

Detection of Valvular Heart Diseases combining Orthogonal Non-negative Matrix Factorization and Convolutional Neural Networks in PCG signals

J. Torre-Cruz^{a,*}, F. Canadas-Quesada^a, N. Ruiz-Reyes^a, P. Vera-Candeas^a, S. Garcia-Galan^a, J. Carabias-Orti^a and J. Ranilla^b

^aDepartment of Telecommunication Engineering. University of Jaen, Campus Científico-Tecnológico de Linares, Avda. de la Universidad, s/n, Linares (Jaen), 23700, Spain

^bDepartment of Computer Science, University of Oviedo, Campus de Gijón s/n, Gijón (Asturias), 33203, Spain

ARTICLE INFO

Keywords:

Phonocardiogram
Abnormal heart sounds
Orthogonal Non-negative Matrix Factorization
Convolutional Neural Network
Spectral pattern
Valvular heart sound

ABSTRACT

Background and objective: Valvular heart diseases (VHD) are associated with elevated mortality rates globally. Despite transthoracic echocardiography (TTE) being the gold standard detection tool, phonocardiography (PCG) could be an alternative as it is a cost-effective and non-invasive method for cardiac auscultation. A lot of researchers have dedicated their efforts to improve the decision-making process and developing robust and precise approaches to assist physicians in providing reliable diagnoses of VHDs. **Methods:** This research proposes a novel approach for the detection of anomalous valvular heart sounds from PCG signals. The proposed approach combines Orthogonal Non-negative Matrix Factorization (ONMF) and Convolutional Neural Network (CNN) architectures in a three-stage cascade, with the aim of improving the learning process and identifying the optimal ONMF temporal or spectral patterns for accurate detection. The first stage computes the time-frequency representation of the input PCG signal and performs a band-pass filtering to locate the spectral range most relevant for the presence of such cardiac abnormalities. In the second stage, the ONMF approach extracts the temporal and spectral cardiac structures which are used in the third stage to feed into the CNN architecture for detecting abnormal heart sounds. **Results:** Several state-of-the-art CNN architectures such as, LeNet5, AlexNet, ResNet50, VGG16 and GoogLeNet have been evaluated to determine the effectiveness of using ONMF temporal features for VHD detection. The results reported that the integration of ONMF temporal features with a CNN classifier produced significant improvements for VHD detection. Specifically, the proposed approach achieved an accuracy improvement of approximately 45% compared to using ONMF spectral features and 35% compared to using time-frequency features from the STFT spectrogram. Additionally, the feeding ONMF temporal features into low-complexity CNN architectures yielded competitive results comparable to those obtained with more complex architectures. **Conclusions:** The temporal structure factorized by ONMF plays a critical role in distinguishing between normal and abnormal heart sounds since the repeatability of normal heart cycles is disrupted by the presence of cardiac abnormalities. Consequently, results highlight the importance of appropriate input data representation in the learning process of CNN models in the biomedical field of the valvular heart sounds detection.

1. Introduction


Currently and according to World Health Organization (WHO) [1], cardiovascular disease (CVD) remains the largest cause of death globally with an estimated 17.9 million people dying from this cause in 2019 which accounted for approximately 32% of all deaths worldwide. Specifically, more than 85% of CVD deaths are due to myocardial infarction and stroke, more than 33% occur prematurely in people under the age of 70 and more than 75% of them occur

in low- and middle-income countries due to late medical detection. If we focus on Spain, CVDs were responsible for 24.3% of total deaths during 2020 [2].

In general, CVDs are usually associated with physiological malfunctioning of the arteries in the cardiovascular system, however, there are another type of specific CVDs called valvular heart diseases (VHDs) that are growing rapidly in the last years and are associated with structural or functional abnormalities in one or more of the four valves in the heart. The task of VHD detection is an urgent priority across Europe [3] because VHDs are also considered a major heart disease due to their high mortality rates because many people live with these cardiac disease undetected for several years until the patient requires immediate medical attention [4] or receive treatment too late, which can lead to premature death [3, 5, 6]. Several medical equipments are applied to VHDs such as, Electrocardiogram (ECG), Chest X-ray or the TransThoracic Echocardiography (TTE), the latter being the most reliable diagnostic tool for the detection of VHD at a costly investment in medical equipment, qualified

*This work was supported in part under grant PID2020-119082RB-C21,C22 funded MCIN/AEI/10.13039/501100011033, grant 1257914 funded by Programa Operativo FEDER Andalucía 2014–2020, grant P18-RT-1994 funded by the Ministry of Economy, Knowledge and University, Junta de Andalucía, Spain, grant AYUD/2021/50994 funded by Gobierno del Principado de Asturias, Spain and QUANTUM SPAIN project funded by the Ministry of Economic Affairs and Digital Transformation of the Spanish Government and the European Union through the Recovery, Transformation and Resilience Plan - NextGenerationEU

*Corresponding author

 jtorre@ujaen.es (J. Torre-Cruz)

ORCID(s):

personnel for its analysis and a long period of time for the acquisition of cardiac data [7]. Because cardiac sound signals contain relevant information associated with CVDs [8, 9], auscultation is still applied due to low cost and together with its non-invasive nature makes such analysis very attractive to minimize healthcare costs at the expense of the physician's expertise to recognise and interpret the meaning of the sounds heard through the medical devices [10, 11]. Access to basic health technologies in all primary care centers is essential to ensure that people in need receive specialized medical treatment and advice so both the medical and e-health engineering communities are investing a lot of effort to provide early and reliable diagnosis of VHD from phonocardiogram (PCG) signal analysis, a PCG signal being a recording of heart sounds using a digital stethoscope with the advantage that it can be subsequently analyzed by several physicians [10]. Optimal treatment of VHD is complex and resource-intensive, but the cost of not treating it effectively is much higher, both from the point of view of the patient and the economic cost to the sanitary system. Specifically, people suffering from VHD who do not receive proper medical treatment will typically require long hospital stays, possible admission to intensive care units and returns to the hospital, not to mention a reduction in the patient's quality of life [3].

A healthy heart is characterized by a periodic sequence of two primary heart sounds, S1 (known as "lub") and S2 (known as "dub") throughout each cardiac cycle. Each S1 sound occurs just before a short gap known as systole in which both mitral and tricuspid valves close while each S2 sound occurs just before a short gap known as diastole in which both the aortic and pulmonary valves close [10]. In this manner, the healthy heart valves generate an inaudible blood flow that circulates in a forward direction due to the correct closing and opening of the heart valves. The normal heart sounds are associated to the presence of the S1 and S2 sounds, located most of the relevant spectral content in the range of 20-150 Hz [12, 13] as shown in the first row of Figure 1. However, the presence of any VHD implies the circulation of the blood flow in a backward direction, appearing some of the main types of VHDs such as, aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP). The term stenosis indicates a narrowing of the valve that prevents adequate blood flow, while the term regurgitation is associated with the inability to prevent backflow of blood when the valves do not join properly. As a consequence, any VHD generates an audible blood flow because it becomes turbulent due to an anomalous heart valve [14], appearing abnormal heart sounds such as clicks, snaps and murmurs whose significant spectral components can often be located at higher spectral range between 500-600 Hz [12, 13]. These abnormal heart sounds can be categorized according to some features such as timing (systole or diastole), intensity (holosystolic, crescendo and decrescendo) and what it sounds like (harsh, high-pitched, low-pitched or blowing) as shown in the second and third row of Figure 1.

Several signal processing and machine learning approaches have been proposed to address heart detection and classification such as, Hidden Markov Models (HMM) [15, 16, 17, 18, 19], Chirplet transform [20], Cepstrum [21, 22], Cochleagram-based spectral clustering [23], Linear Predictive Coding (LPC) [24], Envelopgrams [25], K-Nearest Neighbour (KNN) [26], Support Vector Machine (SVM) [27, 28, 29, 30, 31, 32], Wavelet entropy [33, 34, 35], Wavelet transform [36, 37, 38], Entropy [39], Empirical mode decomposition (EMD) [40], Principal Component Analysis (PCA) [41, 42], Spectrograms [43, 44] and Mel-Frequency Cepstrum Coefficients (MFCC) [45, 46] and Spectral dissimilarity [10]. Recently, deep learning (DL) techniques have been used in order to save time and avoid manual time-frequency feature extraction that was performed in most of the previous approaches, where ingrained knowledge of signal processing is required. In this sense, DL attempts to automate the feature extraction process without the intervention of the signal processing engineer through the training of models most of them being based on Convolutional Neural Networks (CNN) [47, 48, 49, 50, 51, 52, 53, 54, 55] and Recurrent Neural Networks (RNN) [56, 57, 58, 59, 60, 61, 62]. Das et al. [23] use spectral clustering as unsupervised cardiac sound segmentation method achieving a remarkable level of accuracy when applied on Cochleagram feature. In [32], several features from heart spectrogram images are extracted using pre-trained CNN models followed by SVM classifier. In [34], authors employed time-frequency representation based on wavelet in order to select three two-dimensional feature distributions comprising six features that included energy and entropy information to classify and identify normal heart sounds and systolic heart murmurs. In [38], a system to classify cardiac disorder is presented in which the feature extraction is performed using MFCCs and wavelets features from the heart sounds and applying SVM, deep neural network (DNN) and KNN. Li et al. [51] addressed the task of normal and abnormal heart sounds classification combining a set of feature and a compact CNN model, including a study to determine the best feature selection and classification performance based on CNN architecture. Baghel et al. [54] applied a CNN model to reduce misclassification of several heart disorders using a denoising stage based on a Gaussian filter and data augmentation technique to increase the robustness of the proposal. In order to automate the feature extraction and selection process in the analysis of the PCG signal, Chen et al. [62] combined one-dimensional convolutional neural networks (1D-CNN) and long short-term memory networks (LSTM) to classify normal and abnormal heart sounds. Non-negative matrix factorization (NMF) has been successfully applied in several sound signal processing areas such as, audio [63, 64, 65, 66, 67], image [68, 69, 70] and biomedicine [71, 72, 73, 74, 75, 76, 77] confirming its high potential to find hidden spectral and temporal structures into the raw data. Canadas et al. [71] combined similarities and energy distributions from spectral and temporal NMF patterns to separate heart and lung

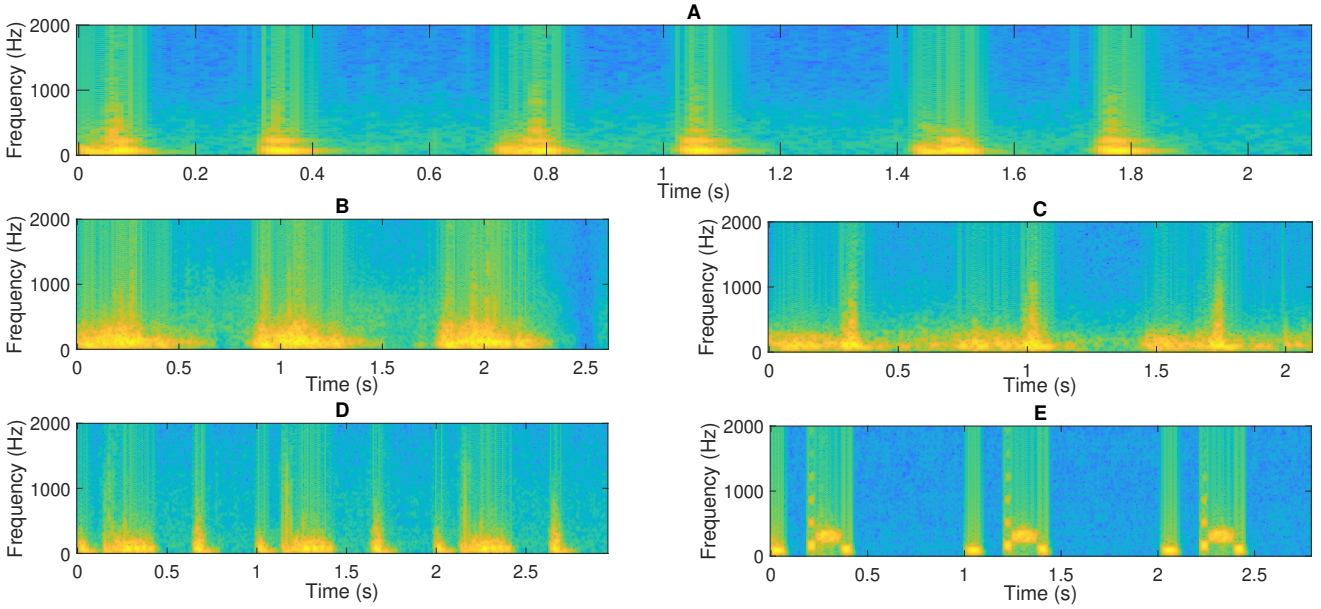


Figure 1: Time-frequency representations (spectrograms) associated to normal and abnormal heart sounds in PCG signals: normal (A), AS (B), MR (C), MS (D) and MVP (E).

sounds. Dia et al. [73] estimated the heart rate from PCG signals using an NMF approach based on a source-filter model to model the quasi-harmonic structure of heart sounds. The study cited as [74] employs non-smooth NMF (nsNMF) to identify the genes and pathways that are linked with specific types of cancer. The dimensionality of the data is reduced by means of nsNMF, which enables the subsequent training of a cancer type classifier using a SVM model.

