



Research paper

Valvular Heart Disease Classification through Hierarchical Decomposition via Matrix Factorization of Scalogram-Based Phonocardiogram Representations

Samira Moghani, Hossein Marvi* and Zeynab Mohammadpoory

Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran.

Article Info

Article History:

Received 20 April 2025

Revised 13 May 2025

Accepted 15 June 2025

DOI:10.22044/jadm.2025.16092.2726

Keywords:

Phonocardiogram (PCG),
Valvular Heart Disease (VHD),
Deep Orthogonal Non-Negative
Matrix Factorization (Deep
ONMF), Scalogram, Time-
Frequency Analysis.

*Corresponding author:
h.marvi@shahroodut.ac.ir (H. Marvi).

Abstract

This study introduces a novel classification framework based on Deep Orthogonal Non-Negative Matrix Factorization (Deep ONMF), which leverages scalogram representations of phonocardiogram (PCG) signals to hierarchically extract structural features crucial for detecting valvular heart diseases (VHDs). Scalograms, generated via the Continuous Wavelet Transform (CWT), serve as the foundational input to the proposed feature extraction pipeline, which integrates them with Deep ONMF in a unified and segmentation-free architecture. The resulting scalogram-Deep ONMF framework is designed to hierarchically extract features through two complementary perspectives: Scale-Domain Analysis (SDA) and Temporal-Domain Analysis (TDA). These extracted features are then classified using shallow classifiers, with Random Forest (RF) achieving the best results, particularly when paired with SDA features based on the Bump wavelet. Experimental evaluations on two public PCG datasets—one with five heart sound classes and another with binary classification—demonstrate the effectiveness of the proposed method, achieving high classification accuracies of up to 98.40% and 97.23%, respectively, thereby confirming its competitiveness with state-of-the-art techniques. The results suggest that the proposed approach offers a practical and powerful solution for automated heart sound analysis, with potential applications beyond VHD detection.

1. Introduction

The Phonocardiogram (PCG) signal records the heart's acoustic activity and is visually interpreted via phonocardiography. Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, highlighting the need for accurate and timely diagnosis [1, 2]. As a non-invasive and cost-effective modality, phonocardiography is vital for early detection of heart diseases. Valvular heart disease (VHD), caused by structural valve abnormalities such as aortic stenosis (AS), mitral regurgitation (MR), and mitral stenosis (MS), alters the primary heart sounds S1 and S2, which correspond to valve closures during the cardiac cycle [3–6]. However, manual auscultation is often insufficient, prompting interest in AI-based PCG

analysis [7]. Methods involving signal processing, machine learning, and deep learning are increasingly used for PCG classification [8], with segmentation posing a significant challenge due to the complexity of detecting heart sound phases [9, 10]. In this study, we propose a Deep Orthogonal Non-negative Matrix Factorization (Deep ONMF) approach for the effective extraction of hierarchical features from scalogram representations of PCG signals. Unlike traditional methods that rely on explicit segmentation of heart sounds, the proposed framework eliminates the segmentation stage entirely, enabling direct analysis of the raw PCG signal's time-frequency representation. The method comprises three main

stages: data preparation, deep feature extraction using the proposed Deep ONMF model, and classification. In the feature extraction stage, the scalograms are decomposed hierarchically from two complementary perspectives: Scale-Domain Analysis (SDA) and Temporal-Domain Analysis (TDA), which respectively extract deep and meaningful spectral and temporal patterns from the signal. The extracted features are then used to classify normal and four types of valvular heart diseases using shallow classifiers. Experimental results indicate that among the various classifiers evaluated, RF and KNN achieved comparatively better performance, with RF even surpassing KNN, demonstrating the robustness of the proposed segmentation-free approach for automated heart sound classification.

2. Related works

The automated detection of valvular heart diseases (VHDs) from phonocardiogram (PCG) signals generally involves three main stages: segmentation of cardiac cycles, extraction of informative features, and classification of heart conditions [3]. Li et al. [11] combined traditional feature engineering techniques with deep learning frameworks to differentiate between normal and abnormal heart sounds. They utilized the Hidden Semi-Markov Model (HSMM) for segmentation of PCG signals, followed by the application of a convolutional neural network (CNN) for feature extraction. Their approach was evaluated on the PhysioNet/CinC Challenge 2016 dataset, achieving average metrics of 86.8% accuracy, 87% sensitivity, 86.6% specificity, and a Matthews Correlation Coefficient (MCC) of 72.1%. Ghosh et al. [3] performed heart sound segmentation by extracting the envelope of the PCG signal using Shannon energy, aiming to detect valvular disorders on the publicly available Yaseen database (Y-18) [12]. They applied the chirplet transform to extract time-frequency domain features, followed by the derivation of Local Energy (LEN) and Local Entropy (LENT) features from the time-frequency matrix. Their proposed multi-class classification approach achieved an overall accuracy of 98.33%. In contrast, Nawaz Khan et al. [13] proposed a segmentation-free framework for detecting heart rhythm abnormalities using CNNs. Their models, trained on the PhysioNet dataset with the aid of transfer learning, demonstrated strong performance, with the best model achieving 96.8% accuracy, 95.8% sensitivity, 98% specificity, 98.29% precision, and an F1-score of 97.05%. Tiwari et al. [14] developed a CNN-based model for heart sound

classification using Mel Frequency Cepstral Coefficients (MFCCs). They compared three types of Discrete Cosine Transform (DCT) in the feature extraction stage, reporting that DCT Type-I yielded the best results with a loss value of 0.21 and classification accuracy of 95%. Al-Naami et al. [15] introduced a method for detecting abnormal valve sounds using spectral features and an Adaptive Neuro-Fuzzy Inference System (ANFIS). Their approach consisted of signal preprocessing, feature extraction through Discrete Fourier Transform (DFT) and Higher Order Spectra (HOS), and training of the ANFIS classifier. Under optimal conditions, the model achieved 89% accuracy, 100% sensitivity, and 100% specificity.

