines represent manual expert annotation.

B. Influence of denoising on classification performance

Using the 32 features extracted from systolic and diastolic regions (described above), we evaluated classification performance with a weighted KNN classifier. Table II reveals how denoising impacts murmur classification performance. On raw data (CirCor-Raw), both segmentation methods yield modest classification results. The LR-HSMM model produces the lowest metrics (Acc: 0.693, Se: 0.603, Sp: 0.774, F1-Score: 0.652), with the CNN-LSTM model performing only slightly better (Acc: 0.714, Se: 0.640, Sp: 0.781, F1-Score: 0.680). This confirms that noise in the PCG signals significantly affects classification accuracy regardless of the segmentation method used. The most significant finding appears in the CirCor-Denoised results. After applying graph-based denoising, both segmentation methods show substantial improvements. For the proposed two-step pipeline, the LR-HSMM achieving metrics (Acc: 0.754, Se: 0.709, Sp: 0.799, F1-Score: 0.741) nearly identical to CNN-LSTM (Acc: 0.756, Se: 0.712, Sp: 0.802, F1-Score: 0.745). This eliminates the performance gap between

the two approaches, demonstrating that proper denoising can enable the simpler, more interpretable LR-HSMM model to match the classification performance of complex DL models. Note that classification performance was enhanced for both LR-HSMM and CNN-LSTM because the graph-based denoising successfully preserves the low-amplitude murmur patterns (visible in the systolic phase between S1 and S2, Figure 2) while effectively removing random noise. This selective preservation of clinically significant features is a key advantage of the graph-based approach, as it maintains diagnostically important signals and improves classification performance.

TABLE II
AVERAGE CLASSIFICATION PERFORMANCE ACROSS 10 FOLDS

Dataset	Model	Acc	Se	Sp	F1-Score
Raw	LR-HSMM	0.693	0.603	0.774	0.652
	CNN-LSTM	0.714	0.640	0.781	0.680
Denoised	LR-HSMM	0.754	0.709	0.799	0.741
	CNN-LSTM	0.756	0.712	0.802	0.745

These results indicate that the well-organized graph-based denoising approach could improve the direct segmentation of heart sounds, and also can enhance the quality of features extracted for downstream diagnostic tasks. The nearly identical performance between the proposed two-step pipeline and CNN-LSTM suggests that with appropriate preprocessing, simpler and more interpretable models can be as effective as complex DL approaches for PCG analysis. Although denoising didn't significantly improve DL-based segmentation, it still leads to enhanced subsequent classification performance.

C. Computational complexity

The computational comparison summarized in Table III highlights significant practical differences between the proposed two-step pipeline (Graph-based denoising and LR-HSMM) and CNN-LSTM models. Theoretical complexities in terms of numerical flops are $O(k_1 S_L)$ and $O(k_2 S_L)$ for the two-step pipeline and the CNN-LSTM model, respectively, where S_L is the sequence length and $k_i, i \in \{1, 2\}$ denotes the number of operations per timestep. Notably, the proposed two-step pipeline requires approximately 660 times fewer operations per timestep and fewer parameters (44 vs. 600,900) compared to the ones of CNN-LSTM. This results directly in faster inference (0.8 sec vs. 2.3 sec for a 10-sec recording) and memory consumption (24 MB vs. 112 MB). Consequently, the LR-HSMM, supported by graph-based denoising, offers an efficient and more interpretable alternative to CNN-LSTM, without sacrificing performance quality.