The main contribution of this work is to demonstrate the efficacy of ONMF in improving the learning process of CNN-based architectures for the detection of abnormal heart sounds from PCG signals. Thus, ONMF is capable of extracting relevant temporal structures that contain meaningful cardiac content, thereby allowing for the identification of significant deviations from the periodic pattern exhibited by a healthy heart, which are indicative of cardiac malfunction [21]. To this end, the proposed method consists of three stages aimed at improving the learning of CNN architectures for VHD detection. The first stage involves the computation of a time-frequency representation and the application of several band-pass filters to determine the most relevant spectral band for detecting abnormal heart sounds. The second stage decomposes the input signal and extracts the time-domain and frequency-domain features associated to the heart sounds. Finally, the third stage employs a CNN architecture to detect the presence of valvular heart sounds using either the temporal or spectral features factorized by ONMF. Results show that the use of temporal ONMF features improves the learning of CNN architectures in this biomedical field since achieves that low-complexity CNN models provide comparable performance to more complex CNN models. This underscores the importance of selecting an appropriate representation of input data during the training stage of a CNN model.

The remaining sections of the paper are structured as follows: Section 2 gives a review of the fundamentals of Non-negative Matrix Factorization and Convolutional Neural Networks. Section 3 illustrates in detail the proposed method applied to VHD detection. Section 4 details the setup, metrics and discusses the experimental results with other state-of-the-art approaches. Finally, Section 5 describes the main conclusions and the main directions in future work.

2. Background

2.1. Notations and basic concepts

Consider an input signal $x(t)$ composed of normal heart sounds $x_N(t)$ and abnormal heart sounds $x_A(t)$, being their sampled versions $x[n]$, $x_N[n]$ and $x_A[n]$ where the n is the sample index using a sampling rate f_s Hz. We assume the input magnitude spectrogram $\mathbf{X} \in \mathbb{R}_+^{F \times T}$, composed of F frequency bins and T frames, as a linear mixing model using $\mathbf{X}_N \in \mathbb{R}_+^{F \times T}$ and $\mathbf{X}_A \in \mathbb{R}_+^{F \times T}$ which denotes the normal and abnormal heart sound magnitude spectrogram such as, $\mathbf{X} = \mathbf{X}_N + \mathbf{X}_A$. Each input spectrogram \mathbf{X} has been computed by means of the magnitude of the Short-Time Fourier Transform (STFT) applying a Hamming window of size N with 25% overlap. In this work, a normalized spectrogram $\bar{\mathbf{X}}$ has been computed by enforcing its L_1 -norm equals to the unity in order to be independent with respect to input spectrogram and to the observation interval,

$$\bar{\mathbf{X}} = \frac{\mathbf{X}}{\left(\frac{\sum_{f=1}^F \sum_{t=1}^T X_{f,t}}{FT} \right)} \quad (1)$$

Hereafter, the normalized spectrogram $\bar{\mathbf{X}}$ will be noted as \mathbf{X} to simplify the nomenclature throughout the paper.

2.2. Non-Negative Matrix Factorization

Classical Non-negative matrix factorization (NMF) [78], unconstrained NMF, is a technique for multidimensional data reduction to extract hidden spectral and temporal structures by means of parts-based representation of objects with non-negativity of the data. NMF approximates \mathbf{X} as a linear combination of the K most relevant components (rank) by means of the product of a non-negative basis matrix $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ and a non-negative activation matrix $\mathbf{H} \in \mathbb{R}_+^{K \times T}$. Each column or basis vector W_i represents the i -th spectral pattern associated to physical properties of sounds active in the input spectrogram while each row or activation vector H_i reports the time intervals in which each basis W_i is active. Next, the NMF factorization is detailed in Eq. (2),

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{W}\mathbf{H} \quad (2)$$

$\hat{\mathbf{X}}$ being the estimated spectrogram and it is well-known the choice $K(F+T) \ll FT$ to reduce the dimensionality of the data.

Focusing on the NMF decomposition, it is performed by minimizing a cost function $D(\mathbf{X}|\hat{\mathbf{X}})$ that measures the difference between the input and the estimation so, it ensures the nonnegativity of the bases and activations applying a gradient descent algorithm based on multiplicative update rules [79] which are obtained calculating the partial derivatives of the cost function associated to a given parameter \mathbf{Z} as follows,

$$\mathbf{Z} = \mathbf{Z} \odot \frac{\left[\frac{\partial D(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{Z}} \right]^-}{\left[\frac{\partial D(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{Z}} \right]^+} \quad (3)$$

Several cost functions [80, 81] have been previously used in sound processing such as Euclidean distance, the Itakura–Saito divergence and the Cauchy distribution, however, in this work we propose to minimize the generalized Kullback-Liebler divergence $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$ according to its promising results [71, 75, 76, 10].

$$\begin{aligned} D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) &= \mathbf{X} \log \frac{\mathbf{X}}{\hat{\mathbf{X}}} - \mathbf{X} + \hat{\mathbf{X}} = \\ &= \mathbf{X} \log \frac{\mathbf{X}}{\mathbf{W}\mathbf{H}} - \mathbf{X} + \mathbf{W}\mathbf{H} \end{aligned} \quad (4)$$

The update process of the matrices \mathbf{W} and \mathbf{H} can be seen in Eq. (5) and Eq. (6),

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{((\mathbf{W}\mathbf{H})^{-1}\mathbf{X})\mathbf{H}^T}{\mathbf{1}\mathbf{H}^T} \quad (5)$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T((\mathbf{W}\mathbf{H})^{-1}\mathbf{X})}{\mathbf{W}^T\mathbf{1}} \quad (6)$$

where \mathbf{W} and \mathbf{H} are randomly initialized with positive values, $\mathbf{1}$ is the all-ones matrix, \odot and division represent the element-wise product and division.

The main drawback of NMF is that it calculates the reconstruction of the input spectrogram without guaranteeing physical meaning of the factorized components [82] or even factorizing components that are the result of mixing different parts of several sound sources. An alternative to overcome this problem is to develop a constrained NMF that incorporates prior information into the factorization process by means of additional constraints (see Section 2.3) that help to converge in a better solution.

2.3. Orthogonal Non-Negative Matrix Factorization

Orthogonal Non-negative Matrix Factorization (ONMF) is a constrained NMF that can be considered as the K-means algorithm, where each basis vector W_i or activation vector H_i just correspond to clustering centroids [83]. ONMF has demonstrated to be a powerful tool applied to sound source separation [84, 85] and detection [75] since it minimizes the redundancy between the components of the basis matrix \mathbf{W} , the activations matrix \mathbf{H} or both corresponding to a unique sparse area in the solution region, which learns the most distinct parts [86]. Considering as an example the orthogonality $\phi(\mathbf{W})$ applied to bases, $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ must be fulfilled, in other words, $\phi(\mathbf{W}) = \mathbf{W}^T\mathbf{W} - \mathbf{I}$ must be minimized being T the transpose operator and \mathbf{I} the identity matrix. Incorporating $\phi(\mathbf{W})$ into the previous NMF factorization procedure, the global objective function $D(\mathbf{X}|\hat{\mathbf{X}})$ is obtained in Eq. (7) where the basis matrix \mathbf{W} and the activation matrix \mathbf{H} are computed using a gradient descent algorithm based on multiplicative update rules [87, 88] until the convergence of the factorization after I iterations.

$$\begin{aligned} D(\mathbf{X}|\hat{\mathbf{X}}) &= D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) + \lambda\phi(\mathbf{W}) = \\ &= D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) + \lambda\frac{1}{2}\text{Trace}(\mathbf{W}^T\mathbf{W} - \mathbf{I}) \end{aligned} \quad (7)$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \sqrt{\frac{\mathbf{X}\mathbf{H}^T}{\mathbf{W}\mathbf{W}^T\mathbf{X}\mathbf{H}^T}} \quad (8)$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T\mathbf{X}}{\mathbf{W}^T\mathbf{W}\mathbf{H}} \quad (9)$$

where \mathbf{W} and \mathbf{H} are randomly initialized with positive values, λ controls the importance of the orthogonality constraint and the operator Trace computes the sum of diagonal elements of the square matrix $\mathbf{W}^T\mathbf{W}$. As a result, each basis is orthogonal to the rest of them, that is $W_i \perp W_j, i \neq j$. In this manner, ONMF improves the clustering performance since it factorizes a wider set of true structures that can be found in the input spectrogram [89].

2.4. CNN architecture

Convolutional Neural Networks (CNNs) represent a category of sophisticated neural network architectures that have significantly transformed the landscape of image and audio processing tasks [90, 91, 92, 93, 94, 95]. CNNs are inspired

by the visual cortex of animals, which uses a hierarchical organization of receptive fields to process visual information. Specifically, CNNs are designed to handle large-scale image data extracting meaningful features from images while preserving the spatial and temporal dependencies inherent in such data by employing a set of convolutional layers, pooling layers, and fully connected layers, as follows:

Convolutional layers are the core building blocks of CNNs and consist of a set of filters \mathbf{S} that convolve with the input image to extract local spatial features. These filters slide over the image in a step-wise manner, and the output of each filter is then passed typically through the most used non-linear activation functions $\sigma(\cdot)$, enhancing the model's generalization capabilities by enabling it to learn and recognize patterns in the input data more effectively, such as, the Sigmoid function, the hyperbolic tangent (Tanh) function, the Rectified Linear Unit (ReLU) function and the Leaky ReLU function [90, 93]. Mathematically, this can be represented as follows:

$$C_{i,j,k} = \sigma\left(\sum_{m=1}^M \sum_{n=1}^N I_{i+m-1,j+n-1} \cdot S_{m,n,k} + b_k\right) \quad (10)$$

Here, $C_{i,j,k}$ represents the output of the k^{th} feature map at position (i, j) . $I_{i+m-1,j+n-1}$ represents the input image at position $(i + m - 1, j + n - 1)$, and $S_{m,n,k}$ represents the weight of the k^{th} filter at position (m, n) . b_k represents the bias term of the k^{th} filter.

Pooling layers are typically used after convolutional layers to downsample the feature maps and reduce the dimensionality of the representation. These layers perform a spatial aggregation of nearby features and can be implemented using max-pooling or average-pooling operations. Considering the most common type of pooling (max-pooling) can be expressed as follows:

$$M_{i,j,k} = \max_{(m,n) \in R_{i,j}} C_{i+m-1,j+n-1,k} \quad (11)$$

where $M_{i,j,k}$ represents the output of the k^{th} feature map at position (i, j) after max-pooling. $R_{i,j}$ represents the pooling region around the position (i, j) , and typically has a size of 2x2 or 3x3 [90, 92].