Wavelet transform-based analysis techniques have become widely adopted for processing non-stationary signals, and they are considered particularly suitable for analyzing PCG signals [16]. In this context, study [17] focuses on the classification of fundamental heart sounds (S1 and S2) by utilizing Continuous Wavelet Transform (CWT) scalograms in conjunction with convolutional neural networks (CNNs). The model achieved a classification accuracy of 86% using CNN, and when the CNN-extracted features were classified with a support vector machine (SVM), an accuracy of 85.9% was attained on the PhysioNet dataset. In another study [18], the authors employed Discrete Wavelet Transform (DWT) for denoising, the Teager Energy Operator (TEO) and autocorrelation for signal segmentation, and CWT for feature extraction. This approach was evaluated using the PhysioNet dataset, where it achieved an accuracy of 91.19% with ResNet152 and 90.75% with VGG16 architectures. Paper [19] proposed a classification method based on wavelet transform and ensemble deep learning models. Their method demonstrated high accuracy across multiple datasets, including 98.57% on PhysioNet, 96.55% on Dataset A, and 89.19% on Dataset B of the PASCAL Challenge. In [12], the authors extracted features from PCG signals using both Mel Frequency Cepstral Coefficients (MFCCs) and DWT, aiming to improve classification performance.

Similarly, in study [20], a combination of Wavelet Packet Decomposition (WPD) and CWT was used for the analysis of PCG signals, with a focus on detecting valvular heart diseases (VHDs). Among wavelet-based methods, the CWT is often preferred for analyzing non-stationary biomedical signals because it provides a more detailed time-frequency representation. Unlike the DWT, which uses a limited set of scales and positions, CWT can

capture all possible scales of the mother wavelet. This allows for more accurate analysis of the signal. In contrast, the limited resolution of DWT might miss important signal details and lead to less reliable feature extraction [21].

3. Proposed method

In this study, Deep Orthogonal Non-negative Matrix Factorization (Deep ONMF) is used to extract deep features from scalogram representations of PCG signals, effectively utilizing scalograms as input to improve the classification of normal and four types of heart valve diseases. The proposed method consists of three main stages: data preparation including preprocessing of PCG signals and scalogram generation, hierarchical feature extraction using ONMF and classification. These stages are illustrated in Figure 1.

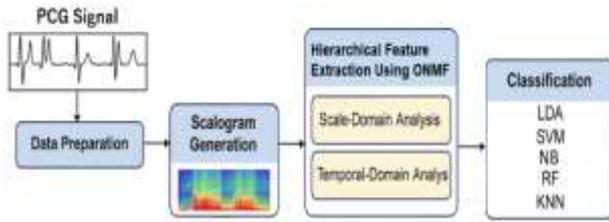


Figure 1. Flowchart of the proposed method.

3.1. Data preparation

PCG signals are first divided into two-second segments with a one-second overlap. Since the cardiac cycle lasts approximately 0.8 seconds and some heart diseases occur within one of two consecutive heart cycles [8], this approach ensures the inclusion of at least two cardiac cycles within each two-second sliding window, making it suitable for detecting heart conditions. Next, scalograms are computed to obtain a time-frequency representation of each segment. Unlike spectrograms, which use the Short-Time Fourier Transform (STFT), scalograms rely on the Continuous Wavelet Transform (CWT), providing better time-frequency localization and capturing transient cardiac events more effectively. The scalogram representation $PCG_w(\alpha, \beta)$ of a continuous PCG signal can be expressed as shown in equation (1):

$$PCG_w(\alpha, \beta) = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{\infty} PCG(t) \bar{\psi}\left(\frac{t-\beta}{\alpha}\right) dt \quad (1)$$

Where α is the scaling parameter and should be greater than 0; β is translation parameter; $\bar{\psi}$ is continuous mother wavelet. The computed scalograms are then used as input for Deep ONMF decomposition.

3.2. Hierarchical Feature Extraction Using ONMF

Deep ONMF decomposes the scalogram matrix into multiple hierarchical components. This decomposition is performed layer by layer, employing two complementary analyses: scale-domain analysis (SDA) and temporal-domain analysis (TDA). The SDA focuses on extracting multiscale spectral patterns from the PCG signal, while the TDA captures time-dependent activations. Both analyses utilize a hierarchical factorization scheme to extract meaningful features across multiple levels.

- **Scale-Domain Analysis (SDA).** Scale-domain analysis (SDA) is performed on the scalogram of the PCG signal by factorizing the coefficient matrix H . The scalogram of each PCG signal frame, represented as $X^{F \times T}$, where F is the number of frequency bins (representing the frequency resolution or scale) and T is the number of time frames (representing the temporal resolution), is used as input for the decomposition process. First layer: The factorization $X \approx W_1^{F \times d_1} H_1^{d_1 \times T}$ is applied, where W_1 extracts primary spectral features from the scalogram, and H_1 preserves corresponding temporal information. Second layer: The matrix H_1 is further decomposed as $H_1 \approx W_2^{d_1 \times d_2} H_2^{d_2 \times T}$, where W_2 extracts more refined spectral patterns from the scalogram, and H_2 enhances the temporal representation. This process continues up to the L -th layer, $H_{L-1} \approx W_L^{d_{L-1} \times d_L} H_L^{d_L \times T}$, and W_L captures the most detailed spectral features from the scalogram. To integrate the spectral features extracted from all layers into a single representative matrix, the feature matrices from each layer are multiplied, forming the final spectral feature matrix $W_{integ.}$, as expressed in equation (2), Where d_l denotes the inner rank at each layer, $l = 1, 2, \dots, L$ and L represents the total number of layers.

$$W_{(integ.)}^{(F \times dL)} = [W_1^{(F \times d1)} \times W_2^{(d1 \times d2)} \times \dots \times W_L^{(dL-1 \times dL)}] \quad (2)$$

