www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 416-430. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2025.v16i4.1031

A New Robust Model for Heart Disease Detection on PCG signals using Entropygrams and a CNN

Monserrat Castro-Coria[1], Antonio Camarena-Ibarrola[1] and Karina Figueroa[2]
[1] División de Estudios de Posgrado de la Facultad de Ingeniería Eléctrica, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México
[2] Facultad de Ciencias Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México
monserrat.castro@umich.mx, antonio.camarena@umich.mx, karina.figueroa@umich.mx

Abstract. Every year, 17.9 million people die from heart-failure–related conditions, making it the leading cause of mortality worldwide; early diagnosis could prevent many deaths. The most common way to detect cardiac abnormalities is through medical auscultation. Accurate diagnosis often depends on clinicians' auscultation skills; however, they frequently have to listen in noisy environments. We propose a method for the automatic diagnosis of cardiovascular disease that is robust to noise. We extract entropy spectrograms from phonocardiograms to reliably convert audio signals into images, and then use a two-dimensional convolutional neural network (2D-CNN) to classify patients as healthy or unhealthy. To evaluate the method, we added white noise to the original recordings. The results show that entropy spectrograms are more robust than conventional feature-extraction techniques such as energy spectrograms or mel-frequency spectrograms.
Keywords: Phonocardiograms, Heart Sound, Entropygrams, Spectral Entropy, Convolutional Neural Networks.

Article Info
Received February 25, 2025
Accepted July 6, 2025

1 Introduction

Cardiovascular Diseases (CVD) are among the leading causes of mortality, according to the World Health Organization (WHO). Most CVDs are treatable when detected at early stages. The diagnosis normally relies on the physician's ability to hear as well as his/her training on distinguishing subtle differences in pitch and duration of heart sounds (HS). These sounds result from disturbances in blood flow and vibrations of cardiovascular structures, these vibrations are produced by events occurring in the heart valves and myocardium, which are influenced by the function, electrical activity, and hemodynamics of the cardiac muscle (Subasi, 2019). Auscultation or HS examination are the leading non-invasive, low-cost first-contact method to detect heart diseases.

The heart is an organ that pumps blood through all organs of the human body, it receives low-pressure blood from the veins, increases pressure by contracting its cardiac chambers, and expells blood into the arteries. Blood enters the heart through valves that open and close. These movements of contraction (systole) and relaxation (diastole) are audible; in the absence of any abnormal conditions, they are typically heard as *lub...dub..... lub...dub.....* (Chakrabarti et al., 2015). If any valves in the heart are damaged, additional sounds may be heard between the first heart sound (S1) and the second heart sound (S2), as well as between S2 and S1. These abnormal sounds are known as heart murmurs, which are audible vibrations caused by changes in blood flow through a damaged valve. There are also artifact sounds, which can occur during HS assessments. These noises result from factors such as coughing, hiccups, the movement of the stethoscope, breathing, or other body movements (Kumar & Saha, 2018). To determine the type of murmur a patient has, healthcare professionals consider several factors, including the timing, intensity, pattern, pitch, quality, and location of the sounds.

Phonocardiography is a non-invasive diagnostic technique that registers the sounds produced by the heart using a specialized microphone called a phonocardiograph. Phonocardiograms (PCG) are recordings that capture the acoustic signals produced by the heart's mechanical activity, especially the opening and closing of the heart valves. In a healthy adult, the systolic phase lasts approximately one-third of the total cardiac cycle time, as shown in Fig. 1, while the diastolic phase lasts the remaining two-

thirds. The duration of systole is considered constant. A heart murmur is classified according to its pattern (Chaudhuri & Jayanthi, 2016) as: crescendo-decrescendo (ascending and then descending sound, diamond-shaped), crescendo (intense sound that goes from weak to strong), decrescendo (sound that goes from strong to faint), and holosystolic (sharp sound throughout systole).

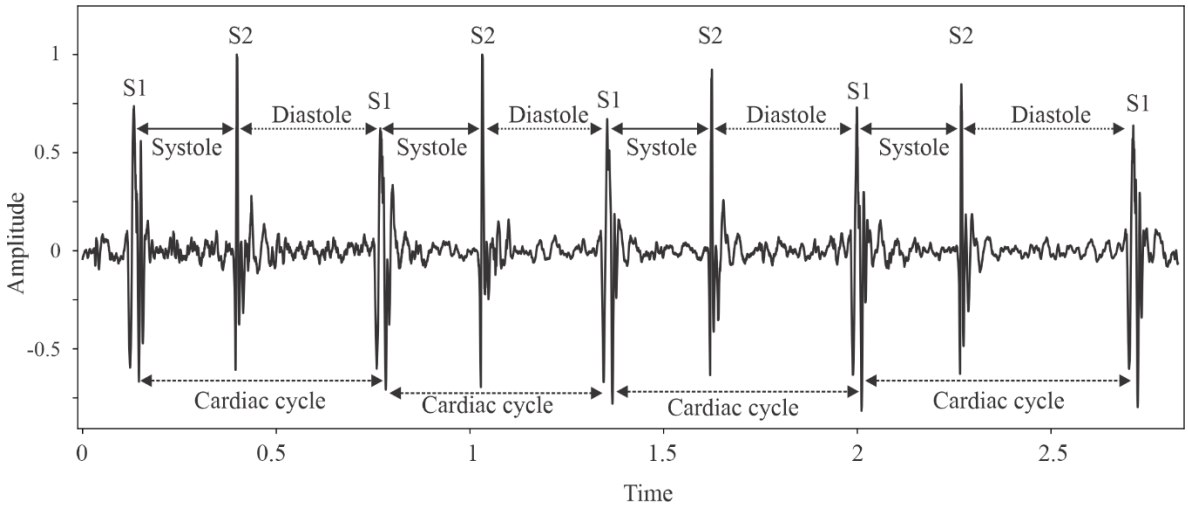

Fig. 1 PCG signal from dataset PASCAL (normal__128_1306344005749_B.wav) with annotations of the parts that compose each cardiac cycle.

Murmurs are also classified according to their intensity, following the Levine scale: I, soft, only identified by highly experienced cardiologists; II, audible but very weak; III, easily audible; IV, strong, accompanied by a sound produced by strong blood flow turbulence; V, strong enough to be heard with a stethoscope barely touching the auscultation site; and VI, strong enough to be heard with a stethoscope not in contact with the auscultation area.

The heart's valves are aortic, mitral, pulmonary, and tricuspid, each of which can cause problems such as regurgitation or stenosis. The last one, stenosis, emerges when the valve leaflets thicken and become rigid, narrowing the valve opening and causing less blood flow. Conversely, regurgitation occurs when valve leaflets don't close properly, causing blood to flow backward. Detecting these types of murmurs is crucial, depending on where and when sounds arise.