Fully connected (FC) layers are typically added at the end of the CNN to perform the final classification or regression task. These layers are similar to those used in traditional neural networks and consist of a set of neurons that connect every input to every output allowing for a more nuanced interpretation of the data. The mathematical equation for a fully connected layer can be represented as:

$$Y_p = f\left(\sum_{i=0}^{N-1} G_{i,p} D_i + b_p\right) \quad (12)$$

where $f(\cdot)$ denotes the non-linear activation function, N is the number of input neurons, Y_p represents the output of the p^{th} neuron, $G_{i,p}$ is the weight of the connection between

the i^{th} input and the p^{th} neuron, D_i is the i^{th} input, and b_p is the bias term for the p^{th} neuron.

Different CNN architectures have been applied in the field of image, audio and biomedicine as previously mentioned in Section 1 and Section 2.4, where some of the classical CNN architectures that are still widely used as a benchmark for classification tasks are described below:

- **LeNet-5** [96] was one of the first CNN architectures proposed by Yann LeCun in 1998 for handwritten digit recognition. It consists of two convolutional-pooling layers, followed by two fully connected layers, and achieves an accuracy equals 99.2% on the MNIST¹ dataset. However, its limited depth and small receptive field reduce its ability to learn complex features and patterns in large databases.
- **AlexNet** [97], proposed by Krizhevsky et al. in 2012, was the first CNN to win the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [98]. It consists of five convolutional layers, three pooling layers, and three fully connected layers, confirming the effectiveness of deep CNNs and the use of ReLU activation functions. However, AlexNet exhibits some remarkable drawbacks such as, the high computational cost to train and consequently, the high risk of overfitting due to its large number of parameters.
- **VGG16** [99], proposed by Simonyan and Zisserman in 2014, has a very deep architecture consisting of 16 convolutional layers, five pooling layers, and three fully connected layers. VGG16 demonstrated the importance of small filter sizes and the use of max-pooling layers for down-sampling. However, VGG16 has also a high computational cost due to its large number of parameters so, it can make challenging to train the model on resource-constrained devices or with limited memory in order to achieve real-time applications. Finally, VGG16 can suffer overfitting since it has a large number of parameters which can converge to poor generalization performance limiting the ability of the architecture to be applied to new and unseen data.
- **ResNet50** [100], proposed by He et al. in 2016, introduced the concept of residual connections, which allow the network to learn residual functions instead of directly learning the underlying mapping. ResNet50 consists of 50 layers and achieved state-of-the-art performance on the ILSVRC² and COCO³ datasets. ResNet50 demonstrated the importance of depth and residual connections in CNNs. Nevertheless, ResNet50 exhibits notable limitations, particularly with regards to memory requirements and interpretability of the learned features. The high memory requirements arise from the large number of layers,

¹<http://yann.lecun.com/exdb/mnist/>

²<http://www.image-net.org/>

³<https://cocodataset.org/>

Network	Conv Layers	Pool Layers	FC Layers	Activation	Parameters	Accuracy	Database
LeNet-5 [96]	2 (5x5)	2 (2x2)	2	Tanh	60k	99.2%	MNIST
AlexNet [97]	5 (11x11, 5x5, 3x3)	3 (3x3, 2x2)	3	ReLU	60M	84.7%	ILSVRC
VGG16 [99]	13 (3x3)	5 (2x2)	3	ReLU	138M	92.7%	ILSVRC
ResNet50 [100]	50 (7x7, 3x3, 1x1)	1 (3x3)	1	ReLU	25.5M	92.5%	ILSVRC
GoogLeNet [101]	22 (1x1, 3x3, 5x5)	3 (3x3, 2x2)	2	ReLU	6.8M	93.3%	ILSVRC

Table 1

A detailed description, with a specific emphasis on their fundamental layers, relevant parameters, performance metric and databases used for evaluation, of some classical CNN architectures.

which necessitates a significant amount of memory to store the activations and gradients during the training phase. In addition, the high number of layers hinders the interpretation of the learned features, which can be crucial for tasks such as feature selection and transfer learning. These challenges limit the ability to gain insights into the underlying patterns and structure of the data.

- **GoogLeNet** [101], proposed by Szegedy et al. in 2015, introduced the concept of Inception modules, which consist of multiple convolutional filters with different kernel sizes and pooling operations. GoogLeNet achieved state-of-the-art performance on the ILSVRC dataset and demonstrated the effectiveness of using multiple filter sizes in a single layer. However, GoogLeNet can generate significant computational costs during the training phase mainly due to its complex design, comprising 22 layers with several Inception modules and multiple paths, which increases the probability of encountering leakage gradients during the training process leading to slow convergence and poor performance.

Finally, Table 1 shows a summary of the main characteristics of the classical CNN architectures described above.

3. Proposed method

The majority of biomedical signal processing algorithms that rely on CNNs commonly utilize standard time-frequency representations, such as STFT, wavelet, scalogram, Mel, or log-mel, without considering the physiological behavior of the target signal. In this paper, we propose applying the ONMF approach to extract relevant temporal information from PCG signals in order to improve the performance of abnormal heart sound detection facilitating the learning process of any CNN architecture. The proposed method consists of three stages: (i) Signal processing; (ii) ONMF-based feature extraction; and (iii) CNN architecture. The flowchart of the proposed method is shown in Figure 2, and details are depicted in the following Sections 3.1, 3.2 and 3.3.

3.1. Signal processing

Time-frequency representation using spectrograms obtained through the Short-Time Fourier Transform (STFT)

has been proven to be useful for visualizing the characteristics and behavior of biomedical sounds [75, 76, 10, 77]. In this regard, the procedure described in Section 2.1 has been applied to obtain the magnitude spectrogram \mathbf{X} for each input PCG signal $x(t)$. Specifically, each spectrogram \mathbf{X} describes the temporal evolution of the spectral patterns that composed the input PCG signal, characterizing both in time and frequency the normal or abnormal heart sounds active in the recording.

Next, we propose to apply different band-pass filters on the magnitude spectrogram \mathbf{X} to find the most relevant spectral range for the abnormal heart detection. Therefore, Table 2 reports the cut-off frequencies used to design the band-pass filters assuming that most of the energy of normal and abnormal heart sounds can be found between 20-200 Hz and 200-700 Hz respectively [12, 13]. Figure 3 shows the magnitude spectrogram \mathbf{X} for a PCG signal $x(t)$ composed of normal and abnormal heart sounds analyzing the spectral bands $B_C \in [20 - 700]$ Hz, $B_N \in [20 - 200]$ Hz and $B_A \in [200 - 700]$ Hz, noting some observations: i) Focusing on the spectral band B_A , it is discernible that the amplitude of abnormal sounds is notably greater than the amplitude of normal sounds, which consequently facilitates the differentiation between a subject suffering cardiac abnormalities (as depicted in subfigure 3F) and a healthy subject (as portrayed in subfigure 3C); ii) Most of the normal heart sounds are found in the spectral band B_N , while only a small proportion of abnormal heart sounds are present in this frequency range, which appears to make it difficult to detect the presence of VHD (subfigure 3E) and a healthy PCG signal (subfigure 3B); and iii) Likewise, the spectral band denoted by B_C poses a challenge in discriminating between a PCG signal exhibiting the presence (subfigure 3D) or absence (subfigure 3A) of cardiac disorders, given that the energy distribution of normal and abnormal cardiac sounds is relatively equitable within this range. The process of hyperparameter optimization, as presented in Section 4.4, provides corroborating evidence that the cardiac content located in the spectral band B_A is the optimal frequency range for this detection task.

3.2. ONMF-based feature extraction

We propose to extract the temporal and spectral features of the heart sounds by means of ONMF applied to the magnitude spectrogram \mathbf{X} of the input PCG signal $x(t)$, as described in Section 2.3. As can be seen in Eq. (7), the ONMF approach allows to decompose the input magnitude spectrogram \mathbf{X} into the product of two non-negative

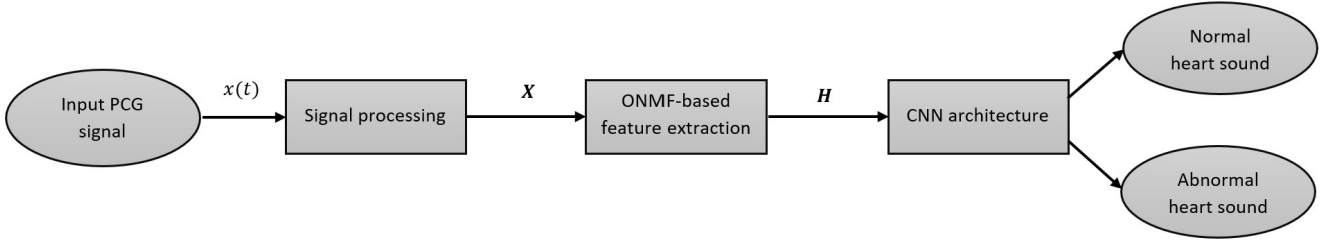


Figure 2: Flowchart of the proposed method applied to the detection of abnormal heart sounds that reveal the presence of VHD.

Band identifier	Frequency band (Hz)	Spectral content
B_F	[20 – 2048]	Full spectral band using a sampling frequency f_s Hz
B_C	[20 – 700]	Most energy of both normal and abnormal heart sounds
B_N	[20 – 200]	Most energy of normal heart sounds
B_A	[200 – 700]	Most energy of abnormal heart sounds

Table 2

Description of the proposed frequency bands.

matrices: basis matrix \mathbf{W} (frequency characteristics) and activation matrix \mathbf{H} (time characteristics). Concretely, \mathbf{W} provides a dictionary composed of the spectral patterns that are active in the input PCG signal and \mathbf{H} stores the temporal activity of the previous spectral patterns over time. The procedure to obtain the basis and activation matrix using the ONMF approach is summarized in Algorithm 1.

Considering the temporal repetitiveness exhibited by the periodic rhythm of a normal heart sounds, we propose to use the ONMF activation matrix \mathbf{H} (temporal features) in combination with a CNN architecture to improve the abnormal cardiac detection. In order to clarify the proposal, Figure 4 shows an example of the ONMF basis \mathbf{W} and activation \mathbf{H} matrices for a PCG signal with and without

Algorithm 1 ONMF

Require: $x(t)$, K , N and I .

- 1: Compute the input PCG magnitude spectrogram \mathbf{X} .
- 2: Normalize the magnitude spectrogram \mathbf{X} using Eq. (1).
- 3: Apply a band-pass filtering.
- 4: Initialize \mathbf{W} and \mathbf{H} with random non negative values.
- 5: Update \mathbf{W} using Equation (8).
- 6: Update \mathbf{H} using Equation (9).
- 7: Repeat steps 5-6 until the algorithm converges (or until the maximum number of iterations I is reached).

return \mathbf{W} and \mathbf{H}

abnormal sounds. On the one hand, both dictionaries \mathbf{W}

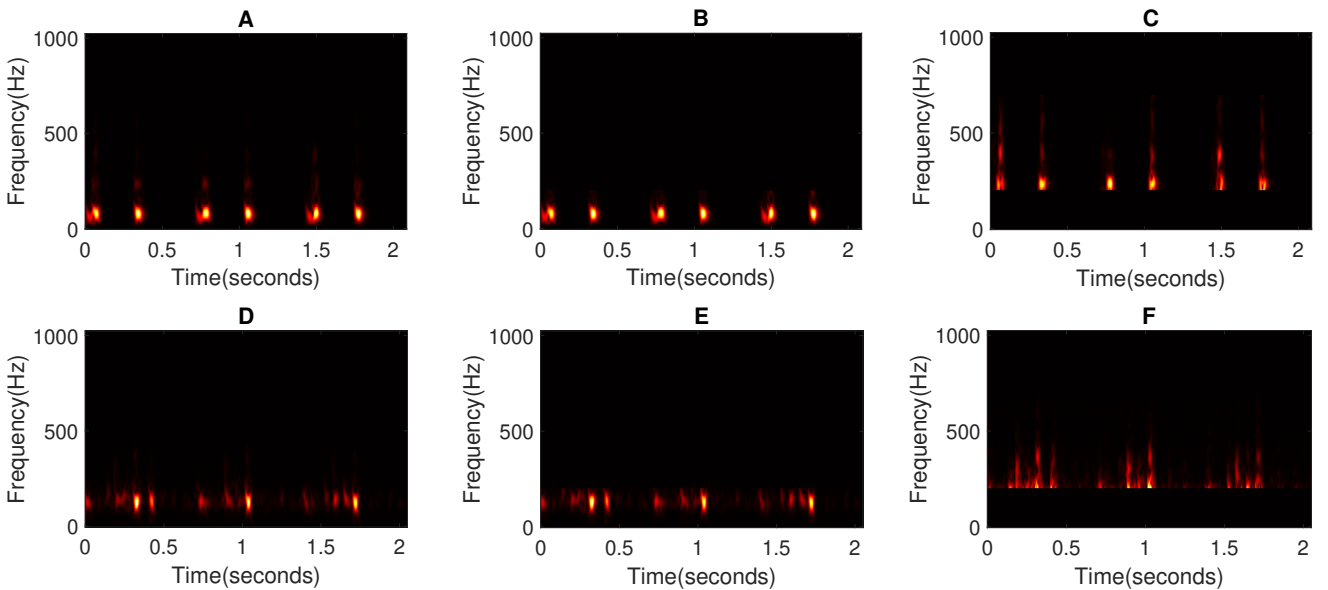


Figure 3: Magnitude spectrogram associated with normal (A, B and C) and abnormal (D, E and F) heart sounds in a PCG signal analyzing the spectral bands B_C (A and D), B_N (B and E) and B_A (C and F).