- Temporal-Domain Analysis (TDA).** Temporal-Domain Analysis (TDA) is performed on the PCG signal by factorizing the basis matrix W . In this approach, the scalogram $X^{(F \times T)}$ is decomposed in a hierarchical manner. First layer: The factorization $X \approx W_1^{(F \times d_1)} H_1^{(d_1 \times T)}$ is performed. Second layer: The matrix W_1 undergoes further decomposition as $W_1 \approx W_2^{(F \times d_2)} H_2^{(d_2 \times d_1)}$. This process continues up to the L -th layer, where $W_{(L-1)} \approx W_L^{(F \times d_L)} H_L^{(d_L \times d_{L-1})}$, with H_l capturing increasingly refined temporal features from the scalogram. The extracted temporal features from all layers are then integrated into a single matrix $H_{integ.}$, as given by equation (3).

$$H_{(integ.)}^{(dL \times T)} = [H_L^{(dL \times d_{L-1})} \times \dots \times H_2^{(d_2 \times d_1)} \times H_1^{(d_1 \times T)}] \quad (3)$$

In both processes, the matrices W_l extract spectral patterns at different levels, while H_l refines temporal structures in a hierarchical manner. Typically, the ranks d_1, d_2, \dots, d_L are chosen in descending order ($d_1 > d_2 > \dots > d_L$) ensuring that the initial layers extract fundamental and high-variance features, while subsequent layers capture finer details. This hierarchical structure enables the identification of complex patterns at multiple levels and facilitates dimensionality reduction while preserving essential information. The multiplicative combination of feature matrices effectively compresses data into a lower-dimensional space while retaining discriminative features for heart sound classification. By integrating the matrices from all layers into a single matrix ($W_{integ.}$ and $H_{integ.}$), a feature matrix is generated for each frame. The overall procedure for extracting spectral and temporal features using the Deep ONMF model is outlined in Algorithm 1.

3.2. Classification

Classification plays a crucial role in diagnostic systems, significantly impacting their overall accuracy. Despite the growing prominence of deep learning methods, shallow classifiers remain valuable tools in various applications [7]. This study utilizes several commonly employed shallow classifiers for the detection of abnormal heart sounds. The features extracted through the proposed SDA and TDA were provided to traditional classifiers for a multi-class classification task (normal vs. four types of heart valve diseases), including Linear Discriminant

Analysis (LDA), Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), and k-Nearest Neighbors (KNN). Experimental results indicate that among the classifiers tested, RF and KNN achieved better performance compared to the others.

Algorithm 1. Deep ONMF with SDA and TDA.

Pseudocode for Deep ONMF with SDA & TDA

Input: Non-negative Scalogram $X \in \mathbb{R}^{F \times T}$; number of layers L ; layer ranks $d_1 > d_2 > \dots > d_L$; \max_iter ; threshold ϵ ;

Output: Spectral features $W_{integ.} \in \mathbb{R}^{F \times dL}$, Temporal features $H_{integ.} \in \mathbb{R}^{dL \times T}$;

Stage 1 – SDA: Initialize $W_1 \in \mathbb{R}^{F \times d_1}, H_1 \in \mathbb{R}^{d_1 \times T}$; for $l = 2$ to L : initialize $W_l \in \mathbb{R}^{d_{l-1} \times d_l}, H_l \in \mathbb{R}^{d_l \times T}$; update W_l, H_l via projected gradient with Nesterov until $\|H_{l-1} - W_l H_l\|_F^2 < \epsilon$; finally, compute $W_{integ.} = W_1 W_2 \dots W_L$.

Stage 2 – TDA: Reuse W_l, H_l ; for $l = 2$ to L : initialize $W_l \in \mathbb{R}^{F \times d_l}, H_l \in \mathbb{R}^{d_l \times d_{l-1}}$; update W_l, H_l until $\|W_{l-1} - W_l H_l\|_F^2 < \epsilon$; finally, compute $H_{integ.} = H_L H_{L-1} \dots H_1$.

Return: $W_{integ.}, H_{integ.}$

3.3. Dataset

In this study, two publicly available datasets were used to evaluate the efficiency of the proposed method. The first is the Yaseen and Kwon dataset (Y-18), which was introduced by Yaseen and Kwon [12] and is widely utilized for the detection and classification of valvular heart disease (VHD). It consists of 1,000 audio samples, categorized into five heart sound classes: normal (N), mitral regurgitation (MR), aortic stenosis (AS), mitral stenosis (MS), and mitral valve prolapse (MVP). Each class contains 200 recordings with a sampling frequency of 8 kHz, 16-bit resolution, and a bit rate of 128 kbps. The duration of each recording is approximately 3 seconds, covering at least three complete cardiac cycles [22].

In addition, the PhysioNet/CinC Challenge 2016 dataset [23] was used as the second dataset to further assess the generalizability of the proposed method. It comprises 3,126 heart sound recordings collected from various clinical and non-clinical settings, with durations ranging from 5 to 120 seconds. The recordings, obtained from key auscultation sites (aortic, pulmonic, tricuspid, mitral), are labeled as normal or abnormal, and are standardized to a 2,000 Hz sampling rate in .wav format.

4. Experimental setup

All experiments were conducted using MATLAB 2019b software. Similar to standard NMF, deep ONMF employs an alternating optimization scheme, iteratively updating each factor while keeping the others fixed. Optimization is performed using projected gradient steps with Nesterov's acceleration to minimize the Frobenius

norm of the reconstruction error. The stopping criterion is defined as either a maximum of 100 iterations or a minimum reduction of 10^{-6} in the objective function between successive iterations. The default settings of MATLAB software were also utilized for the classification process. To assess the performance of the models, a 10-fold cross-validation strategy was employed. The dataset was partitioned into ten equal-sized subsets; in each iteration, nine folds were used for training while the remaining fold served as the test set.