Heart murmurs, as shown in Fig. 2, are classified according to their type (Liu et al., 2016) as:

a) Mitral regurgitation that occurs during the systolic cycle with a uniform pattern (holosystolic) when the mitral valve doesn't close properly, a sound between 60 and 600 Hz is produced.
b) Tricuspid regurgitation occurs during systolic cycle with a uniform pattern (holosystolic) when the tricuspid valve doesn't close properly, in that case a sound between 100 and 300 Hz is produced.
c) Aortic stenosis occurs during systolic cycle and presents a crescendo-decrescendo pattern when the aortic valve doesn't open properly, then a sound between 300 and 500 Hz is produced.
d) Pulmonary stenosis occurs during the systolic cycle with a crescendo-decrescendo pattern when the pulmonary valve doesn't open properly, and a sound between 100 and 400 Hz is produced.
e) Mitral stenosis occurs during the diastolic cycle with an opening snap followed by a decrescendo-crescendo sound when the mitral valve doesn't open properly, in that event, a sound between 50 and 150 Hz is produced.
f) Tricuspid stenosis occurs during the diastolic cycle with a crescendo pattern when the tricuspid valve doesn't open properly and a sound between 50 and 150 Hz is produced.
g) Aortic regurgitation occurs during the diastolic cycle with a decrescendo pattern when the aortic valve doesn't close properly, a sound between 200 and 300 Hz is produced.
h) Pulmonary regurgitation occurs during the diastolic cycle with a decrescendo pattern when the pulmonary valve doesn't close properly, a sound between 200 and 400 Hz is produced.

S3 and S4 sounds occur during diastole and are low-frequency sounds, also known as *extracardiac sounds*. They are normally heard in children, pregnant women, and athletes without posing a health risk.

Machine learning techniques have been used recently to identify heart diseases, but there is a lack of research on the classification of heart problems when background noise is present, for example, the noise produced by air coolers, fans, traffic noise, rain, etc.

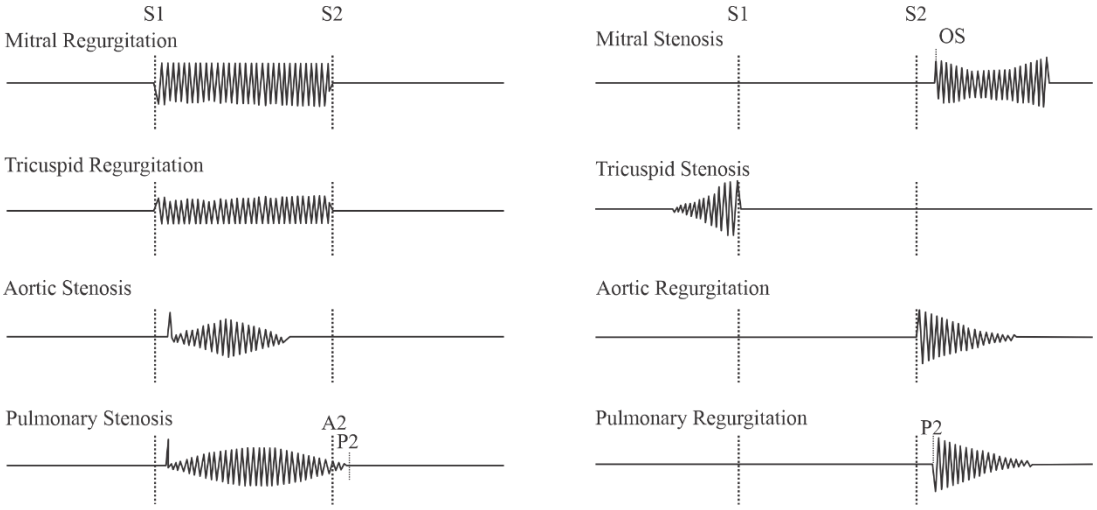

Fig. 2 Types of murmurs.

1.1 Related studies

During the Hippocratic practice (Hanna & Silverman, 2002), 400 years B.C., to auscultate a person, the ear was placed directly on the patient's chest while the body was shaken to listen to the internal movement of the patient's fluids; this was called *immediate auscultation*. It was during the 15th century when Leonardo Da Vinci (Wells, 2014), intrigued by the functioning of the body, inserted fourteen wires into a live pig that reached the heart, to describe venous and arterial circulation, seeking to understand the movement produced by the heart. He noted that the movements were so rapid that it was impossible to quantify them in time. Robert Hooke predicted the usefulness of auscultation …*I have been able to hear very plainly the beating of a Man's heart... Who knows, I say, but that it may be possible to discover the Motions of the Internal Parts of Bodies... by the sound they make, that one may discover the Works performed in the several Offices and Shops of a Man's Body, and thereby discover what Instrument or Engine is out of order…* (Hooke, 1705, p. 39), and to this day, more than five hundred years later, it remains the most widely used method by medical doctors to examine a patient.

It was not until 1816 when Laennec (Geddes, 2005), while examining a young woman's chest and feeling uncomfortable putting his ear on her chest, it occurred to him, knowing that it was possible to hear sounds through solid objects, to roll up a piece of paper forming a cylinder. He placed one ear against the paper and the other on the young woman's chest; he was not surprised to find that the sound produced by the heart could be heard even better. Later, he tested various materials, finding that wood propagated sound better, and called it a *stethoscope*.

Laennec (Laennec, 1819), Forbes (Forbes et al., 1824), G. Camman, and J. Hope (Hanna & Silverman, 2002), among others, contributed to the analysis and classification of cardiovascular diseases through the produced sound. Einthoven (Einthoven & Geluk, 1894), father of the electrocardiogram (ECG), relied on the sounds emitted by the heart to create an instrument that would allow visualization of electrical records of the heart. Thomas Lewis (Lewis, 1920), based on Einthoven's studies, used a method that allowed recording heart sounds, using a microphone and the stethoscope created by Laennec, connected through tubes to transmit sound (vibrations), and these vibrations were recorded at frequencies of 200 to 300 per minute. Fahr (Sprague, 1962), a disciple of Einthoven, working in his laboratory, analyzed the relationship between ECG and recorded sounds; his observations and subsequent publications significantly contributed to cardiology.

In the year 1949, Aubrey Leatham (Leatham, 1949) emphasized the need to analyze sounds produced by the heart through phonocardiography, mainly affected by the limitations of the human ear in perceiving certain frequencies. Additionally, in PCG sounds such as a third or fourth *beat* (S3 or S4) can easily be identified, which the ear might not hear as they occur in small time intervals, and which are not observable in other methods such as ECG.

McKusick (McKusick et al., 1955) began studying heart problems by using recordings that generated visual representations of HS through PCG. This method enabled detailed observation of the timing, frequency, and intensity of these sounds, thereby facilitating

the detection of various cardiovascular diseases. This was important because the human ear detects sounds at intervals of 0.05-0.06 seconds, while some cardiac alterations occur in 0.03 seconds, making medical training crucial for auscultation.

One advantage of recording HS is that the recordings can be reviewed later to identify various problems (McKusick et al., 1956). However, to classify new recordings, it is necessary to preprocess the signal by removing noise or artifacts to ensure correct classification.

Groch (Groch et al., 1992) proposed using HS as a physiological parameter for cardiac synchronization as an alternative to ECG. This method involves using PCG along with an equation that estimates the R and T frequencies of an ECG. It is assumed that the interval between the first heart sound (S1) and the second heart sound (S2) is shorter than the interval between S2 and S1. The analysis is conducted with window intervals of 200 to 400 milliseconds, which offers a significant advantage in detecting additional heart sounds such as S3 and S4.

Liang (Liang et al., 1997) determined that the ideal approach would be to record these sounds and analyze them in a computerized and objective manner, first segmenting the signals into components for individual analysis.