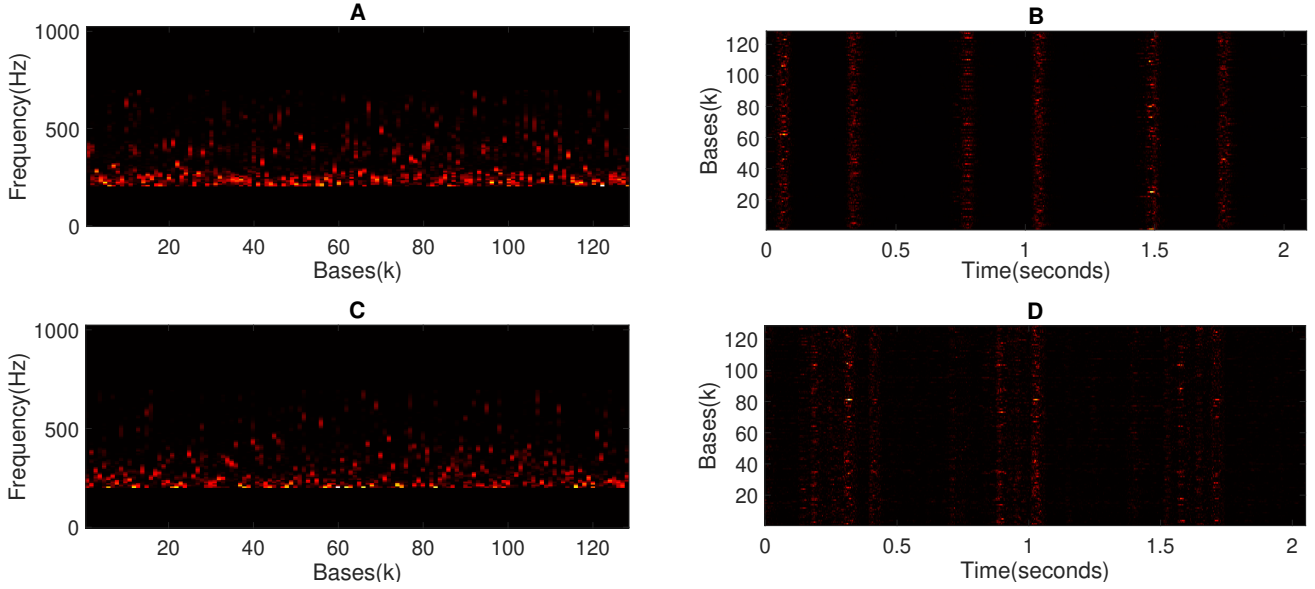


Figure 4: ONMF basis W (A and C) and activation H (B and D) matrices obtained from a PCG signal with normal (A and B) and abnormal (C and D) heart sounds analyzing the spectral band B_A .

show spectral patterns with very similar frequency content, which suggests that the frequency features extracted from the ONMF approach are not relevant to differentiate a PCG signal with cardiac abnormalities (subfigure 4C) from a healthy PCG signal (subfigure 4A). However, the temporal features extracted from the ONMF in the activations matrices H seem to be crucial to determine whether a PCG signal has a VHD (subfigure 4D) or lacks it (subfigure 4B). In this regard, preliminary analysis suggested that temporal information provides clearer insights into cardiac abnormalities by modeling the disruption of temporal repeatability in the periodic rhythm of a normal heart sound since it is widely known that cardiac abnormalities are highly dependent on temporal information encoded between successive heartbeats, as well as the temporal location in the cardiac cycle of the heart sound events S1, systole, S2, and diastole [10].

Unlike the classical NMF method, the ONMF is able to factorize with the most predominant and true spectral structures active in the cardiac signal. This feature reduces the problem of fragmentation of a true cardiac spectral pattern into multiple spectral components. Consequently, ONMF facilitates the factorization of cardiac spectral patterns with greater physiological relevance, a feature that is further demonstrated in its more coherent modelling of the temporal dynamics of cardiac function.

3.3. CNN architecture: UjaNet

As previously mentioned, the main contribution of this paper is to combine the temporal information provided by ONMF with a CNN architecture to facilitate the learning process in the detection between normal and abnormal heart sounds from PCG signals. To this end, we have implemented a simple CNN architecture, denoted as UjaNet, to demonstrate that the enhancement in learning facilitated by feeding

Layer type	Kernel Attribute	Activation
Conv2D	5 × 5 16 Filters	LeakyReLU
MaxPool2D	2 × 2	-
Conv2D	5 × 5 32 Filters	LeakyReLU
MaxPool2D	2 × 2	-
Flatten	-	-
Dense	100 units	LeakyReLU
Dropout	0.5	-
Dense	50 units	LeakyReLU
Dropout	0.5	-
Dense	1 units	Sigmoid

Table 3

A detailed description of the layers and parameters of the proposed UjaNet architecture.

the ONMF activations is also applicable to basic architectures that are relatively simple. Specifically, the architecture UjaNet has been developed based on well-known established guidelines [90, 93], which are: i) at least two convolutional layers to facilitate feature extraction process; ii) the size of the filter kernels should not be too long (not greater than or equal to 5x5) to reduce the number of parameters and computations; iii) a reduction layer after each convolutional layer to reduce the spatial dimensions of the generated feature map reducing the computational cost of the model; iv) use a flattening layer before the dense layers to convert the multidimensional data of the generated feature maps into a feature vector that can be used by the artificial neural network for classification; v) use the dropout regularisation technique to avoid overfitting in the training dataset; and vi) use the sigmoid function in the final output layer for a binary CNN classification model. The guidelines, previously mentioned, are summarised in the Table 3.

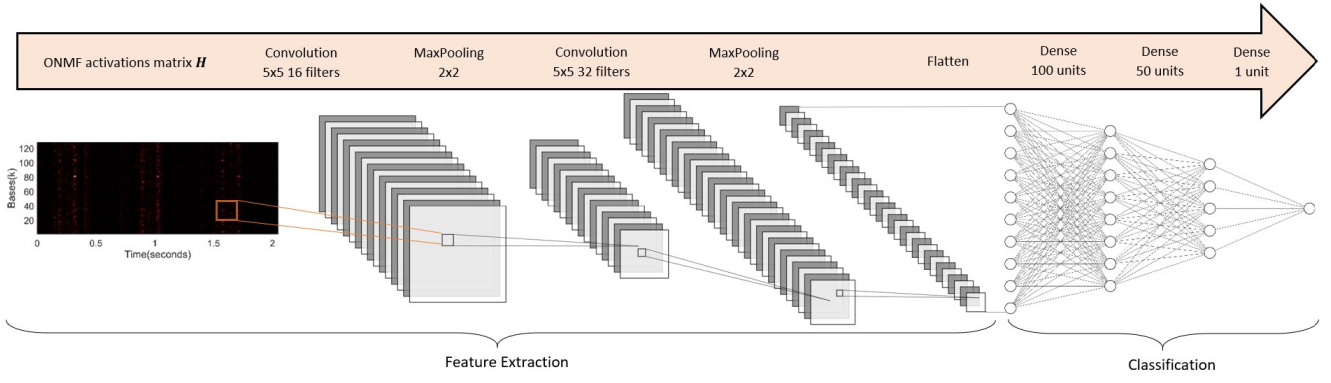


Figure 5: An overview diagram of the proposed method consisting of combining ONMF and CNN architecture.

In summary, Figure 5 shows a complete scheme of the proposal combining the ONMF activation matrix \mathbf{H} with the CNN architecture UjaNet to discriminate between normal and abnormal heart sounds associated to VHD.

4. Experimental results

In this section, an optimization procedure is employed to derive the optimal parameter configuration for the ONMF technique in order to distinguish between normal and abnormal heart sounds. Once the optimal ONMF configuration has been determined, the performance of a set of standard CNN architectures will be evaluated using two different data inputs: i) from the ONMF activations; ii) from the STFT spectrogram. In this manner, the results of this evaluation will be used to determine the most effective input data to improve the learning in the abnormal heart sounds detection using CNN architectures.

4.1. Dataset

In this study, the publicly accessible database D_γ ⁴ was utilized for evaluating the proposed method. This dataset has been widely used for detection and classification of VHD in several previous studies [38, 54, 102, 6]. It contains 1000 PCG signals obtained from multiple sources, including 2 books and 48 websites [38], and has a resolution of 16-bits per sample, a sampling frequency of 8 kHz, and a bit rate of 128 kbps. Each recording is approximately 3 seconds long, covering at least three cardiac cycles. The dataset was divided into two groups: a normal set composed of 200 audio samples from healthy individuals (20% of the total) and an abnormal set of 800 audio samples from patients with one of the following four major category VHDs: aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS), or mitral valve prolapse (MVP) (80% of the total, with 200 audio samples per type of VHD). A more detailed description of the dataset can be found in Table 4.

⁴<https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features/blob/master/README.md>

Patient's condition	VHD	Total recordings
Healthy (normal sound)		200
	Aortic stenosis (AS)	200
Unhealthy (abnormal sound)	Mitral regurgitation (MR)	200
	Mitral stenosis (MS)	200
	Mitral valve prolapse (MVP)	200
		200
Total		1000

Table 4

Summary of the evaluated database D_γ [38], detailing type of patient associated to normal or abnormal heart sounds, type of VHD and number of recordings per type of VHD.

4.2. Experimental setup

Next, we will outline the initialization of the parameters utilized in the three blocks of the proposed method: signal processing, feature extraction, and CNN classifier.

- **Signal processing.** In a preliminary study, the parameters used were based on prior work [10] as they demonstrated the optimal balance between classification performance and computational efficiency: a sampling rate $f_s = 4096$ Hz, a Hamming window with $N=128$ samples length, 25% overlap (resulting in a temporal resolution of 7.8 ms) and a Discrete Fourier Transform (DFT) using 2N points (yielding a frequency resolution of 4 Hz).
- **ONMF-based feature extraction.** The convergence both the NMF and ONMF approaches was empirically observed after 60 iterations for each signal. As a result, the number of iteration I was set to 60. The parameters involved in the decomposition process, including the type of factorization (NMF or ONMF), the basis matrix \mathbf{W} , the activation matrix \mathbf{H} , and the rank or number of components K , were computed through a comprehensive analysis based on a hyperparametric optimization, as described in Section 4.4.
- **CNN architecture.** A 10-fold cross-validation methodology has been implemented during the entire training procedure and repeated 5 times to guarantee the robustness of the model, and a consequence, the average results are presented in a similar way as occurs in recent biomedical machine learning works [103]. For

each fold, the database D_γ was divided into training–testing subsets using a 75%-25% distribution, respectively. Once both subsets were defined, 25% of the training subset was used for validation. Considering that the database D_γ is composed of 4 types of VHDs, as can be seen in Table 4, we ensure that each type of VHD was distributed proportionally in each fold. In addition, a total of 30 epochs were used with a batch size of 16, a learning rate of 0.001 and the adaptive data momentum (ADAM) optimization algorithm. Besides, in order to avoid an over-fitting, the early stopping strategy employed during the training was set at 10 consecutive epoch, taking as monitor parameter the validation loss. Finally, the experimental works were applied using Tensorflow and Keras installed on a computer with an intel(R) Core(TM) i9-12900HK CPU @2.9 GHz with 14 core, NVIDIA GeForce RTX 3080Ti GPU and 16 GB RAM.