This process was repeated ten times, and the average accuracy across all folds was computed to ensure the robustness and generalizability of the results. In this phase, the RF and KNN classifiers were selected, as they previously demonstrated superior performance compared to other classification algorithms.

5. Experimental Results

5.1. Optimal Selection of the Number of Layers in Deep ONMF

The selection of parameters in deep matrix factorization models is contingent upon the specific application. The parameters of the deep ONMF model significantly influence its performance in detecting abnormal heart sounds. To optimize these parameters, a hyperparameter tuning process was conducted, evaluating key factors such as the number of layers and the internal ranks in the Deep ONMF structure. Increasing the number of layers beyond a certain threshold typically does not lead to the extraction of additional meaningful information or novel patterns. Existing research suggests that the optimal number of layers rarely exceeds three.

To empirically validate this, a PCG signal scalogram was processed through the proposed SDA of the deep ONMF model, utilizing a six-layer architecture.

Figure 2 illustrates the generated W matrices across six layers (a) to (f), with internal ranks of 9, 8, 7, 6, 5 and 4. The consistent patterns observed up to the third layer suggest that deeper ONMF models with more than three layers may not yield additional informative features. The selection of ranks follows a descending order, as detailed in the following.

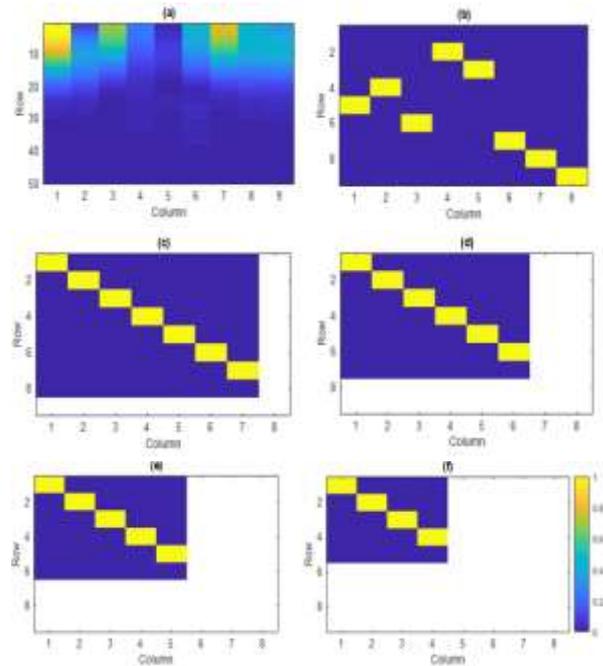


Figure 2. Matrix W related to: (a) the First Layer, (b) the Second Layer, (c) the Third Layer, (d) the Fourth Layer, (e) the Fifth Layer and (f) the Sixth Layer by the proposed SDA of the Deep ONMF model.

5.2. Optimal Selection of Internal Ranks in Deep ONMF

Ranks were chosen such that

$$d_L = d_{L-1} - 1, d_{L-1} = d_{L-2} - 1, \dots, d_2 = d_1 - 1 \text{ and } L = 3 \text{ as previously explained and}$$

$$d_1 \in \{6, 9, 12, 15, 18, 21, 24, 27, 30\} \text{ is considered.}$$

Figure 3 illustrates the accuracy and computation time trade-off for different first-layer rank values (d_1) using the first approach, a RF classifier, and 10-fold cross-validation on the Y-18 dataset.

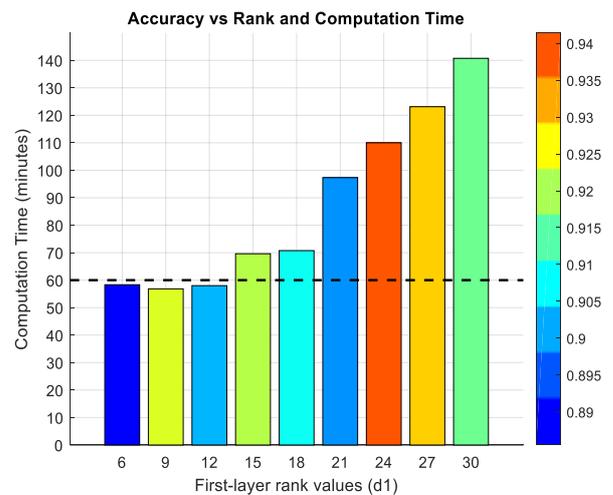


Figure 3. Trade-off between Accuracy and Computational Cost for Varying Ranks in Deep ONMF by proposed SDA.

Each bar represents a different d_1 value, with accuracy indicated by color and computation time on the y-axis. For instance, a rank of 6 corresponds to a three-layer deep ONMF model with ranks of 6, 5, and 4 for the first, second, and third layers, respectively. The results indicate that a first-layer rank of 9 offers the best balance between accuracy and computational efficiency within the desired time constraint of below 60 minutes.

5.1. Evaluation on Multiple Datasets

To further assess the generalizability of the proposed SDA and TDA methods, experiments were conducted on two well-established benchmark datasets, Y-18 and PhysioNet. These evaluations aim to demonstrate the robustness and effectiveness of the proposed approach across different and diverse datasets. An optimized Deep ONMF model—with three layers and internal ranks of 9, 8, and 7—was employed to extract hierarchical features. Using this configuration, a comparative analysis was performed across several wavelet functions to evaluate the proposed SDA and TDA.

5.2. Performance Comparison of Wavelet Functions

The wavelet functions assessed in this analysis include the analytic Morlet wavelet (referred to as amor), the Morse wavelet (morse), and the Bump wavelet (bump). These wavelet functions were employed to generate scalogram representations of Phonocardiogram (PCG) signals. Each wavelet offers unique characteristics that contribute to capturing different temporal and frequency information from the PCG signals, enabling an in-depth evaluation of the classification methods.