El-Hanjouri (El-Hanjouri et al., 2002) analyzed PCG using Hidden Markov Models (HMM) to classify different components. Since recordings often contain noise, they filtered the signal to eliminate non-cardiac sounds. They determined that the frequency components S1 and S2 of normal sounds are between 50-140 Hz and 80-200 Hz respectively. Any deformation or additional sound is called a murmur, caused by turbulent blood flow associated with improperly functioning valves. Murmurs have higher frequencies (up to 600 Hz).

Kumar (Kumar et al., 2006) points out that several cardiovascular problems can be diagnosed efficiently through auscultation. The objective was identifying the limits of each of the segments produced by the HS, using wavelets of the Daubechies family of order 6, since he considered that their shape adapts to the natural recordings of HS. This method effectively identifies the S1 and S2 moments of the cardiac cycle in a healthy individual; however, several inaccuracies still arise when diagnosing CVD.

Schmidt (Schmidt et al., 2008) mentions the importance of identifying the moments S1, Systole, S2, and Diastole, considering that these four states that occur in a chain HMM are usually very useful for identifying the segments of the cardiac cycle. HMM was able to detect 5 types of murmurs (Zhong et al., 2013), crucial information for diagnosing specific diseases.

Chakrabarti (Chakrabarti et al., 2015) analyzed the challenges that arise in the analysis of PCG, pointing out that the segmentation of an audio signal from the sounds produced by the heart was the most difficult challenge due to the anomalies that may occur.

PCG signal analysis consists of six steps: start, data acquisition, segmentation, feature extraction, classification, and end (Subasi, 2019). One of the biggest challenges has been preprocessing the audio signal to eliminate noise that accompanies the recordings, typically using a band-pass filter. As shown in Fig. 3, the mechanical movements produced by the heart are recorded in an audio file from which a spectrogram can be extracted, where its most relevant frequencies can be depicted. Several variations such as Mel-spectrograms, Scalograms, Energy spectrograms, and bispectrum magnitude have been used. However, noisy audio signals continue to be one of the greatest challenges for good results in the next stage (classification).

Springer (Springer et al., 2015) identifies that segmenting an audio signal without the presence of noise is relatively simple, however recording an audio signal of the sounds emitted by the heart without the presence of noise is impossible. For this reason, it uses different combined methods (Wavelets, Hilbert Transform, Spectral Density, FFT) using both PCG and ECG samples.

Chaudhuri and Jayanthi (Chaudhuri & Jayanthi, 2016) took two signals to identify the location of an abnormality in the patient's chest, the first sound was taken from the atrial area and the second one from the ventricular area, aligning both signals to discover the S1 and S2 moments. They identified three kinds of abnormal sounds, the first from S1 to S2 in its entire range (holosystolic), the second from S1 to S2 in a diamond shape, and the third from S2 to S1 in a decreasing shape, deducing that the place where one of these shapes is most visible can be diagnosed with some type of cardiovascular disease, using a neural network to carry out the classification.

In 2016, a group of researchers created a public dataset (Liu et al., 2016), as a byproduct of their research where they determined that the sounds emitted by the heart are between the frequencies 10-140 Hz for the S1 and 10-200 Hz for S2, abnormal sounds are found up to frequencies of 600 Hz, while sounds emitted by breathing occur in the range of 200 to 700 Hz, this complicates further the segmentation of the sounds emitted by the heart. They evaluated the use of several techniques for extracting features and

classification methods such as Wavelets, Shannon Entropy, Machine Learning methods, HMM, Neuronal Networks, Support Vector Machine, and Clustering, among others, being a dataset that allows further research on HS.

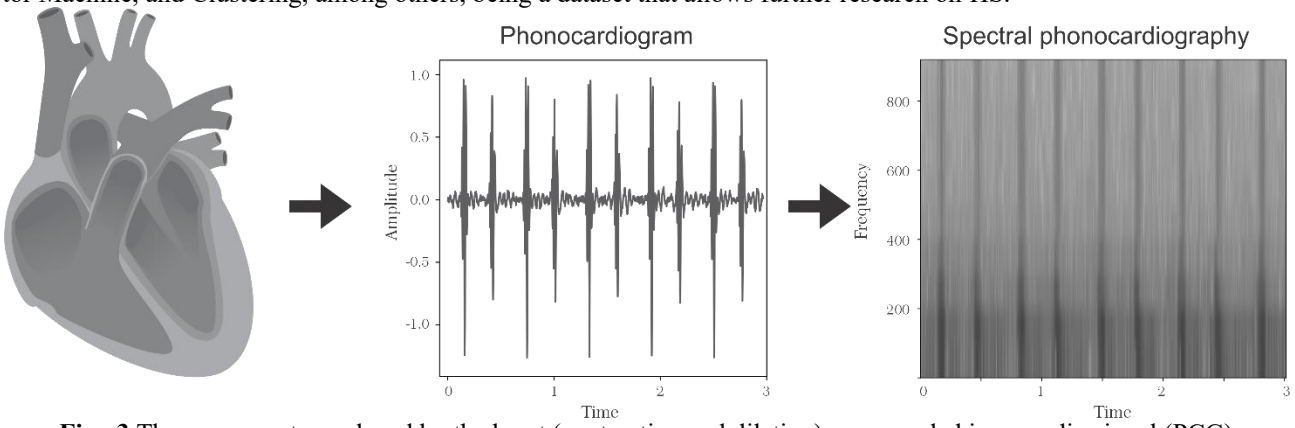

Fig. 3 The movements produced by the heart (contraction and dilation) are recorded in an audio signal (PCG) to produce an image of energy values in the frequency domain.

Table 1. Comparison in terms of accuracy among works on murmur diagnosis from heart sound signals

Reference & Year	Noise reduction	Features	Classification	Accuracy
Meintjes et al. (2018)	Low-pass filter 25 Hz	CWT Scalograms	CNN	86.00%
Low & Choo (2018)	Normalization	Spectrograms of energy	CNN	80.30%
Han et al. (2018)	-	MFCC	CNN	91.50%
Wibawa et al. (2018)	-	Spectrograms of energy	CNN	82.75%
Ren et al., 2018	-	Scalograms	CNN	56.20 %
Normal et al., 2019	Band-pass filter 25-400 Hz	MFCC	CNN	88.82 %
Banerjee & Majhi (2020)	Band-pass filter 30-900 Hz	MFCC	CNN	83.00 %
Chen et al. (2020)	-	Modified WT	CNN	93.91%
Alqudah et al. (2020)	Low-pass filter 500 Hz	Spectrograms of energy	CNN	93.70%
Takezaki & Kishida (2021)	Normalization	Bispectrum Magnitude	CNN	98.70%
Cheng & Sun (2023)	Band-pass filter 25-400 Hz	-	CNN Transformer	95.70%
Torre-Cruz et al. (2023)	Bior-4.4 discrete filter	Mean, variance, kurtosis, energy, power, etc.	MSV, RT, DT, KNN	98.00%
Proposed	None	Entropygrams	2D-CNN	100.00%

Thanks to the accessibility of HS databases, several researchers were able to test various, increasingly effective methods for classifying the sounds produced by the heart. Whether using Energy Spectrograms (Low & Choo, 2018; Takezaki & Kishida, 2021; Wibawa et al., 2018), MFCC (Banerjee & Mahji, 2020; Norman et al., 2019), Wavelet Transform Scalograms (Chen et al., 2020; Ren et al., 2024), Bispectral Magnitude (Alqudah et al., 2020) or using various types of NN, the most commonly used being CNN (Alqudah et al., 2019; Norman et al., 2019; Renna et al. 2019; Banerjee & Majni, 2020; Chen et al., 2021; Takezaki & Kishida, 2021; Yildrim, 2022; Bao et al., 2023) and using Transformers (Cheng et al., 2023) to classify the type of disease. Denoising the audio signal has been one of the biggest challenges.