TABLE III
COMPUTATIONAL COMPARISON BETWEEN THE PROPOSED TWO-STEP PIPELINE AND CNN-LSTM

Metric	LR-HSMM	CNN-LSTM
Theoretical complexity	$O(k_1 S_L)$	$O(k_2 S_L)$
Operations per timestep	$k_1 = 920$	$k_2 = 608 \times 10^3$
Learnable parameters	44	600.9×10^3
Inference (10-sec record)	0.8 sec	2.3 sec
Memory (10-sec record)	24 MB	112 MB

VII. CONCLUSION

This paper proposed a two-step pipeline consisting of (i) a graph-based denoising method that exploits the pseudo-periodicity of cardiac cycles and (ii) the standard LR-HSMM technique for heart sound segmentation. The goal was to bridge the performance gap between simpler probabilistic models, supported by classical machine learning methods, and deep learning-based approaches for heart sound analysis. Results on the CirCor dataset demonstrated that, following denoising, the segmentation performance of the proposed pipeline closely matched that of the CNN-LSTM model, while offering better interpretability and requiring fewer computational resources. These findings suggest that, with proper preprocessing, simpler probabilistic models can achieve performance comparable to deep learning models for heart sound segmentation. Future work should focus on extending this approach to other pseudo-periodic physiological signals, exploring additional models, and optimizing the construction of the graph structure.

REFERENCES

[1] L. Laranjo, F. Lanas, M. C. Sun, D. A. Chen, L. Hynes, T. F. Imran, D. S. Kazi, A. P. Kengne, M. Komiyama, M. Kuwabara *et al.*, "World heart federation roadmap for secondary prevention of cardiovascular disease: 2023 update," *Global Heart*, vol. 19, no. 1, 2024.

[2] R. Luengo-Fernandez, M. Walli-Attaei, A. Gray, A. Torbica, A. P. Maggioni, R. Huculeci, F. Bairami, V. Aboyans, A. D. Timmis, P. Vardas *et al.*, "Economic burden of cardiovascular diseases in the european union: a population-based cost study," *European Heart Journal*, vol. 44, no. 45, pp. 4752–4767, 2023.

[3] A. Desiderio, M. Pastorino, M. Campitelli, M. Longo, C. Miele, R. Napoli, F. Beguinot, and G. A. Raciti, "Dna methylation in cardiovascular disease and heart failure: novel prediction models?" *Clinical Epigenetics*, vol. 16, no. 1, p. 115, 2024.

[4] F. Renna, J. Oliveira, and M. T. Coimbra, "Convolutional neural networks for heart sound segmentation," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 757–761.

[5] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2015.

[6] Z. Ren, Y. Chang, T. T. Nguyen, Y. Tan, K. Qian, and B. W. Schuller, "A comprehensive survey on heart sound analysis in the deep learning era," *IEEE Computational Intelligence Magazine*, vol. 19, no. 3, pp. 42–57, 2024.

[7] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiological Measurement*, vol. 31, no. 4, p. 513, 2010.

[8] S. A. Alali, A. Kachenoura, L. Senhadji, A. I. Hernandez, C. Michel, L. Albera, and A. Karfoul, "An efficient strategy for the denoising of heart vibration signals acquired from an implantable device," in *2024 IEEE 12th International Conference on Intelligent Systems (IS)*. IEEE, 2024, pp. 1–6.

[9] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares *et al.*, "The circor digiscope dataset: from murmur detection to murmur classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2524–2535, 2021.

[10] X. Han, "Robust low-rank tensor approximations using group sparsity," Ph.D. dissertation, Université de Rennes, 2019.

[11] H. Areiza-Laverde, C. Dopierala, L. Senhadji, F. Boucher, P. Y. Gumery, and A. Hernández, "Analysis of cardiac vibration signals acquired from a novel implant placed on the gastric fundus," *Frontiers in Physiology*, vol. 12, p. 748367, 2021.

[12] D. Kumar, "Automatic heart sound analysis for cardiovascular disease assessment," Ph.D. dissertation, Universidade de Coimbra (Portugal), 2014.

[13] L. Albera, P. Comon, L. C. Parra, A. Karfoul, A. Kachenoura, and L. Senhadji, "Biomedical applications," in *Handbook of blind source separation*. Elsevier, 2010, pp. 737–777.