The choice of wavelets—Morse, Bump, and Analytic Morlet (Amor)—was informed by both prior studies and the specific characteristics of PCG signals. CWT is known for its superior time–frequency resolution in analyzing non-stationary biomedical signals compared to STFT and Wigner distribution [24, 17]. Ergen et al. [25] identified the Analytic Morlet wavelet as offering the most reliable TFR for detecting heart abnormalities. Morse and Bump wavelets also demonstrated strong performance in PCG classification tasks in recent studies [26, 16]. The performance of the proposed methods was quantitatively assessed using five evaluation metrics: Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision (Pre), and F1-Score (F1), all reported in percentage. The selection of these wavelet functions is crucial for understanding the underlying frequency components and temporal

patterns within the heart sounds, which are vital for accurate disease classification and analysis. Table 1 and Table 2 summarize the classification performance of the proposed SDA method on two datasets, Y-18 and PhysioNet, respectively. As shown in Table 1, the results clearly indicate that the Random Forest (RF) classifier consistently outperforms k-Nearest Neighbors (KNN) across all evaluated wavelet types. RF achieves higher overall accuracy—ranging from 96.71% to 98.40% on the Y-18 dataset, and 94.65% to 97.23% on the PhysioNet dataset—alongside superior sensitivity, specificity, and F1-score when compared to KNN. These findings highlight RF’s stronger capability in correctly classifying heart sound signals while minimizing classification errors. Although KNN demonstrates acceptable performance in some cases, it remains less competitive than RF in terms of both accuracy and class balance. Regarding wavelet type, the Bump wavelet yields the highest performance for both classifiers. In particular, RF achieves its best accuracy of 98.40% on Y-18 and 97.23% on PhysioNet with the Bump wavelet. In contrast, the Morse wavelet results in the weakest performance for KNN on both datasets, with an accuracy of only 68.90% for Y-18 and 63.10% for PhysioNet, while RF still maintains high classification performance (98.11% and 95.27%, respectively).

Table 1. Classification Performance of Wavelet-Based Scalograms using SDA on Y-18 Dataset.

Wavelet	Classifier	Acc	Sen	Spe	Pre	F1
Morse	RF	98.11	98.95	99.76	98.97	98.93
	KNN	68.90	68.19	96.21	81.01	73.56
Amor	RF	96.71	98.95	98.90	95.52	97.15
	KNN	65.41	68.65	95.84	78.69	73.06
Bump	RF	98.40	98.36	99.88	99.50	98.90
	KNN	75.37	77.43	96.09	82.79	79.52

Table 2. Classification Performance of Wavelet-Based Scalograms using SDA on PhysioNet Dataset.

Wavelet	Classifier	Acc	Sen	Spe	Pre	F1
Morse	RF	95.27	95.44	97.85	94.90	95.10
	KNN	63.10	65.12	92.56	76.85	70.32
Amor	RF	94.65	94.81	97.42	93.55	94.50
	KNN	61.82	64.27	91.84	75.38	69.12
Bump	RF	97.23	96.96	98.50	97.33	96.89
	KNN	72.41	74.03	94.91	80.21	77.28

The Morlet wavelet leads to intermediate outcomes, where RF again outperforms KNN but with slightly reduced accuracy compared to the Bump and Morse wavelets—96.71% for Y-18 and 94.65% for PhysioNet. Comparing RF performance with different wavelets, the Bump

wavelet provides marginally better results than the Morse wavelet in terms of overall accuracy on both datasets (98.40% vs. 98.11% for Y-18 and 97.23% vs. 95.27% for PhysioNet) and specificity (99.88% vs. 99.76% for Y-18, 98.50% vs. 97.85% for PhysioNet). However, the difference in F1-score between them remains minimal (98.90% vs. 98.93% for Y-18, and 96.89% vs. 95.10% for PhysioNet), indicating that both configurations are highly effective, with Bump having a slight advantage in overall performance and generalizability across datasets. Tables 3 and 4 report the performance of the proposed TDA method on the Y-18 and PhysioNet datasets, respectively, both showing consistently weaker results than SDA when scalograms are used as input representations. Specifically, RF achieves lower overall accuracies with TDA compared to SDA, across all wavelet types. On the Y-18 dataset, the best TDA accuracy with RF is 97.55% using the Bump wavelet, which is lower than SDA’s 98.40%. On the PhysioNet dataset, TDA with the Bump wavelet and RF achieves 95.11%, again lower than SDA’s 97.23%. The same trend is observed for KNN, with TDA consistently showing lower performance across all evaluation metrics compared to SDA for each wavelet and dataset. This suggests that the spectral features extracted via SDA are more informative and discriminative for heart sound classification in this setting. This suggests that the spectral features extracted via SDA are more informative and discriminative for heart sound classification in this setting.

Table 3. Classification Performance of Wavelet-Based Scalograms using TDA on Y-18 Dataset.

Wavelet	Classifier	Acc	Sen	Spe	Pre	F1
Morse	RF	97.21	97.88	98.16	97.23	97.91
	KNN	67.30	67.10	95.23	80.98	72.34
Amor	RF	95.51	97.39	97.45	94.33	96.19
	KNN	64.21	67.15	94.34	77.18	72.68
Bump	RF	97.55	97.44	98.65	98.45	97.37
	KNN	74.36	76.21	95.11	81.67	78.27

Table 4. Classification Performance of Wavelet-Based Scalograms using TDA on PhysioNet Dataset.