4.3. Metrics

In order to determine the type of heart sounds, normal or abnormal, the performance of the proposed method has been evaluated using a set of well-known metrics widely used in the field of biomedical signal processing [104, 6, 105, 42, 106, 103]: i) Accuracy (Acc), the ability to correctly identify the type of heart sounds, that is, normal or abnormal; ii) Sensitivity (Sen), the ability to correctly identify abnormal heart sounds; iii) Specificity (Spe), the ability to correctly identify normal heart sounds taking into account the number of misclassified normal heart sounds; iv) Precision (Pre), the ability to correctly identify abnormal heart sounds taking into account the number of misclassified normal heart sounds; and v) Score (Sco) averages the performance obtained by the Sensitivity and Specificity. Although the metric Acc has been chosen as the most relevant measure through the hyperparametric optimization process of the proposed method (see section 4.4) similarly as occurs in [23], the four remainder metrics have also been considered to show specific performances provided by different CNN architectures when are combined with the proposed method based on ONMF (see section 4.5).

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$Sen = \frac{TP}{TP + FN} \quad (14)$$

$$Spe = \frac{TN}{TN + FP} \quad (15)$$

$$Pre = \frac{TP}{TP + FP} \quad (16)$$

$$Sco = \frac{Sen + Spe}{2} \quad (17)$$

For the purpose of these metrics, TP (True Positive) represent the number of abnormal heart sounds correctly detected, TN (True Negative) represent the number of normal heart sounds correctly detected, FP (False Positive)

represent the number of normal heart sounds misclassified as abnormal heart sounds, and FN (False Negative) represent the number of abnormal heart sounds misclassified as normal heart sounds.

In the training/testing scheme used in this paper, the confusion matrix has been generated for each fold sequentially. In addition, after all iterations we have calculated the average confusion matrix. In this sense, all metrics have been obtained from the average confusion matrix in order to show the overall detection results and from the confusion matrix of each fold to show the variability of the cross-validation methodology used.

4.4. NMF optimization

The use of the proposed ONMF approach is based on a range of parameters that can significantly impact the performance of the factorization in order to ensure the accurate and effective detection of abnormal heart sounds. To address this, a hyperparametric optimization process has been implemented to determine the best combination of parameters that will improve the learning of the CNN architecture. The optimization process evaluated four key parameters: i) The type of non-negative matrix factorization in order to improve the extraction of spectro-temporal features from the input signal. In this case, the optimization analyzes the performance of the unconstrained NMF with the constrained NMF based on orthogonality, that is, ONMF; ii) The input data fed into the CNN architecture. The factorized information can be derived from the spectral domain represented by the basis matrix \mathbf{W} by means of spectral patterns and the temporal domain represented by the activations matrix \mathbf{H} which provides temporal features; iii) The bandpass filtering applied to the input signal, specifically, four different filtering ranges have been considered: $B_F \in [20 - 2048]$ Hz (full spectral band), $B_C \in [20 - 700]$ Hz (spectral band with most energy of both normal and abnormal heart sounds), $B_N \in [20 - 200]$ Hz (spectral band with most energy of normal heart sounds), and $B_A \in [200 - 700]$ Hz (spectral band with most energy of abnormal heart sounds); and finally, iv) The rank or number of components K used to factorize the spectral and temporal content of the input signal, that is, $K = [16, 32, 64, 128, 256, 512]$.

Figure 6 shows the performance results obtained from the hyperparametric optimization previously mentioned to discriminate normal and abnormal cardiac sounds:

Focusing on the type of non-negative matrix factorization and comparing Figure 6A-6C and Figure 6B-6D, it can be observed the superiority of the ONMF approach over NMF. The best configuration for both approaches, Figure 6D, was achieved by using the activations \mathbf{H} with the bandpass filter B_A and $K = 128$ components, resulting in approximately 25% improvement in performance for ONMF compared to NMF (see Figure 6B). This is attributed to the use of a constrained NMF based on the orthogonality incorporated in the spectral domain because is able to model the most distinct true cardiac spectral structures hidden in the input signal avoiding to split a true heart spectral pattern into

multiple spectral components as is often the case with NMF so. As a consequence, ONMF factorizes cardiac spectral patterns with more physiological significance which also implies a more physiological significance of the temporal cardiac patterns as it is in line with the results of the following paragraph.

Focusing on the input data fed into the CNN architecture and comparing Figure 6A-6B and Figure 6C-6D, using only temporal information from the activation matrix \mathbf{H} (see Figure 6D) results in a 45% improvement compared to using only spectral information from the basis matrix \mathbf{W} as shown in Figure 6C. The reason seems to be that temporal information more clearly reveals cardiac abnormalities by modeling the break in the temporal repeatability exhibited by the periodic rhythm of a normal heart sound since it is well-known that cardiac abnormalities depend significantly on the temporal information encoded between consecutive heartbeats, as well as the temporal location of the beats in the cardiac cycle (S1, systole, S2 and diastole). Therefore, temporal information can be considered more representative for identifying cardiac abnormalities compared to spectral information, as the spectral behavior of both normal and abnormal heart sounds exhibit similar patterns, such as smoothness across the frequency domain.

Focusing on the bandpass filtering applied to the input signal and comparing all subfigures in Figure 6, it is found the optimal bandpass filtering in the frequency range B_A in which a high percentage of the abnormal heart sounds often exhibit the highest energy. Specifically, Figure 6D indicates a noticeable improvement in detection performance of 18%, 10%, and 9% compared to the band B_F , B_C , and B_N , and are also in concurrence with the findings from previous studies that have analyzed comparable spectral bands [107, 108, 109]. The higher performance in the band B_A seems due to the higher level of distinguishable information it offers for identifying abnormal heart sounds, as normal heart sounds have less energy in this frequency range, resulting in improved detection compared to other frequency bands. In contrast, the least restrictive band, B_F , obtains the lowest performance due to the presence of non-cardiac sounds above 700Hz, which may lead to confusion in the learning of the CNN architecture for detecting cardiac abnormalities. Bands B_C and B_N exhibit similar results (differing by only 1% in terms of average accuracy), with B_N containing most of the normal heart sounds and a non-significant proportion of abnormal heart sounds. Similarly, the band B_C hinders the analysis of the cardiac temporal repetitive pattern as the energy distribution of normal and abnormal heart sounds is more balanced.

Focusing on the rank of the factorization model, all subfigures in Figure 6 report the best performance when the number of components, K , is between 32 and 128, with the best detection observed at $K = 128$. The use of a low number of components, $K \leq 32$, in the decomposition process is insufficient for capturing all the spectro-temporal characteristics of cardiac sounds, showing a abrupt drop of more than 48% (see Figure 6D) compared to the optimal

value. Conversely, a high number of components $K \geq 256$ factorizes spectral patterns that lack of physiological significance, as the cardiac spectral structures are fragmented into multiple components rather than being represented in individual components, displaying a decline in performance with a drop of more than 24% in Figure 6D.

In short, the hyperparametric optimization reports that the optimal performance for detecting normal and abnormal heart sounds is achieved by using only the temporal features factorized in the ONMF activation matrix with $K = 128$ components jointly analyzing the content located in the frequency range B_A as it often shows most of the relevant content found in most of abnormal heart sounds that characterize major VHD such as those considered in this work.

4.5. Results and comparative evaluations

The optimized ONMF proposal is used in order to assess the performance using CNN classifiers, in comparison to a conventional time-frequency representation, specifically, the Short Time Fourier Transform (STFT) spectrogram due to its relevance and efficiency demonstrated in previous works in the field of biomedical signal processing [75, 76, 10, 77, 103]. Several CNN architectures have been evaluated to analyze the robustness of the proposed method with different machine learning models. Hereafter, the results obtained using the STFT spectrogram are referred to as STFT, while those obtained from the hyperparametric optimization are referred to as ONMF. Indicate to the reader that the STFT spectrograms were filtered in the optimal frequency range B_A to ensure fairness in the cardiac detection assessment.

The proposed method for detecting abnormal heart sounds was evaluated using six CNN architectures - UjaNet, LeNet5, AlexNet, ResNet50, VGG16, and GoogLeNet - as illustrated in Figure 7. The assessment was based on the detection performance, which was determined by inputting both the STFT spectrogram and the ONMF activation matrix into the classifiers. Each box plot represents 50 data points, which correspond to a 10-fold cross-validation of the testing portion of the database D_γ . The lower and upper lines of each box represent the first and third quartiles, respectively, while the line in the center of each box represents the median. The diamond shape in the center of each box represents the average value. The lines extending above and below each box represent the extent of the remaining samples, excluding outliers, which are defined as points over 1.5 times the interquartile range from the median and are depicted as crosses. The results demonstrate that the use of ONMF temporal information followed by a CNN classifier significantly improves the detection performance compared to using the STFT spectrogram in cascade with a CNN classifier, regardless of the metric or classifier used. Specifically, the average accuracy results obtained using the ONMF activations outperform those obtained from the STFT spectrogram by 35%, 30%, 23%, 10%, 8%, and 6% for UjaNet, LeNet5, AlexNet, ResNet50, VGG16, and GoogLeNet (see Figure 7A). Additionally, all subfigures of the Figure 7 demonstrate the promising robustness of the proposed approach for

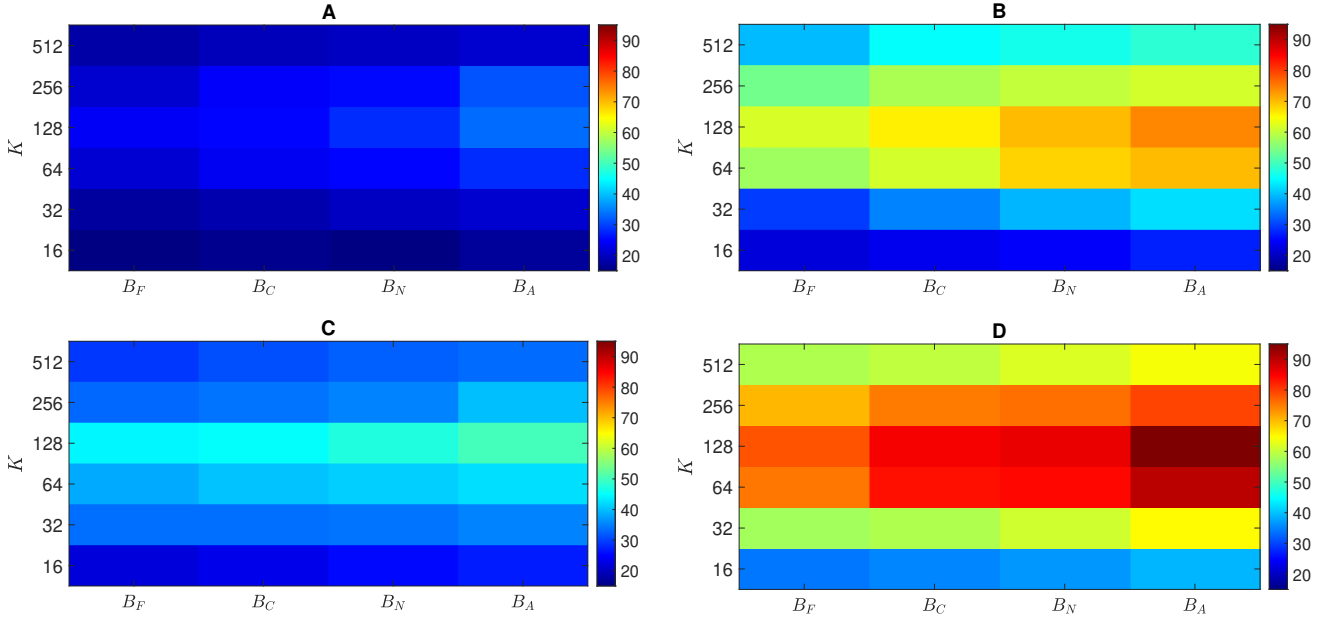


Figure 6: Average accuracy results provided by the hyperparametric optimization process considering as input to the CNN network UjaNet the matrix of bases \mathbf{W} or activations \mathbf{H} of the NMF or ONMF approach: subfigure **A** (NMF- \mathbf{W}), subfigure **B** (NMF- \mathbf{H}), subfigure **C** (ONMF- \mathbf{W}) and subfigure **D** (ONMF- \mathbf{H}). Each pixel represents the metric $Acc(\%)$ obtained from the average confusion matrix associated with the 10-fold cross-validation of the database D_γ . The rows represent the number of components K of the NMF or ONMF factorization process and the columns represent each bandpass filter applied (B_F , B_C , B_N or B_A).