Panah (Panah et al., 2023) conducted a study on the effects of noise on the classification of HS. They observed that capturing a signal free of noise is nearly impossible. They created a synthetic dataset contaminated with different noise levels (-10 dB to 40 dB). They concluded that noises such as coughs, sneezes, or barks are easier to eliminate than long-duration noise that interferes with the frequencies of the recorded signal, making correct classification difficult.

If the noise as a range frequency that is similar with the range of frequencies emitted by HS, it is very difficult to eliminate, and the common practice is to discard these samples, losing valuable information. We instead decided to use entropygrams, a potentially more effective method for analyzing HS in noisy environments. This method is expected to be more robust when dealing with noisy audio signals, capturing the important characteristics of HS even with interference.

For murmur diagnosis from heart sound signals, most researchers use filters to eliminate noise. Since we were interested in assessing spectrograms of entropy (i.e. entropygrams) as features extracted from noisy signals with the purpose of classification, we did not use any filtering, furthermore, we added more noise to the signal since the purpose of our test was to assess the robustness of our proposal. Table 1 shows a comparison in terms of accuracy among works on murmur diagnosis from heart sound signals.

2 Robust Heart Disease Detection

Our method consists in turning cardiac audio signals into images that preserve essential diagnostic information while being resilient to noise. These images are then fed to a 2D convolutional neural network since this kind of networks are known for their effectiveness in image recognition tasks. The network predicts the specific cardiac condition.

The preprocessing pipeline begins with amplitude normalization, constraining values between -1 and 1 to ensure consistency across recordings. The normalized signal is split into 30 milliseconds frames with a two-thirds overlap between consecutive frames. To each frame, a Hamming window is applied to mitigate spectral leakage effects.

2.1 Conversion of audio-signals into images

We use spectrograms of entropy to illustrate how the information content in the signals is distributed across frequencies and evolves over time. This approach has been used before for robust audio analysis in noisy environments (Camarena-Ibarrola et al., 2020). Our implementation uses the psychoacoustic Bark scale (Eq. 1), focusing specifically on the first eight critical bands that encompass the frequency range of primary HS (S1 and S2: 50-400 Hz) and murmurs (up to 600 Hz) as identified by El-Hanjouri et al. (2002).

$$z = 13 \arctan\left(\frac{0.76f}{1000}\right) + 3.5 \arctan\left(\frac{f}{7500}\right)^2 \tag{1}$$

Spectral entropy measures the amount of information in the signal's spectrum, this metric effectively differentiates between sound types, identifies audio transitions, and achieves great noise resilience. In cardiac contexts, spectral entropy reveals subtle patterns indicative of abnormal conditions that may remain undetectable in time-domain analysis. For stationary random processes with Gaussian density N(0,∑), entropy can be determined using the variances of real and imaginary parts of the Fourier coefficients with Eq. 2. We compute entropy for each frame and critical band, arranging them in an array where columns correspond to time progression and rows represent the number of the critical band.

$$H = \ln(2\,\pi e) + \frac{1}{2}\ln(\sigma_x^2\sigma_y^2 - \sigma_{xy}) \tag{2}$$

where σ_x^2 and σ_y^2 represent the variances of the real and imaginary parts of the Fourier coefficients respectively, and σ_{xy} denotes the covariance between these two components of the spectrum.

The procedure is illustrated in Fig. 4. First, the audio signal generated by HS is captured and normalized to values between -1 and 1. Next, the signal is divided into short frames. The Discrete Fourier Transform (DFT) is then applied to each frame, and the Fourier coefficients corresponding to each of the first eight critical bands according to the Bark scale are grouped. For each critical band, its spectral entropy is determined, resulting in a matrix that, when depicted as an image, is referred to as a spectrogram of entropy.

The computational process involves computing entropy values for each frame and critical band. These values populate a two-dimensional matrix where the horizontal axis is time and the vertical axis is the frequency in the Bark scale. Fig. 5 shows spectrograms of entropy for four different classes: cardiac sounds containing artifacts, those exhibiting extrasystole, recordings with murmurs, and those from subjects with normal (healthy) cardiac function.

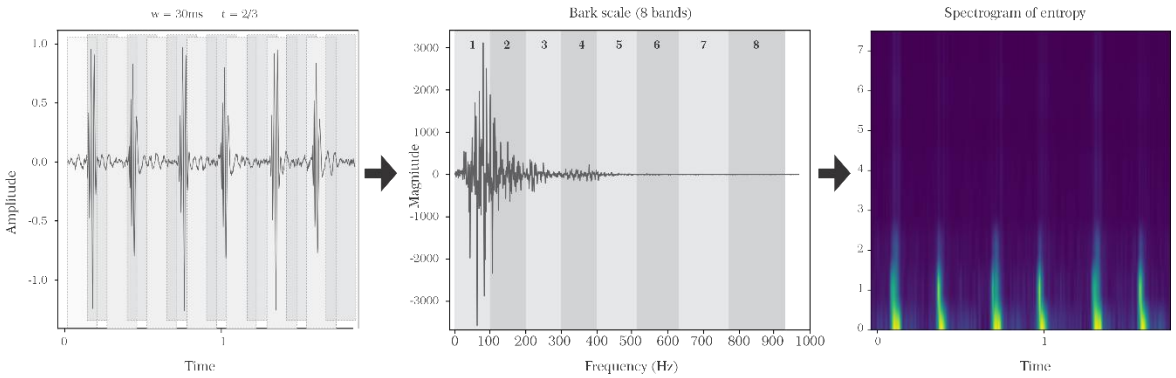

Fig. 4 The audio signal is split into 30-millisecond frames. The DFT is applied to each frame, and the Fourier coefficients are grouped into the first 8 critical bands according to the bark scale. For each band, the spectral entropy is calculated resulting in a spectrogram of entropy.

The visual matrix is organized such that the leftmost column displays entropygrams derived from unmodified source recordings. The central column presents entropygrams generated from signals that have been degraded with white noise, achieving a Signal-to-Noise Ratio (SNR) of 5 dB, and the rightmost column displays entropygrams from signals that are severely contaminated with noise, with SNR of -5 dB. At this extreme noise level, the acoustic interference is so substantial that the human auditory system cannot reliably detect individual cardiac cycles within the recording. SNR quantification follows the formulation presented in Eq. 3.