Wavelet	Classifier	Acc	Sen	Spe	Pre	F1
Morse	RF	93.44	94.03	96.91	93.85	93.72
	KNN	60.25	61.22	90.11	74.98	67.00
Amor	RF	92.88	93.16	96.10	92.17	92.62
	KNN	58.90	60.02	89.80	73.11	66.07
Bump	RF	95.11	95.30	97.62	96.00	95.47
	KNN	70.16	72.44	93.57	78.96	75.33

5.3. Baseline Comparison with Deep Neural Classifier

To further validate the effectiveness of the proposed shallow classifiers, we conducted a baseline comparison using a deep neural network. A simple yet representative convolutional neural network (CNN) was trained on the same set of SDA- and TDA-based features extracted via the proposed Deep ONMF method. The CNN architecture consisted of two convolutional layers (with 16 and 32 filters of size 3×3), each followed by batch normalization, ReLU activation, and 2×2 max pooling. This was followed by a fully connected layer with 64 units, a dropout layer (rate = 0.5), and a softmax output layer for five-class classification. Training was performed using the Adam optimizer over 20 epochs with a mini-batch size of 32. As in the previous section, the input features were derived from scalograms generated using the Bump wavelet, which had already shown superior performance among wavelet types. The results showed that despite being a deep model, the CNN did not outperform the Random Forest classifier, supporting the discriminative power and efficiency of the proposed shallow models. The average performance metrics of the CNN model are summarized in Table 5.

Table 5. CNN Classification Results Using SDA and TDA Features from Deep ONMF.

Feature Type	Classifier	Acc	Sen	Spe	Pre	F1
SDA	CNN	94.00	94.10	98.48	94.64	94.25
TDA	CNN	91.85	92.00	97.60	92.70	92.30

5.4. Evaluation of Time-Frequency Representations: Scalogram vs. Spectrogram

Figure 4 presents a comparative evaluation of the proposed model based on Deep ONMF using SDA, with two different types of input representations: scalogram and spectrogram. It is important to note that the underlying model architecture remains unchanged; only the input representation varies between the two cases. This comparison provides a detailed analysis of how each type of time-frequency representation affects classification accuracy and misclassification patterns across the different heart disease classes. By analyzing the confusion matrices for both scalogram- and spectrogram-based inputs, Figure 4 highlights the crucial impact of input feature representation on the performance of the proposed model. In Figure 4 (a), where the scalogram is used as the input representation, the model demonstrates strong classification performance across all heart

disease classes. For the Normal class, 199 out of 201 samples were correctly classified, with only 2 samples misclassified as MS. In the case of AS, the model achieved perfect classification, correctly identifying all 232 samples without any errors. For MR, 182 samples were correctly classified, while 3 were misclassified as AS and 1 as MS. Regarding the MS class, 195 out of 199 samples were correctly identified, with 1 sample misclassified as Normal (N) and 3 as AS. Finally, the model achieved perfect classification for the MVP class, correctly classifying all 185 samples. In Figure 4 (b), where the spectrogram is used as the input representation, the classification performance of the model shows greater variability across different heart disease classes. For the Normal class, 178 samples were correctly classified, while 17 were misclassified as MS, 4 as MR, and 2 as MVP. In the case of AS, the model correctly classified 231 out of 232 samples, with only 1 sample misclassified as MR. For the MR class, 165 samples were correctly identified, whereas 6 were misclassified as AS, 7 as MS, and 8 as MVP. Regarding MS, 143 samples were correctly classified, but 27 were misclassified as N, 3 as AS, 10 as MR, and 16 as MVP. Finally, for the Mitral Valve Prolapse (MVP) class, 157 samples were correctly classified, with 12 misclassified as AS, 5 as MR, and 11 as MS.

Overall, the results demonstrate that the scalogram-based model (Figure 4 (a)) provides more consistent and reliable performance across all classes, particularly in reducing misclassifications in complex cases such as MS and MVP. This highlights the superior effectiveness of scalogram representations over spectrograms in capturing the distinctive characteristics of PCG signals for heart valve disease classification.

and classification strategies. These include the use of Chirplet Z-transform spectrograms with transfer learning [3], integration of orthogonal NMF with CNN architectures [22], hybrid combinations of DWT and MFCC with SVM classifiers [12], and deep models such as ResNet101, DenseNet201, DarkNet19, and GoogLeNet based on CWT representations [26]. Additionally, lightweight CNN-RNN networks enhanced with attention mechanisms were proposed to improve noise robustness [27]. While many of these methods achieve competitive accuracy—often in the 97-98% range—their performance is frequently tied to complex preprocessing steps, segmentation requirements, or computationally intensive models. However, common challenges persist across these approaches. Many rely on explicit segmentation of cardiac cycles, which can be error-prone and sensitive to signal variability. Others require deep networks with numerous parameters, leading to high computational overhead and reduced interpretability. Additionally, handcrafted feature pipelines may lack adaptability across datasets and conditions. In contrast, the proposed method introduces a novel Deep ONMF framework that eliminates the need for segmentation and heavy preprocessing. By directly leveraging CWT-based scalograms and extracting hierarchical spectral and temporal features through Scale-Domain and Temporal-Domain Analysis, the model offers high classification performance (98.40% accuracy with Random Forest) using shallow classifiers. This not only enhances interpretability and efficiency but also makes the system more robust and easier to deploy in real-world, resource-constrained environments. Furthermore, the integration of hierarchical decomposition enables the discovery of multi-level discriminative features, contributing to superior generalization across all five heart sound classes.

6. Conclusion

The proposed method for classifying valvular heart diseases (VHDs) using phonocardiogram (PCG) signals demonstrates high classification performance while addressing key limitations of conventional approaches. We validated the generalizability of the method using two widely-used benchmark datasets, confirming its robustness across different recording conditions. Unlike traditional methods that require precise segmentation of heart sounds—a process that is often error-prone and sensitive to noise—our approach directly analyzes full PCG scalograms, eliminating the need for segmentation and

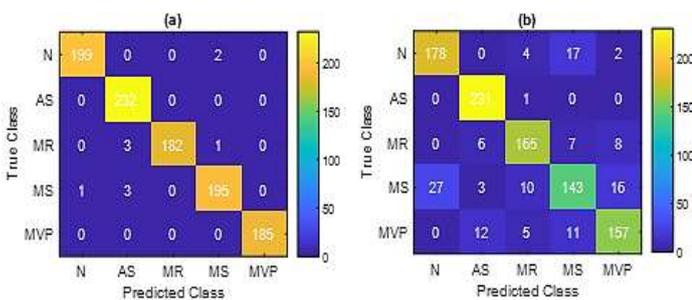


Figure 4. Confusion Matrix for VHD Detection using the Proposed SDA: (a) Scalogram, (b) Spectrogram Input.