each metric being the average accuracy results obtained by ONMF very similar for all the evaluated classifiers and range from 93% (AlexNet) to 98% (GoogLeNet). Moreover, the use of temporal information from the ONMF approach significantly improves the detection performance of UjaNet, LeNet5, and AlexNet, which were not competitive using the STFT spectrogram. In fact, using temporal information from ONMF, UjaNet, LeNet5, and AlexNet not only become competitive but also show behavior similar to that of more complex networks such as ResNet50. This fact reports that the ONMF temporal characteristics play a crucial role in detecting the presence of cardiac abnormalities emphasizing the importance of using an appropriate representation of input data in the learning process of a CNN model as occurs in recent works [110, 103]. Focusing on Figures 7B and 7C, outcomes also indicate that the detection performance using the ONMF approach is higher for all the evaluated CNN architectures. However, it should be noted that this higher sensitivity, Figure 7B, does not necessarily mean that the proposed approach prioritizes the detection of abnormal heart sounds at the expense of reducing the correct diagnosis of healthy patients (see Figure 7C) since it must be a trade-off that should be evaluated by medical physicians based on the specific application and the requirements. Figure 7B shows that the ONMF proposal improves the reliability in the diagnosis of the presence of cardiac abnormalities and provides a lower dispersion, specifically, a dispersion between 13% (VGG16) and 40% (AlexNet) when using the STFT spectrogram, and between 2% (GoogLeNet) and 10% (AlexNet) when using the ONMF approach. This fact

suggests that the ONMF approach improves the training of the CNN models, making the detection less affected by the subset of the database used for training, in other words, the ONMF approach successfully generalizes the behavior of abnormal heart sounds, regardless of the subset of the database used for training.

Figure 8 shows the average confusion matrix generated after all iterations composing the 10-fold cross-validation process for all CNN networks evaluated using the STFT spectrogram and the ONMF activations. Specifically, it reports that the improvement produced by the ONMF model over the STFT spectrogram is higher in terms of FN compared to FP for all evaluated networks so, this leads to a more accurate and more patient-protective diagnosis as the CNN classifier will miss fewer unhealthy patients and thus minimize the number of undetected VHD patients. The results show that the ONMF model improves the performance of the CNN network by adapting its training to better reflect the heart sound patterns, which clearly translates into a decrease in the loss of cardiac anomalies (FN) and occurrence of false heart anomalies (FP) at the expense of increasing the correct detection of normal (NT) and abnormal (TP) cardiac events regardless of the complexity and capacity of the classifier architecture.

5. Conclusions and Future work

This work proposes a combination of factorization and Convolutional Neural Network (CNN) to improve the detection of the presence of VHDs. The proposal combines the temporal information from heart sounds, obtained through

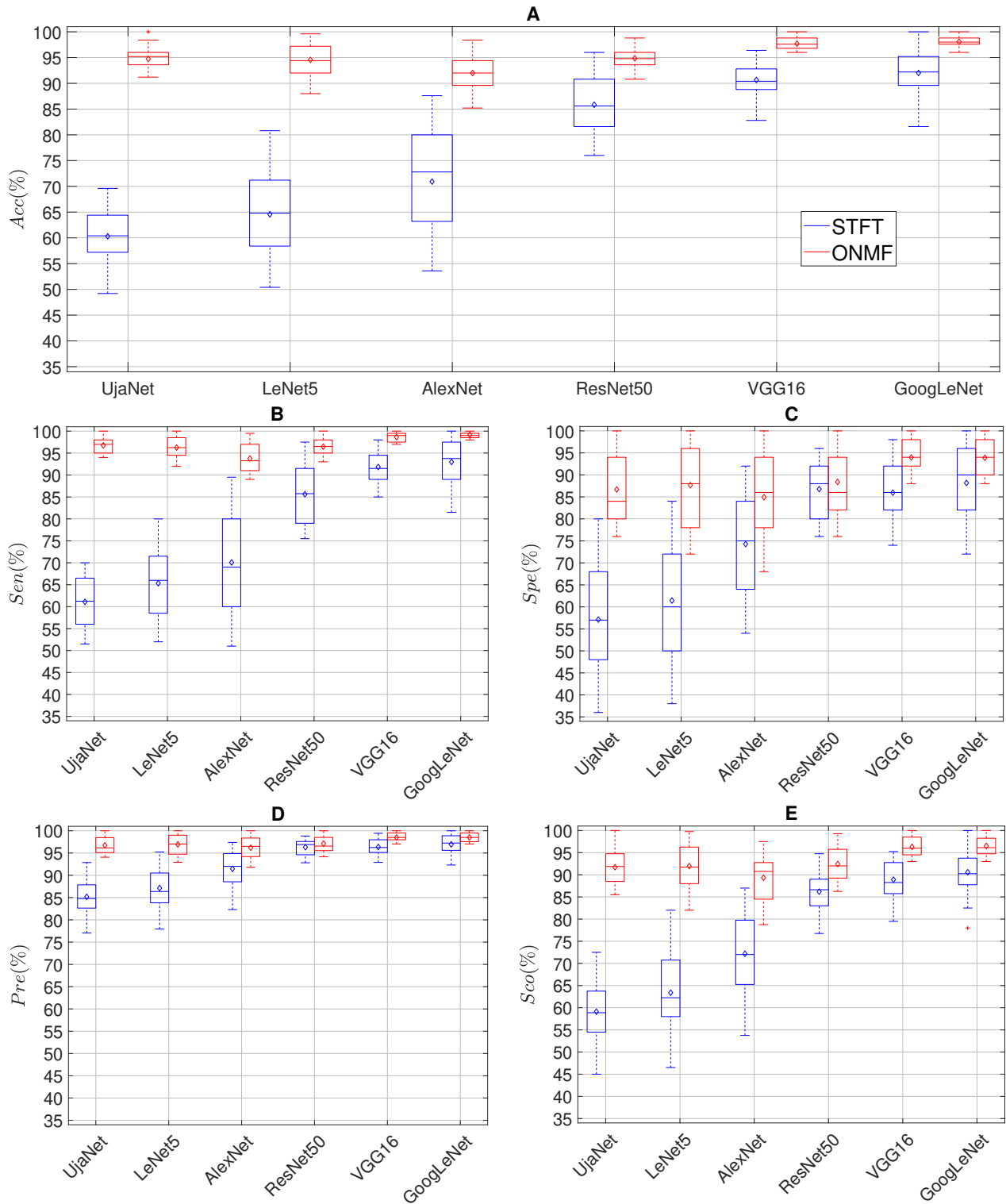


Figure 7: Results related to *Acc* (A), *Sen* (B), *Spe* (C), *Pre* (D) and *Sco* (E) evaluating the database D_γ for different CNN networks in which either the STFT spectrogram or the activation matrix of the ONMF approach is being used as input data.

the temporal activations by means of Orthogonal Non-Negative Matrix Factorization (ONMF), followed by a CNN architecture to achieve better differentiation between normal and abnormal cardiac events from phonocardiogram (PCG) signals. The increasing the training reliability of

the CNN can be attributed to the use of an ONMF in the spectral domain since it models the most distinct true cardiac spectral structures in the input signal, thereby avoiding the fragmentation of true heart spectral patterns into multiple spectral components, which is often observed in traditional

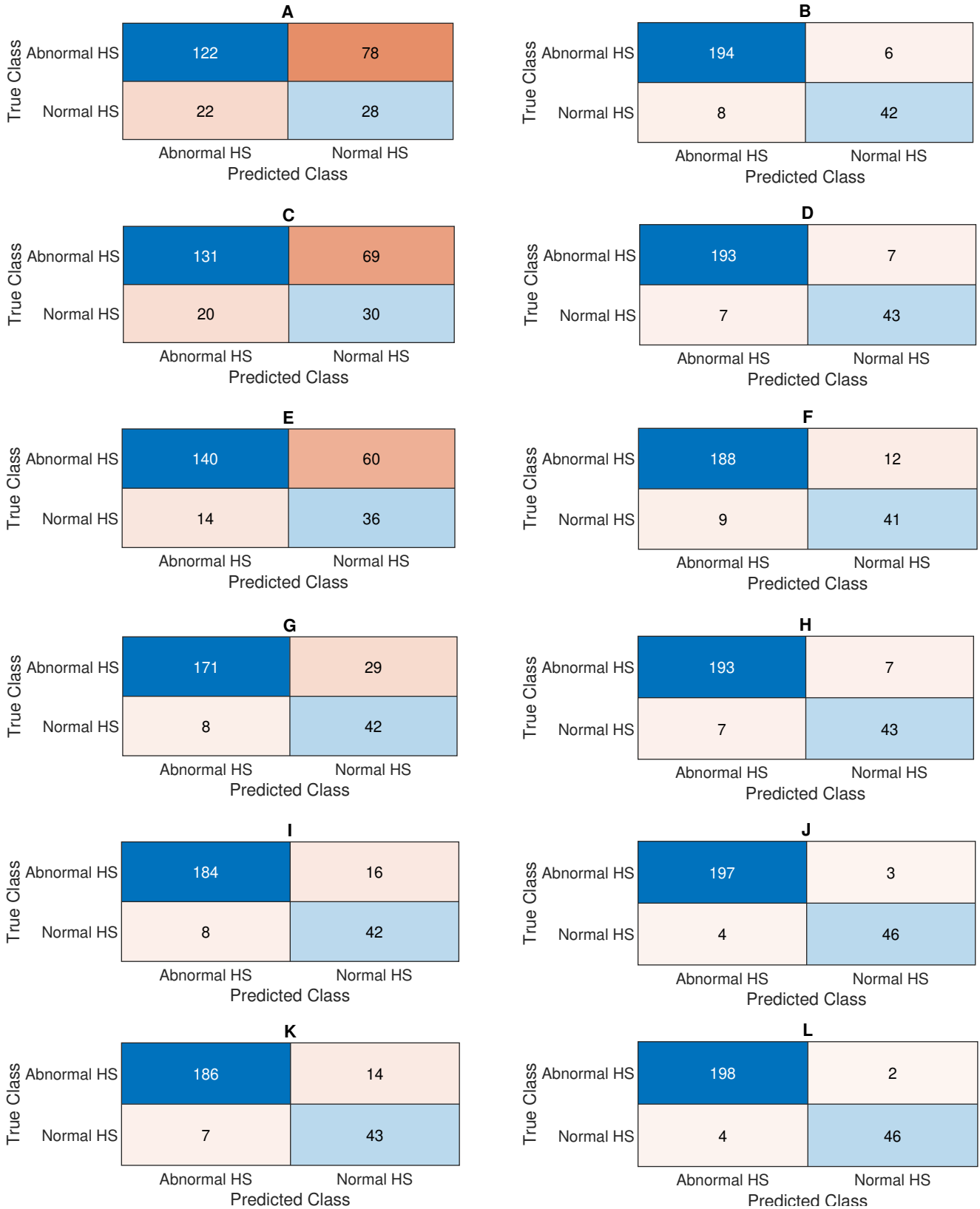


Figure 8: Average confusion matrix in the heart detection performance evaluating the database D_γ for different CNN architectures: UjaNet (A and B), LeNet5 (C and D), AlexNet (E and F), ResNet50 (G and H), VGG16 (I and J) and GoogLeNet (K and L). The left column represents the scenario in which the input is the STFT spectrogram while the right column indicates the scenario in which the input is the ONMF activations.