$$SNR\ (dB) = 10Log\left(\frac{SignalEnergy}{NoiseEnergy}\right) \tag{3}$$

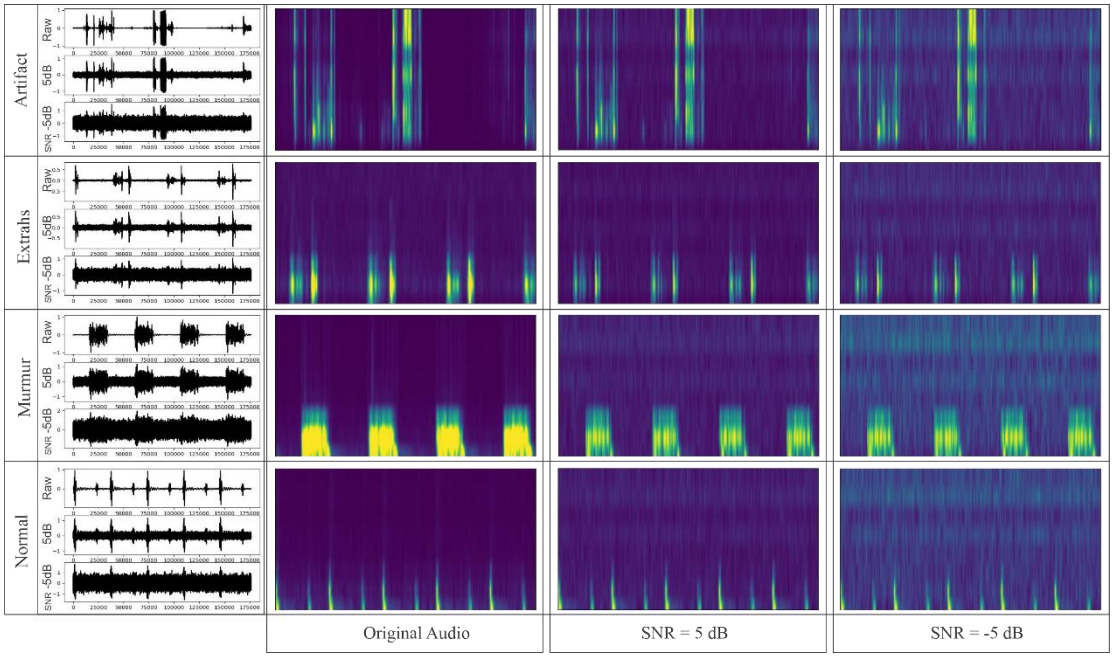

Fig. 5 Spectrograms of entropy from original audio signal and for signals mixed with SNR of 5dB and -5dB for 4 different classes of audio signals (Normal or Healthy; Murmur; ExtraSystole; Artifacts).

For our research, the class imbalance issue that has challenged previous heart sound classification studies (Chen et al., 2020; Han et al., 2018; Noman et al., 2019; Ren et al., 2018; Takezaki & Kishida, 2021; Wibawa et al., 2018). Rather than employing conventional image augmentation techniques (rotation, scaling, shifting), which would not make sense with spectrograms, we implemented dataset interpolation to generate further training data. This method creates new RGB images by interpolating (Eq. 4) between two existing entropygrams of the same class, using a parameter *t* that controls similarity to the source images, as shown in Fig. 6. Once we interpolated the images, only non-interpolated images were used for training.

$$P_{new}\begin{cases} R_{new} = (1-t)\times R_1 + t\times R_2 & \text{for } t\in[0,1] \\ G_{new} = (1-t)\times G_1 + t\times G_2 & \text{for } t\in[0,1] \\ B_{new} = (1-t)\times B_1 + t\times B_2 & \text{for } t\in[0,1] \end{cases} \quad (4)$$

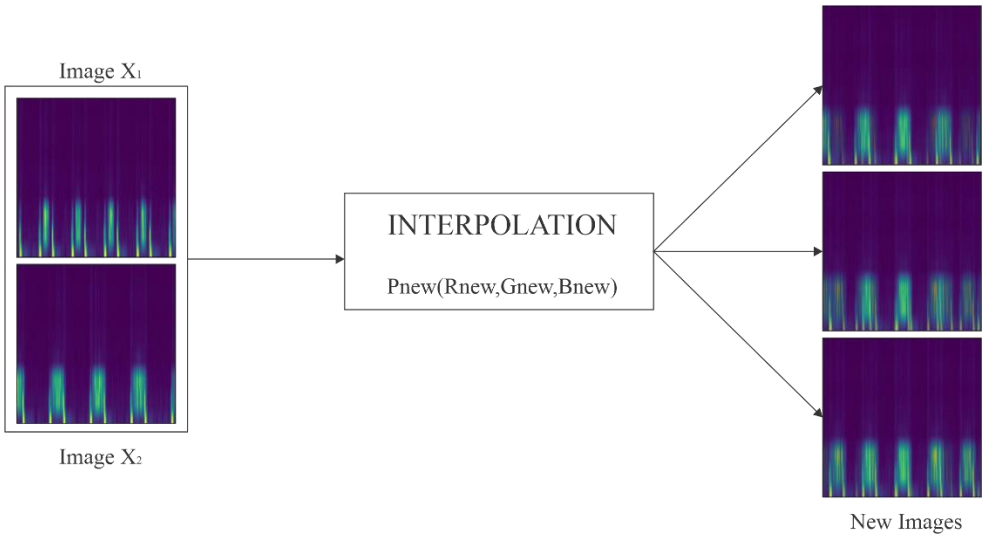

Fig. 6 To balance the dataset, new images are generated through interpolation.

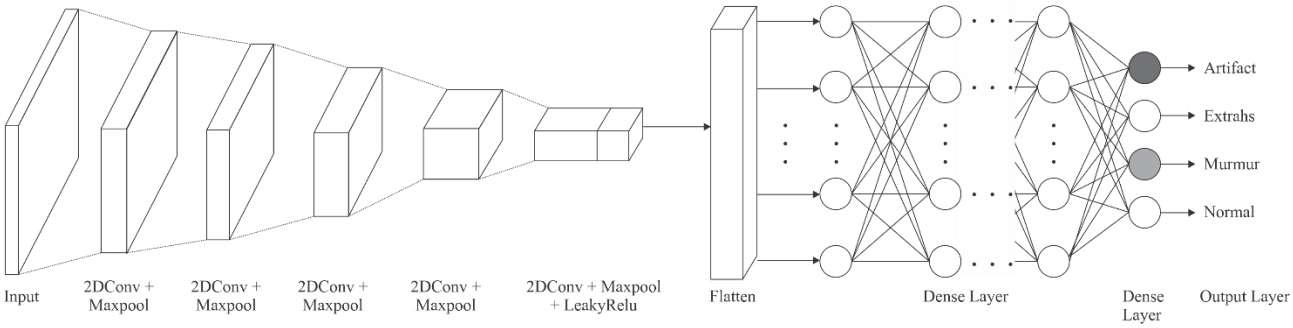

Fig. 7 Convolutional Neural Network used for the diagnosis in the Bentley's dataset.