5.2. Comparison with Recent Approaches

To evaluate the effectiveness of the proposed method, its performance was compared with several recent studies using the same or comparable datasets. As summarized in Table 6, prior works employed various feature extraction

simplifying preprocessing. By applying Deep Orthogonal Non-Negative Matrix Factorization (Deep ONMF), the method extracts hierarchical and interpretable features that capture both scale- and time-domain characteristics, while the scale-domain features consistently outperform the time-domain features in classification. In addition, the use of shallow classifiers enhances computational efficiency. Notably, the Random Forest (RF) classifier outperformed a standard Convolutional Neural Network (CNN) on the same features, reaching an accuracy of 98.40%. This highlights

both the quality of the extracted features and the practicality of the model for real-time or low-resource settings. Among the wavelet functions used to generate scalograms, Morse and Bump provided the best results, both exceeding 98% across evaluation metrics, while Amor showed slightly lower results. These findings position the proposed method as a robust and interpretable tool for automated heart sound analysis, with potential for broader application beyond VHD classification.

Table 6. Comparison of the Proposed Method with Recent Approaches for VHD Classification using Y-18.

Reference	Feature Extraction Technique	Classification Model	Acc (%)	Remarks
[3]	Chirplet Z transform (CZT) based Spectrogram	Pre-trained Networks (PTNs) with Transfer Learning (RBTL)	97	Used CZT transform for deep feature extraction and transfer learning
[22]	Orthogonal Non-negative Matrix Factorization (ONMF)	Convolutional Neural Networks (CNNs)	98	Combined ONMF with CNN for normal/abnormal heart sound detection without segmentation
[12]	Discrete Wavelet Transform (DWT) + MFCC	SVM Classifier	97.9	Used combined features of DWT and MFCC with SVM classifier
[26]	Continuous Wavelet Transform (CWT)	ResNet101, DenseNet201, DarkNet19, GoogLeNet	98	Used CWT for signal processing and deep models for classification
[27]	MFCC, STFT, CQT, CWT	NRC-Net (Convolutional Recurrent Neural Network with Attention Block)	97.4	Proposed NRC-Net with CWT and attention block for noise-robust classification
Our work	Deep ONMF using SDA and TDA on Scalograms	RF	98.40%	Proposed SDA features using BUMP wavelet achieved the highest classification accuracy with RF classifier

References

- [1] El-Dahshan, E.-S. A., El-Bakry, H. M., & Al-Saadi, Y. (2021). PCG signals for biometric authentication systems: an in-depth review. *Computer Science Review*, 41, 100420.
- [2] Ismail, S., & Ismail, B. (2023). PCG signal classification using a hybrid multi round transfer learning classifier. *Biocybernetics and Biomedical Engineering*, 43(1), 313–334.
- [3] Ghosh, S. K., Sinha, D., & Ghosh, R. (2020). Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals. *Computers in Biology and Medicine*, 118, 103632.
- [4] Abbas, S., Ismail, B., Al-Mashaqbeh, I. A., & Khan, M. A. (2024). Artificial intelligence framework for heart disease classification from audio signals. *Scientific Reports*, vol. 14, no. 1, 3123.
- [5] Milani, M. G. M., Abas, P. E., & De Silva, L. C. (2022). A critical review of heart sound signal segmentation algorithms. *Smart Health*, 24, 100283.
- [6] Randhawa, S. K., & Singh, M. (2015). Classification of heart sound signals using multi-modal features. *Procedia Computer Science*, vol. 58, pp. 165–171.
- [7] Zeng, W., Zhao, X., Hu, Y., & Lu, H. (2023). Abnormal heart sound detection from unsegmented phonocardiogram using deep features and shallow classifiers. *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26859–26883.
- [8] Ismail, S., Khan, K. N., & Ismail, B. (2023). PCG classification through spectrogram using transfer learning. *Biomedical Signal Processing and Control*, vol. 79, 104075.
- [9] Sabouri, Z., Yousefi, M. R., & Shamsollahi, M. B. (2023). Effective features in the diagnosis of cardiovascular diseases through phonocardiogram. *Multidimensional Systems and Signal Processing*, vol. 34, no. 3, pp. 595–632.
- [10] Narváez, P., Gutierrez, S., & Percybrooks, W. S. (2020). Automatic segmentation and classification of heart sounds using modified empirical wavelet transform and power features. *Applied Sciences*, vol. 10, no. 14, 4791.