NMF methods. As a result, the patterns decomposed by ONMF provide a greater physiological significance, and this is reflected in the resulting temporal cardiac patterns as well.

In order to enhance the performance of the ONMF approach, a hyperparameter optimization process was executed. The results of this process indicated that the optimal configuration for detecting heart abnormalities was achieved by utilizing the ONMF activation matrix with 128 components and a [200-700]Hz bandpass filter as input to the proposed CNN classifier UjaNet.

The data feeding of the CNN classifier using the ONMF activations was evaluated and compared to the performance obtained using the STFT spectrogram as input. The results indicated that: i) The ONMF model demonstrated superior performance compared to the STFT spectrogram for all analyzed CNN architectures. This highlights the importance of incorporating the temporal information provided by the ONMF approach in order to effectively distinguish between normal and abnormal heart sounds, as the repetitive pattern present in normal heart cycles is disrupted in the presence of abnormal heart sounds; ii) The utilization of ONMF activations as inputs to CNNs results in low-complexity CNN models, such as LeNet5 or UjaNet, achieving comparable performance to more complex CNN models, such as VGG16 or GoogLeNet. This highlights the critical role that ONMF temporal features play in the detection of VHDs, emphasizing the significance of employing an appropriate representation of input data in the training process of a CNN model, as observed in prior studies in the domain of biomedical signal processing as previously detailed.

Future work will focus on two challenging topics. Firstly, the development of novel time-frequency representations of input data with the aim of improving the learning capabilities of CNN models while reducing computational requirements and achieving performance comparable to current state-of-the-art CNN models. Secondly, the development of innovative data augmentation strategies based on the time-frequency characteristics of abnormal heart sounds, with the goal of improving the classification performance for major VHDs such as aortic stenosis, mitral stenosis, mitral regurgitation, and mitral valve prolapse.

References

- [1] World Health Organization, Cardiovascular diseases, [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), online. Accessed: 2023-03-20.
- [2] Instituto Nacional de Estadística, Ministerio de Asuntos Económicos y Transformación Digital de España, Defunciones según la Causa de Muerte (Año 2020), https://www.ine.es/prensa/edcm_2020.pdf, online. Accessed: 2023-03-20.
- [3] Health Policy Partnership, Heart valve disease, <https://www.healthpolicypartnership.com/project/heart-valve-disease-2/>, online. Accessed: 2023-03-20.
- [4] World Health Organization, Cardiovascular diseases, https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1, online. Accessed: 2023-03-20.
- [5] Geriatricarea, Las enfermedades cardiovasculares son las responsables del 24,3% del total de defunciones en España, <https://acortar.link/wHm7ZA>, online. Accessed: 2023-03-20.

- [6] M. Alkhdari, L. Fraiwan, Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings, *Computer Methods and Programs in Biomedicine* 200 (2021) 105940.
- [7] T. K. M. Wang, P. Cremer, A. Goyal, W. A. Jaber, Advanced echocardiography in the evaluation of aortic valve disease, *eJ. Cardiol. Pract.* 18 (2020).
- [8] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, et al., American heart association council on epidemiology and prevention statistics committee and stroke statistics subcommittee, Heart disease and stroke statistics-2020 update: a report from the American Heart Association. *Circulation* 141 (9) (2020) e139–e596.
- [9] W. Wang, X. Yu, B. Fang, D.-Y. Zhao, Y. Chen, W. Wei, J. Chen, Cross-modality lge-cmr segmentation using image-to-image translation based data augmentation, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2022).
- [10] J. Torre-Cruz, D. Martínez-Muñoz, N. Ruiz-Reyes, A. Muñoz-Montoro, M. Puentes-Chiachio, F. Canadas-Quesada, Unsupervised detection and classification of heartbeats using the dissimilarity matrix in pcg signals, *Computer Methods and Programs in Biomedicine* 221 (2022) 106909.
- [11] T.-E. Chen, S.-I. Yang, L.-T. Ho, K.-H. Tsai, Y.-H. Chen, Y.-F. Chang, Y.-H. Lai, S.-S. Wang, Y. Tsao, C.-C. Wu, S1 and s2 heart sound recognition using deep neural networks, *IEEE Transactions on Biomedical Engineering* 64 (2) (2016) 372–380.
- [12] M. F. Khan, M. Atteeq, A. N. Qureshi, Computer aided detection of normal and abnormal heart sound using pcg, in: *Proceedings of the 2019 11th international conference on bioinformatics and biomedical technology*, 2019, pp. 94–99.
- [13] R. Banerjee, A. Ghose, A semi-supervised approach for identifying abnormal heart sounds using variational autoencoder, in: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, pp. 1249–1253.
- [14] D. Warriner, J. Michaels, P. D. Morris, Cardiac auscultation: normal and abnormal, *British Journal of Hospital Medicine* 80 (2) (2019) C28–C31.
- [15] S. Chauhan, P. Wang, C. S. Lim, V. Anantharaman, A computer-aided mfcc-based hmm system for automatic auscultation, *Computers in biology and medicine* 38 (2) (2008) 221–233.
- [16] H. Wu, S. Kim, K. Bae, Hidden markov model with heart sound signals for identification of heart diseases, in: *Proceedings of 20th International Congress on Acoustics (ICA)*, Sydney, Australia, 2010, pp. 23–27.
- [17] R. Saraçoğlu, Hidden markov model-based classification of heart valve disease with pca for dimension reduction, *Engineering Applications of Artificial Intelligence* 25 (7) (2012) 1523–1528.
- [18] C. Kwak, O.-W. Kwon, Cardiac disorder classification by heart sound signals using murmur likelihood and hidden markov model state likelihood, *IET signal processing* 6 (4) (2012) 326–334.
- [19] H. Fahad, M. U. Ghani Khan, T. Saba, A. Rehman, S. Iqbal, Microscopic abnormality classification of cardiac murmurs using anfis and hmm, *Microscopy Research and Technique* 81 (5) (2018) 449–457.
- [20] S. K. Ghosh, R. Ponnalagu, R. Tripathy, U. R. Acharya, Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with pcg signals, *Computers in biology and medicine* 118 (2020) 103632.
- [21] J. Vepa, Classification of heart murmurs using cepstral features and support vector machines, in: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE*, 2009, pp. 2539–2542.
- [22] T. Nilanon, J. Yao, J. Hao, S. Purushotham, Y. Liu, Normal/abnormal heart sound recordings classification using convolutional neural network, in: *2016 computing in cardiology conference (CinC)*, IEEE, 2016, pp. 585–588.
- [23] S. Das, S. Pal, M. Mitra, Acoustic feature based unsupervised approach of heart sound event detection, *Computers in Biology and*

- Medicine 126 (2020) 103990.
- [24] D. S. Ediriweera, A. Kasturiratne, A. Pathmeswaran, N. K. Gunawardena, B. A. Wijayawickrama, S. F. Jayamanne, G. K. Isbister, A. Dawson, E. Giorgi, P. J. Diggle, et al., Mapping the risk of snakebite in Sri Lanka—a national survey with geospatial analysis, *PLoS neglected tropical diseases* 10 (7) (2016) e0004813.
 - [25] T. J. Ahmad, H. Ali, S. A. Khan, Classification of phonocardiogram using an adaptive fuzzy inference system., in: *IPCV, Citeseer*, 2009, pp. 609–614.
 - [26] A. Quiceno-Manrique, J. Godino-Llorente, M. Blanco-Velasco, G. Castellanos-Dominguez, Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic signals, *Annals of biomedical engineering* 38 (2010) 118–137.
 - [27] I. Maglogiannis, E. Loukis, E. Zafropoulos, A. Stasis, Support vectors machine-based identification of heart valve diseases using heart sounds, *Computer methods and programs in biomedicine* 95 (1) (2009) 47–61.
 - [28] S. Vernekar, S. Nair, D. Vijaysenan, R. Ranjan, A novel approach for classification of normal/abnormal phonocardiogram recordings using temporal signal analysis and machine learning, in: *2016 computing in cardiology conference (CinC)*, IEEE, 2016, pp. 1141–1144.
 - [29] W. Zhang, J. Han, S. Deng, Heart sound classification based on scaled spectrogram and partial least squares regression, *Biomedical Signal Processing and Control* 32 (2017) 20–28.
 - [30] W. Zhang, J. Han, S. Deng, Heart sound classification based on scaled spectrogram and tensor decomposition, *Expert Systems with Applications* 84 (2017) 220–231.
 - [31] H. Tang, Z. Dai, Y. Jiang, T. Li, C. Liu, Pcg classification using multidomain features and svm classifier, *BioMed research international* 2018 (2018).
 - [32] F. Demir, A. Şengür, V. Bajaj, K. Polat, Towards the classification of heart sounds based on convolutional deep neural network, *Health information science and systems* 7 (2019) 1–9.
 - [33] B. Ergen, Y. Tatar, H. O. Gulcur, Time–frequency analysis of phonocardiogram signals using wavelet transform: a comparative study, *Computer methods in biomechanics and biomedical engineering* 15 (4) (2012) 371–381.
 - [34] Y. Zheng, X. Guo, X. Ding, A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification, *Expert Systems with Applications* 42 (5) (2015) 2710–2721.
 - [35] P. Langley, A. Murray, Heart sound classification from unsegmented phonocardiograms, *Physiological measurement* 38 (8) (2017) 1658.
 - [36] Z. Dokur, T. Ölmez, Heart sound classification using wavelet transform and incremental self-organizing map, *Digital Signal Processing* 18 (6) (2008) 951–959.
 - [37] V. N. Varghees, K. Ramachandran, Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope, *IEEE Sensors Journal* 17 (12) (2017) 3861–3872.
 - [38] G.-Y. Son, S. Kwon, Classification of heart sound signal using multiple features, *Applied Sciences* 8 (12) (2018) 2344.
 - [39] D. Boutana, M. Benidir, B. Barkat, Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis, *IET signal processing* 5 (6) (2011) 527–537.
 - [40] C. D. Papadaniil, L. J. Hadjileontiadis, Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features, *IEEE journal of biomedical and health informatics* 18 (4) (2013) 1138–1152.
 - [41] E. Pretorius, M. L. Cronje, O. Strydom, Development of a pediatric cardiac computer aided auscultation decision support system, in: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, 2010, pp. 6078–6082.
 - [42] O. El Badlaoui, A. Benba, A. Hammouch, Novel pcg analysis method for discriminating between abnormal and normal heart sounds, *Irbm* 41 (4) (2020) 223–228.
 - [43] S. E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, J. J. Struijk, Acoustic features for the identification of coronary artery disease, *IEEE Transactions on Biomedical Engineering* 62 (11) (2015) 2611–2619.
 - [44] W. Zhang, J. Han, S. Deng, Heart sound classification based on scaled spectrogram and partial least squares regression, *Biomedical Signal Processing and Control* 32 (2017) 20–28.
 - [45] G. Petschenka, A. A. Agrawal, How herbivores coopt plant defenses: natural selection, specialization, and sequestration, *Current opinion in insect science* 14 (2016) 17–24.
 - [46] G.-Y. Son, S. Kwon, Classification of heart sound signal using multiple features, *Applied Sciences* 8 (12) (2018) 2344.
 - [47] C. Potes, S. Parvaneh, A. Rahman, B. Conroy, Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds, in: *2016 computing in cardiology conference (CinC)*, IEEE, 2016, pp. 621–624.
 - [48] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, K. Sricharan, Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients, in: *2016 Computing in cardiology conference (CinC)*, IEEE, 2016, pp. 813–816.
 - [49] V. Maknickas, A. Maknickas, Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients, *Physiological measurement* 38 (8) (2017) 1671.
 - [50] S. A. Singh, S. Majumder, M. Mishra, Classification of short unsegmented heart sound based on deep learning, in: *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE, 2019, pp. 1–6.
 - [51] F. Li, H. Tang, S. Shang, K. Mathiak, F. Cong, Classification of heart sounds using convolutional neural network, *Applied Sciences* 10 (11) (2020) 3956.
 - [52] B. Xiao, Y. Xu, X. Bi, J. Zhang, X. Ma, Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption, *Neurocomputing* 392 (2020) 153–159.
 - [53] S. Kiranyaz, M. Zabihi, A. B. Rad, T. Ince, R. Hamila, M. Gabbouj, Real-time phonocardiogram anomaly detection by adaptive 1d convolutional neural networks, *Neurocomputing* 411 (2020) 291–301.
 - [54] N. Baghel, M. K. Dutta, R. Burget, Automatic diagnosis of multiple cardiac diseases from pcg signals using convolutional neural network, *Computer Methods and Programs in Biomedicine* 197 (2020) 105750.
 - [55] J. S. Khan, M. Kaushik, A. Chaurasia, M. K. Dutta, R. Burget, Cardi-net: A deep neural network for classification of cardiac disease using phonocardiogram signal, *Computer Methods and Programs in Biomedicine* 219 (2022) 106727.
 - [56] C. Thomae, A. Dominik, Using deep gated rnn with a convolutional front end for end-to-end classification of heart sound, in: *2016 Computing in Cardiology Conference (CinC)*, IEEE, 2016, pp. 625–628.
 - [57] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, B.-W. On, Heartbeat sound signal classification using deep learning, *Sensors* 19 (21) (2019) 4819.
 - [58] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, H. Fan, Heart sound classification based on improved mfcc features and convolutional recurrent neural networks, *Neural Networks* 130 (2020) 22–32.
 - [59] J.-K. Wang, Y.-F. Chang, K.-H. Tsai, W.-C. Wang, C.-Y. Tsai, C.-H. Cheng, Y. Tsao, Automatic recognition of murmurs of ventricular septal defect using convolutional recurrent neural networks with temporal attentive pooling, *Scientific Reports* 10 (1) (2020) 1–10.
 - [60] B. Ahmad, F. A. Khan, K. N. Khan, M. S. Khan, Automatic classification of heart sounds using long short-term memory, in: *2021 15th International Conference on Open Source Systems and Technologies (ICOSST)*, IEEE, 2021, pp. 1–6.
 - [61] D. R. Megalmani, B. Shailesh, A. Rao, S. S. Jeevannavar, P. K. Ghosh, Unsegmented heart sound classification using hybrid cnn-lstm neural networks, in: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2021, pp. 713–717.
 - [62] D. Chen, W. Xuan, Y. Gu, F. Liu, J. Chen, S. Xia, H. Jin, S. Dong, J. Luo, Automatic classification of normal–abnormal heart sounds