Table 2. Specifics for the CNN used with 4 and 5 classes

Layer	Bentley's dataset		Yassen's dataset	
	Shape	Parameters	Shape	Parameters
Conv2D	252, 252, 32	2,432	252, 252, 32	2,432
MaxPooling2D	126, 126, 32	0	126, 126, 32	0
Conv2D	122, 122, 64	51,264	122, 122, 64	51,264
MaxPooling2D	61, 61, 64	0	61, 61, 64	0
Conv2D	57, 57, 128	204,928	57, 57, 128	204,928
MaxPooling2D	28, 28, 128	0	28, 28, 128	0
Conv2D	24, 24, 256	819,456	24, 24, 256	819,456
MaxPooling2D	12, 12, 256	0	12, 12, 256	0
Conv2D	8, 8, 512	3,277,312	8, 8, 512	3,277,312
MaxPooling2D	4, 4, 512	0	4, 4, 512	0
LeakyReLu	4, 4, 512	0	4, 4, 512	0
Dropout	4, 4, 512	0	4, 4, 512	0
Flatten	8192	0	8192	0
Dense	256	2,097,408	256	2,097,408
Dense	4	1,028	5	1,285
Total		6,453,828		6,454,085

The classification architecture includes a convolutional neural network with five convolutional layers, each followed by a MaxPooling layer. The classification component comprises two dense layers, the output layer has four neurons for Bentley's dataset corresponding to our target classes: Artifacts, ExtraSystole, Murmur, and Normal (Fig. 7). For tests with Yassen's dataset, the output layer has 5 neurons, each one corresponds to the class: N (Normal); AS (Aortic Stenosis); MS (Mitral Stenosis); MR (Mitral Regurgitation); and MVP (Murmur in systole), as shown in Fig. 8. We incorporated dropout techniques to prevent overfitting during training. Table 2 shows the specifics of CNN. Entropygram images are standardized to 256 × 256 pixels, and the window's size used is 5 × 5 for every convolutional layer.

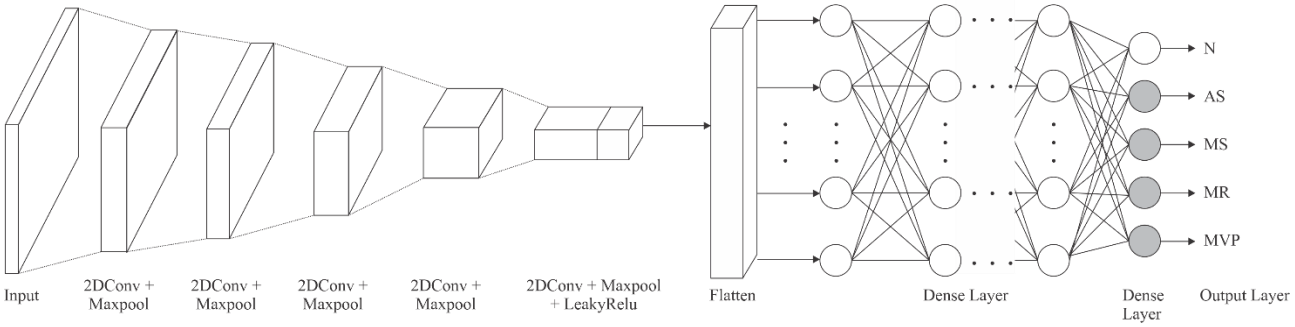

Fig. 8 Convolutional Neural Network used for the diagnosis in the Yanseen's dataset.

3 Experimental procedures

Most researchers have tested their methods on the same datasets since there are few accessible databases, and some of them contain only a few of the different types of murmurs that can be found in a patient. Therefore, two databases with more than two classes were selected to better distribute the classification results. Table 3 shows some open-access databases developed by university researchers with the support of medical specialists.

Table 3. Published and accessible datasets for heart sound classification (audio files)

Dataset	Reference	#Samples	Task
Pascal database A	Bentley et al. (2011)	176	Normal, Murmur, ExtraHS, Artifact
Pascal database B	Bentley et al. (2011)	656	Normal, Murmur, Extrasystole
PhysioNet/CinC database	Liu et al. (2016)	3,240	Normal, Abnormal
GitHub database	Yaseen et al. (2018)	1,000	Normal, AS, MR, MS, MVP
Circor DigiScope database	Oliveira et al. (2021)	3,163	Normal, Abnormal

In our experiments, we used the dataset from Bentley's Classifying Heart Sounds Challenge (Bentley et al., 2011), which comprises four distinct categories of cardiac acoustic signals: those with artifacts, extra-systole, normal (healthy) cardiac function, and murmurs. And Yaseen's dataset (Yaseen et al., 2018) which includes the classes: normal, aortic stenosis (murmur between S1 and S2), mitral stenosis (murmur between S2 and S1), mitral regurgitation (murmur between S1 and S2), and murmur in systole (murmur between S1 and S2), as shown in Table 4.

Table 4. Datasets used

Dataset	Classes	Samples
Pascal - A	Normal	45
	Murmur	48
	Extrahs	27
	Artifact	56
GitHub Open Access	Normal (N)	200
	Aortic Stenosis (AS)	200
	Mitral Stenosis (MS)	200
	Mitral Regurgitation (MR)	200
	Murmur in systole (MVP)	200

To assess the robustness of our classification methodology under adverse acoustic conditions, we systematically introduced white noise to the original cardiovascular audio recordings. Noise mixing was carefully calibrated to produce Signal-to-Noise Ratios

(SNR) spanning from -10 dB to 20 dB. Specifically, we generated test conditions at SNR values of -10 dB, -7 dB, -5 dB, 0.1 dB, 0.5 dB, 1 dB, 3 dB, 10 dB, 15 dB, and 20 dB for Bentley's dataset as Fig. 9, and values of -15 dB, -10 dB, -7 dB, -5 dB, -3 dB, -1 dB, 1 dB, 3 dB, 5 dB, 7 dB, 10 dB, 15 dB, and 20 dB for Yaseen's dataset. It should be noted that lower SNR values, particularly negative ones, indicate proportionally higher noise contamination. Each audio collection was subsequently transformed into corresponding image representations.

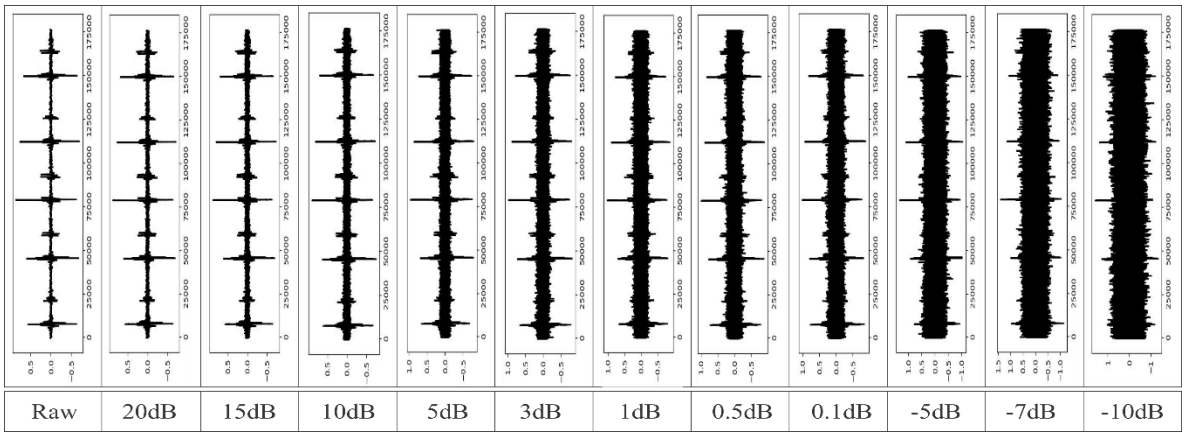

Fig. 9 Different levels of white noise are added to the audio signal, which makes it impossible to identify the segments of the audio signal in the time domain.