- [11] Li, F., Chen, X., Qiu, T., & Wang, Z. (2020). Classification of heart sounds using convolutional neural network. *Applied Sciences*, vol. 10, no. 11, 3956.
- [12] Yaseen, G.-Y. S., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences*, vol. 8, no. 12, 2344.
- [13] Khan, K. N., Ismail, S., & Lee, Y. (2021). Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning. *Physiological Measurement*, vol. 42, no. 9, 095003.
- [14] Tiwari, S., Sapra, V., & Jain, A. (2020). Heartbeat sound classification using Mel-frequency cepstral coefficients and deep convolutional neural network. In *Advances in Computational Techniques for Biomedical Image Analysis* (pp. 115–131). Academic Press.
- [15] Al-Naami, B., Al-Ani, A., & Jalab, H. A. (2020). A framework classification of heart sound signals in PhysioNet challenge 2016 using high order statistics and adaptive neuro-fuzzy inference system. *IEEE Access*, 8, 224852–224859.
- [16] Sugiyarto, A. W., Abadi, A. M., & Sumarna, S. (2021). Classification of heart disease based on PCG signal using CNN. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 5, pp. 1697–1706.
- [17] Meintjes, A., Lowe, A., & Legget, M. (2018). Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1094–1097). IEEE.
- [18] Gelpud, J., Ortega, J. D., Segura, M. P., & Castellanos-Dominguez, G. (2021). Deep learning for heart sounds classification using scalograms and automatic segmentation of PCG signals. In *Advances in Computational Intelligence* (pp. 384–395). Springer.
- [19] Lee, J.-A., & Kwak, K.-C. (2023). Heart sound classification using wavelet analysis approaches and ensemble of deep learning models. *Applied Sciences*, vol. 13, no. 21, 11942.
- [20] Abdollahpur, M., Faez, K., & Khadivi, M. (2017). Detection of pathological heart sounds. *Physiological Measurement*, vol. 38, no. 8, pp. 1616–1632.
- [21] Singh, S. A., Meitei, T. G., & Majumder, S. (2020). Short PCG classification based on deep learning. In *Deep Learning Techniques for Biomedical and Health Informatics* (pp. 141–164). Academic Press.
- [22] Torre-Cruz, J., Perez-Diaz, J. L., Aguilar-Ruiz, J. S., & Rosado-Muñoz, A. (2023). Detection of valvular heart diseases combining orthogonal non-negative matrix factorization and convolutional neural networks in PCG signals. *Journal of Biomedical Informatics*, 145, 104475.
- [23] Shervegar, Vishwanath Madhava. "Heart sound classification technique for early CVD detection using improved wavelet time scattering and discriminant analysis classifiers." *Informatics and Health 2.1* (2025): 49-62.
- [24] Mekahlia, M. S., Fezari, M., & Aliouat, A. (2022). *PCG classification using scalogram and CNN with DAG architecture*. International Journal of Informatics and Applied Mathematics, vol. 5, no. 1, pp. 62–73.
- [25] Ergen, B., Tatar, Y., & Gulcur, H. O. (2012). *Time-frequency analysis of phonocardiogram signals using wavelet transform: A comparative study*. Computer methods in biomechanics and biomedical engineering, vol. 15, no. 4, pp. 371-81.
- [26] Wang, M., Liu, J., & Tang, Q. (2022). Transfer learning models for detecting six categories of phonocardiogram recordings. *Journal of Cardiovascular Development and Disease*, vol. 9, no. 3, 86.
- [27] Shuvo, S. B., Majumder, S., & Alam, M. T. (2023). NRC-Net: Automated noise robust cardio net for detecting valvular cardiac diseases using optimum transformation method with heart sound signals. *Biomedical Signal Processing and Control*, 86, 10527.

طبقه‌بندی بیماری‌های دریچه‌ای قلب از طریق تجزیه سلسله مراتبی با استفاده از تجزیه ماتریسی نمایش‌های فونوکارديوگرام مبتنی بر اسکالوگرام

سمیرا مغانی، حسین مروى* و زینب محمدپوری

دانشکده مهندسی برق، دانشگاه صنعتی شاهرود، شاهرود، ایران.

ارسال ۲۰۲۵/۰۴/۲۰؛ بازنگری ۲۰۲۵/۰۵/۱۳؛ پذیرش ۲۰۲۵/۰۶/۱۵

چکیده:

در این پژوهش، چارچوبی نو برای طبقه‌بندی داده‌ها بر پایه روش تجزیه ماتریس غیرمنفی ارتوگونال عمیق (Deep ONMF) معرفی می‌شود که با بهره‌گیری از نگاشت‌های اسکالوگرام سیگنال‌های صوت قلب (PCG) به صورت سلسله‌مراتبی ویژگی‌های ساختاری مهم را برای بازشناسی بیماری‌های دریچه‌ای قلب استخراج می‌کند. اسکالوگرام‌ها که به کمک تبدیل موجک پیوسته (CWT) ساخته می‌شوند، به عنوان داده‌ی ورودی اصلی در زنجیره‌ی استخراج ویژگی به کار گرفته شده و در ساختاری یکپارچه و بی‌نیاز از قطعه‌بندی، با Deep ONMF درآمیخته می‌شوند. چارچوب پیشنهادی به گونه‌ای طراحی شده است که ویژگی‌ها را از دو دیدگاه مکمل به دست آورد: بررسی در حوزه‌ی مقیاس (SDA) و بررسی در حوزه‌ی زمان (TDA). ویژگی‌های به دست آمده سپس به وسیله طبقه‌بندهای ساده بررسی می‌شوند، که در این میان، طبقه‌بند جنگل تصادفی (Random Forest) همراه با ویژگی‌های SDA بر پایه موجک Bump بهترین کارایی را داشته است. آزمایش‌ها بر روی دو پایگاه داده‌ی باز از سیگنال‌های صوت قلب—یکی دارای پنج گونه صدای قلب و دیگری برای طبقه‌بندی دودویی—نشان‌دهنده‌ی کارآمدی روش پیشنهادی است، به گونه‌ای که دقت‌هایی برابر با ۹۸/۴۰٪ و ۹۷/۲۳٪ به دست آمده است و این روش را با روش‌های پیشرفته موجود، رقابت‌پذیر می‌سازد. یافته‌ها نشان می‌دهد که روش پیشنهادی راهکاری کاربردی و توانمند برای تحلیل خودکار صدای قلب فراهم کرده و توان بهره‌برداری در زمینه‌هایی فراتر از شناسایی بیماری‌های دریچه‌ای قلب را نیز دارد.

کلمات کلیدی: فونوکارديوگرام، بیماری‌های دریچه‌ای قلب، تجزیه ماتریس غیرمنفی ارتوگونال عمیق، اسکالوگرام، تحلیل زمان-فرکانس.