- using convolution neural network and long-short term memory, *Electronics* 11 (8) (2022) 1246.
- [63] J. J. Carabias-Orti, T. Virtanen, P. Vera-Candeas, N. Ruiz-Reyes, F. J. Canadas-Quesada, Musical instrument sound multi-excitation model for non-negative spectrogram factorization, *IEEE Journal of Selected Topics in Signal Processing* 5 (6) (2011) 1144–1158.
- [64] H. Chung, R. Badeau, E. Plourde, B. Champagne, Training and compensation of class-conditioned nmf bases for speech enhancement, *Neurocomputing* 284 (2018) 107–118.
- [65] S. Nie, S. Liang, W. Liu, X. Zhang, J. Tao, Deep learning based speech separation via nmf-style reconstructions, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26 (11) (2018) 2043–2055.
- [66] T. K. Chan, C. S. Chin, Y. Li, Non-negative matrix factorization-convolutional neural network (nmf-cnn) for sound event detection, *arXiv preprint arXiv:2001.07874* (2020).
- [67] A. J. Muñoz-Montoro, J. J. Carabias-Orti, P. Cabañas-Molero, F. J. Cañadas-Quesada, N. Ruiz-Reyes, Multichannel blind music source separation using directivity-aware mnmf with harmonicity constraints, *IEEE access* 10 (2022) 17781–17795.
- [68] W. Liu, K. Yuan, D. Ye, Reducing microarray data via nonnegative matrix factorization for visualization and clustering analysis, *Journal of biomedical informatics* 41 (4) (2008) 602–606.
- [69] N. Kumar, P. Uppala, K. Duddu, H. Sreedhar, V. Varma, G. Guzman, M. Walsh, A. Sethi, Hyperspectral tissue image segmentation using semi-supervised nmf and hierarchical clustering, *IEEE transactions on medical imaging* 38 (5) (2018) 1304–1313.
- [70] T. Aonishi, R. Maruyama, T. Ito, H. Miyakawa, M. Murayama, K. Ota, Imaging data analysis using non-negative matrix factorization, *Neuroscience Research* 179 (2022) 51–56.
- [71] F. Canadas-Quesada, N. Ruiz-Reyes, J. Carabias-Orti, P. Vera-Candeas, J. Fuertes-García, A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds, *Applied Acoustics* 125 (2017) 7–19.
- [72] S. J. Fodeh, A. Tiwari, Exploiting medline for gene molecular function prediction via nmf based multi-label classification, *Journal of Biomedical Informatics* 86 (2018) 160–166.
- [73] N. Dia, J. Fontecave-Jallon, P.-Y. Guméry, B. Rivet, Heart rate estimation from phonocardiogram signals using non-negative matrix factorization, in: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 1293–1297.
- [74] Z. Zeng, A. H. Vo, C. Mao, S. E. Clare, S. A. Khan, Y. Luo, Cancer classification and pathway discovery using non-negative matrix factorization, *Journal of biomedical informatics* 96 (2019) 103247.
- [75] J. D. L. T. Cruz, F. J. C. Quesada, J. J. C. Orti, P. V. Candeas, N. R. Reyes, Combining a recursive approach via non-negative matrix factorization and gini index sparsity to improve reliable detection of wheezing sounds, *Expert systems with applications* 147 (2020) 113212.
- [76] J. De La Torre Cruz, F. J. Cañadas Quesada, N. Ruiz Reyes, S. García Galán, J. J. Carabias Orti, G. Pérez Chica, Monophonic and polyphonic wheezing classification based on constrained low-rank non-negative matrix factorization, *Sensors* 21 (5) (2021) 1661.
- [77] J. D. L. T. Cruz, F. J. C. Quesada, D. Martínez-Munoz, N. R. Reyes, S. G. Galán, J. J. C. Orti, An incremental algorithm based on multichannel non-negative matrix partial co-factorization for ambient denoising in auscultation, *Applied Acoustics* 182 (2021) 108229.
- [78] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (6755) (1999) 788–791.
- [79] D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, *Advances in neural information processing systems* 13 (2000).
- [80] C. Févotte, N. Bertin, J.-L. Durrieu, Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis, *Neural computation* 21 (3) (2009) 793–830.
- [81] A. Liutkus, D. Fitzgerald, R. Badeau, Cauchy nonnegative matrix factorization, in: *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, IEEE, 2015, pp. 1–5.
- [82] C. Laroche, M. Kowalski, H. Papadopoulos, G. Richard, A structured nonnegative matrix factorization for source separation, in: *2015 23rd European Signal Processing Conference (EUSIPCO)*, IEEE, 2015, pp. 2033–2037.
- [83] C. Ding, X. He, H. D. Simon, On the equivalence of nonnegative matrix factorization and spectral clustering, in: *Proceedings of the 2005 SIAM international conference on data mining*, SIAM, 2005, pp. 606–610.
- [84] S. Z. Li, X. W. Hou, H. J. Zhang, Q. S. Cheng, Learning spatially localized, parts-based representation, in: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Vol. 1*, IEEE, 2001, pp. 207–212.
- [85] F. J. Cañadas-Quesada, P. Vera-Candeas, D. Martínez-Munoz, N. Ruiz-Reyes, J. J. Carabias-Orti, P. Cabañas-Molero, Constrained non-negative matrix factorization for score-informed piano music restoration, *Digital Signal Processing* 50 (2016) 240–257.
- [86] Y.-X. Wang, Y.-J. Zhang, Nonnegative matrix factorization: A comprehensive review, *IEEE Transactions on knowledge and data engineering* 25 (6) (2012) 1336–1353.
- [87] C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, in: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 126–135.
- [88] J.-H. Yoo, S.-J. Choi, Nonnegative matrix factorization with orthogonality constraints, *Journal of computing science and engineering* 4 (2) (2010) 97–109.
- [89] E. M. Grais, H. Erdogan, Discriminative nonnegative dictionary learning using cross-coherence penalties for single channel source separation., in: *Interspeech*, 2013, pp. 808–812.
- [90] A. Géron, *Hands-on machine learning with scikit-learn and tensorflow: Concepts, Tools, and Techniques to build intelligent systems* (2017).
- [91] N. Takahashi, Y. Mitsufuji, D3net: Densely connected multi-dilated densenet for music source separation, *arXiv preprint arXiv:2010.01733* (2020).
- [92] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, H. Liu, Attention-guided cnn for image denoising, *Neural Networks* 124 (2020) 117–129.
- [93] F. Chollet, *Deep learning with Python*, Simon and Schuster, 2021.
- [94] C. Hernandez-Olivan, I. Zay Pinilla, C. Hernandez-Lopez, J. R. Beltran, A comparison of deep learning methods for timbre analysis in polyphonic automatic music transcription, *Electronics* 10 (7) (2021) 810.
- [95] A. Chattopadhyay, M. Maitra, Mri-based brain tumor image detection using cnn based deep learning method, *Neuroscience Informatics* (2022) 100060.
- [96] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86 (11) (1998) 2278–2324.
- [97] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Communications of the ACM* 60 (6) (2017) 84–90.
- [98] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, *International journal of computer vision* 115 (2015) 211–252.
- [99] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [100] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [101] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE conference on computer vision*

and pattern recognition, 2015, pp. 1–9.

- [102] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, U. R. Acharya, Classification of heart sound signals using a novel deep wavenet model, *Computer Methods and Programs in Biomedicine* 196 (2020) 105604.
- [103] L. Mang, F. Canadas-Quesada, J. Carabias-Orti, E. Combarro, J. Ranilla, Cochleogram-based adventitious sounds classification using convolutional neural networks, *Biomedical Signal Processing and Control* 82 (2023) 104555.
- [104] X. H. Kok, S. A. Imtiaz, E. Rodriguez-Villegas, A novel method for automatic identification of respiratory disease from acoustic recordings, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 2589–2592.
- [105] F. Demir, A. M. Ismael, A. Sengur, Classification of lung sounds with cnn model using parallel pooling structure, *IEEE Access* 8 (2020) 105376–105383.
- [106] G. Petmezas, G.-A. Cheimariotis, L. Stefanopoulos, B. Rocha, R. P. Paiva, A. K. Katsaggelos, N. Maglaveras, Automated lung sound classification using a hybrid cnn-lstm network and focal loss function, *Sensors* 22 (3) (2022) 1232.
- [107] S. K. Ghosh, R. K. Tripathy, R. Ponnalagu, R. B. Pachori, Automated detection of heart valve disorders from the pcg signal using time-frequency magnitude and phase features, *IEEE Sensors Letters* 3 (12) (2019) 1–4.
- [108] M. Milani, P. E. Abas, L. C. De Silva, N. D. Nanayakkara, Abnormal heart sound classification using phonocardiography signals, *Smart Health* 21 (2021) 100194.
- [109] R. Nersisyan, M. M. Noel, Heart sound and lung sound separation algorithms: a review, *Journal of medical engineering & technology* 41 (1) (2017) 13–21.
- [110] Z. Neili, K. Sundaraj, A comparative study of the spectrogram, scalogram, melspectrogram and gammatonegram time-frequency representations for the classification of lung sounds using the icbhi database based on cnns, *Biomedical Engineering/Biomedizinische Technik* 67 (5) (2022) 367–390.