For the audio-to-image conversion process, we use spectrograms of entropy, but alternative methodologies are used in the field, so for comparative purposes we also produced the Spectrograms of energy; the Wavelet Scalograms; and the Mel Frequency Cepstral Coefficient (MFCC) Spectrograms, for both Bentley's and Yaseen's datasets. Fig. 10 shows representative examples of these four images applied to each of the cardiac sound categories in Bentley's dataset. Similarly Fig. 11, show the corresponding images for Yassen's dataset.

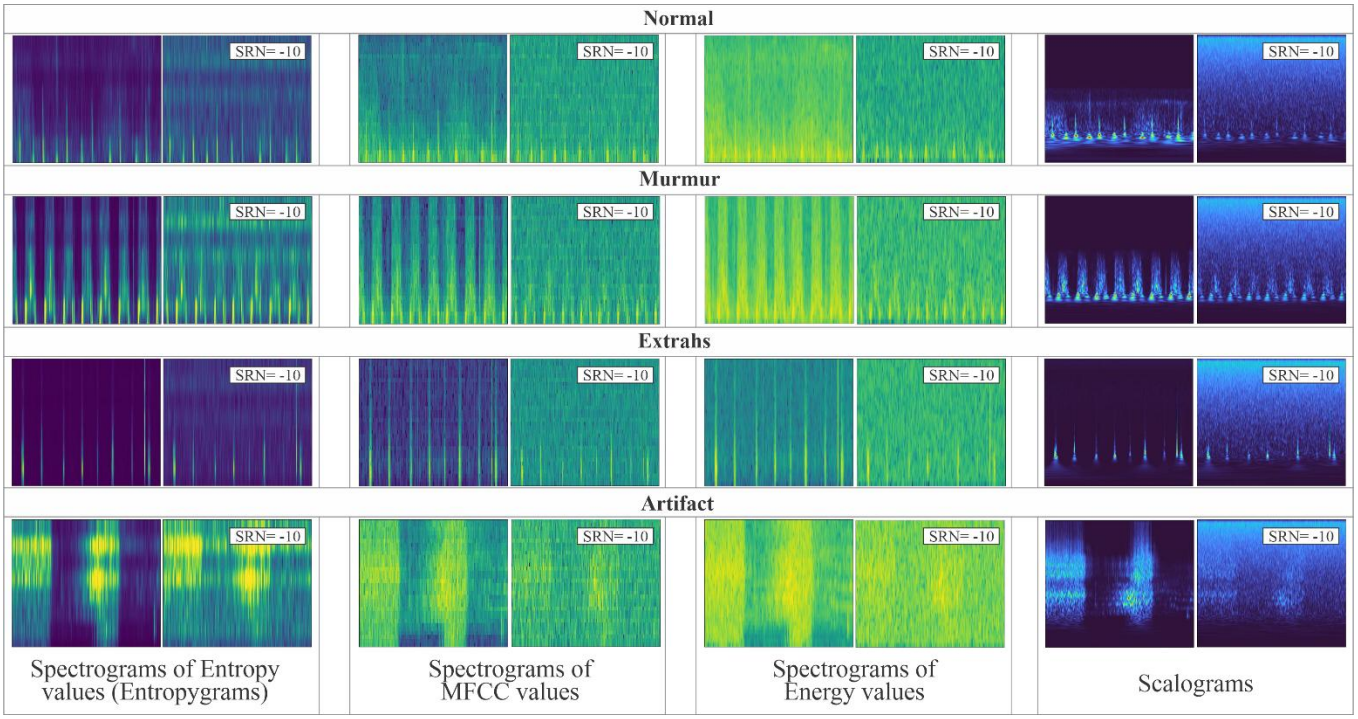

Fig. 10 Samples of the four classes of the Bentley's dataset (Artifact, Extra Heart Sound, Murmur and Normal) using the four different feature extraction methods.

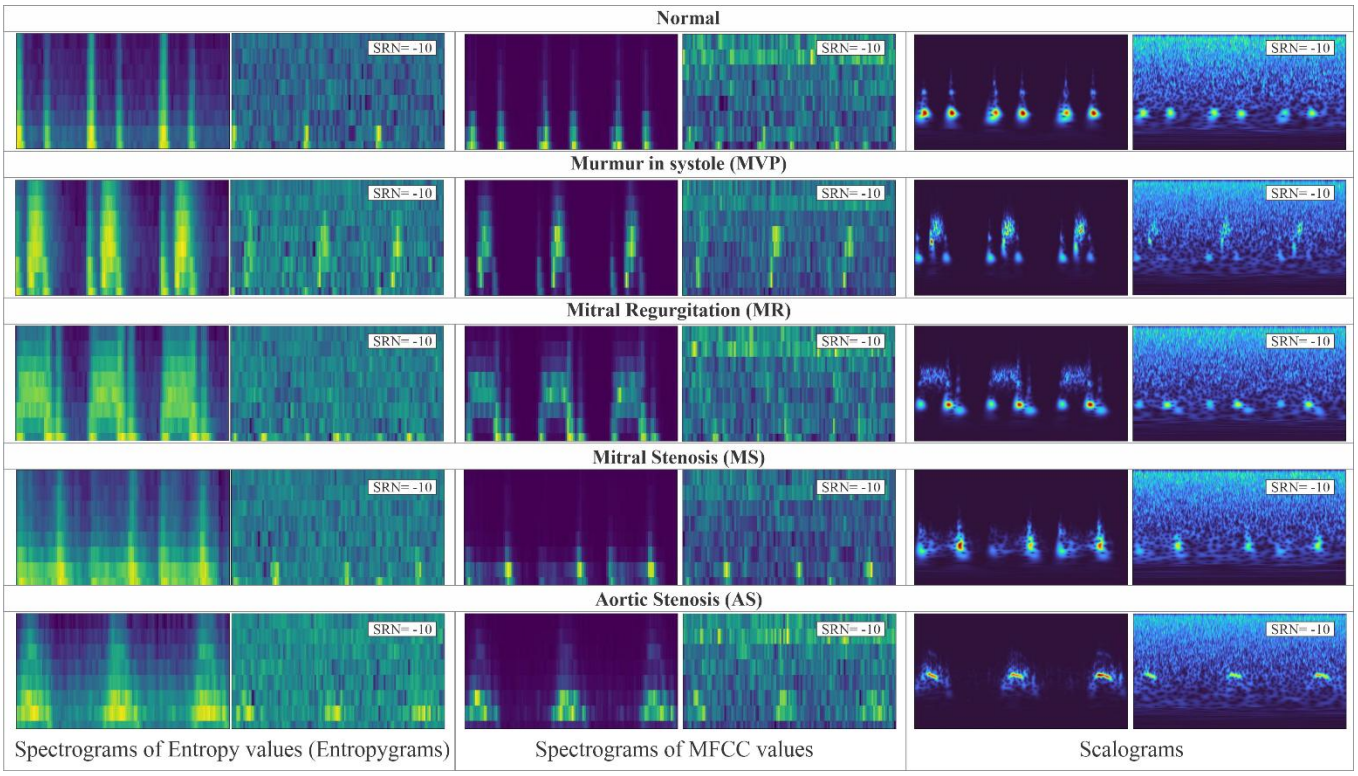

Fig. 11 Samples of the five classes of the Yaseen's dataset (Normal, Murmur in systole, Mitral Regurgitation, Mitral Stenosis and Aortic Stenosis) using three different feature extraction methods.

4 Results

Our experimental protocol used 70% of each image collection training and 30% for tests. Table 5 shows the accuracies obtained for Bentley´s dataset, which is also depicted graphically in Fig. 12.

Table 6 shows the accuracies obtained for Yaseen's dataset, which is also depicted graphically in Fig. 13. The results confirm the general observation, previously noted by Panah, Hines & McKeever (2023), that acoustic noise adversely affects the classification performance across all HS analysis techniques. Notably, our proposed entropy spectrogram approach demonstrated superior noise resilience, maintaining exceptional classification accuracy of 99.43% even under the most challenging noise conditions (SNR of -10 dB).

Table 5. Accuracy achieved for the heart audio-signal image representations (columns) and noise levels (rows); the first rows correspond to higher noise levels for Bentley's dataset

SNR	Spectrogram of Energy	Scalogram	Spectrogram of MFCC	Spectrogram of Entropy (Entropygram)
-10 dB	0.5833	0.2777	0.9166	**0.9444**
-7 dB	0.9166	0.2777	0.8888	**0.9722**
-5 dB	0.8611	0.8611	0.8611	**0.9722**
0.1 dB	0.8888	0.9166	0.9444	**1.0000**
0.5 dB	0.9444	0.8888	0.8888	**1.0000**
1 dB	0.8888	0.9722	0.8611	**1.0000**
3 dB	0.8888	0.9444	0.8888	**1.0000**
5 dB	0.9444	0.9166	0.8888	**0.9722**
10 dB	0.9166	**0.9722**	0.8888	**0.9722**
15 dB	0.9166	**0.9722**	**0.9722**	**0.9722**
20 dB	0.9166	0.9444	0.9166	**1.0000**
Raw	0.9166	0.9444	0.9166	**1.0000**

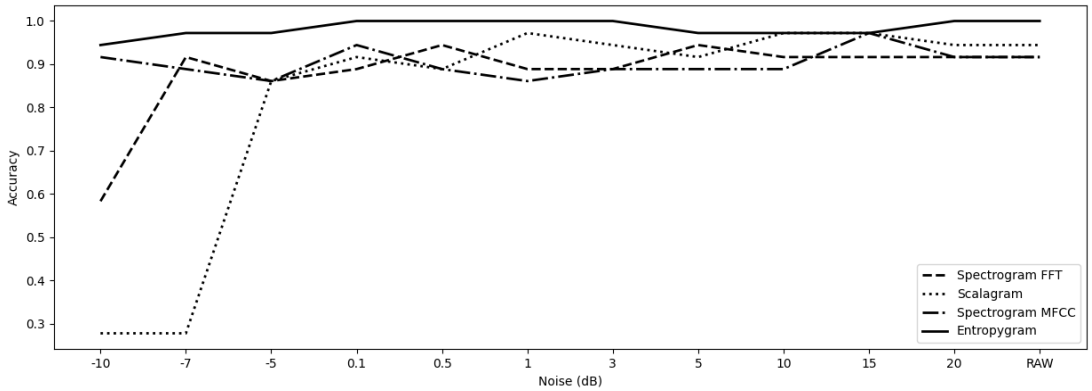

Fig. 12 Accuracy vs SNR for several image representations of features from heart audio-signals for Bentley's dataset with four classes. Increment in horizontal axis means less noise.

For audio signals with such high noise content that the human auditory system would struggle to even discern the heartbeats, our method only reduces the 100.00% accuracy, achieved with uncontaminated signals, to 88.99% for Bentley's dataset which is still hugh enough for practical purposes. In the case of Yaseen's dataset, the accuracy obtained with the noisiest audio signal was 83.45%, remarking the exceptional robustness of our methodology.

Table 6. Accuracy obtained for the image representations of the heart audio-signal (columns) and levels of noise (rows), the first rows correspond to higher levels of noise for Yaseen's dataset

SNR	Scalogram	Spectrogram of MFCC	Spectrogram of Entropy (Entropygram)
-15 dB	0.6833	0.6833	**0.7133**
-10 dB	0.7133	0.7433	**0.7966**
-7 dB	0.7433	0.7599	**0.8033**
-5 dB	0.7533	0.7766	**0.8266**
-3 dB	0.7533	0.7766	**0.8466**
-1 dB	0.7766	0.8133	**0.8366**
1 dB	0.8366	0.8133	**0.8433**
3 dB	0.8366	0.7966	**0.8566**
5 dB	0.8633	0.7966	**0.8700**
7 dB	0.8433	0.8366	**0.8566**
10 dB	0.8433	0.8033	**0.8566**
15 dB	0.8433	0.8033	**0.8733**
20 dB	0.8366	0.7666	**0.8700**
Raw	0.8500	0.8633	**0.8899**

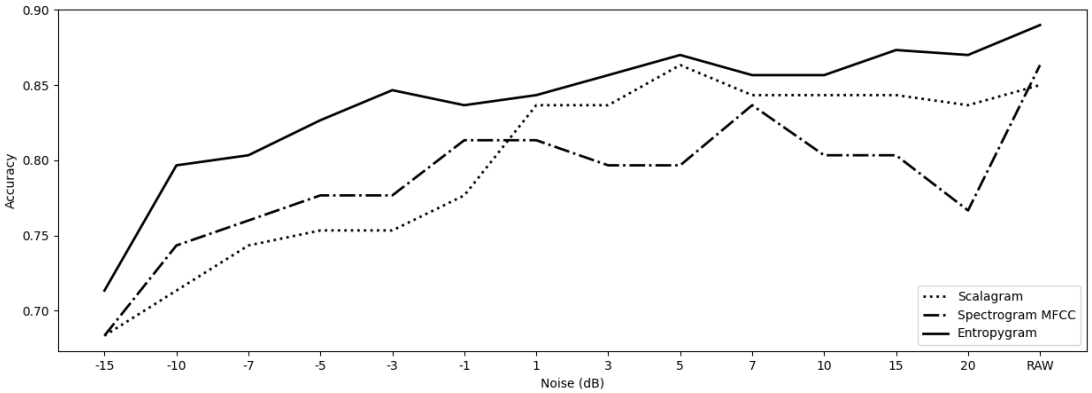

Fig. 13 Accuracy vs SNR for several image representations of features from heart audio-signals for Yanseen's dataset with five classes. Increment in horizontal axis means less noise.

Regarding accuracy obtained using uncontaminated (i.e. original) heart sound signals, we found that the four classes from Bentley's dataset are easily separated. In particular, healthy (i.e. normal) hearts produce sounds where S1 and S2 moments are very clear, while hearts with murmurs contain noise either between the moment s1 and s2, or between s2 and s1; the extras class contains an additional sound between s2 and s1, which are known as s3 and s4; the artifact class contains noise such as echoes, music, speech, these sounds being very noticeable since they are found in frequencies much higher than those produced by HS. This data set is useful for testing whether or not a patient's heart has murmurs, or not and if the sample contains too much noise that should be discarded. Yaseen's dataset has five classes, one normal class and four with different kinds of murmurs, the results show discrimination of healthy patients over those with cardiovascular diseases, and even though the murmur classes could coincide the class between S1 and S2 or between S2 and S1, it manages to classify eight to nine out of ten patients correctly.

4 Conclusions

Our tests have confirmed that spectrograms of entropy are very useful for diagnosing murmurs from phonocardiograms, even in the presence of high noise levels. This approach focuses on the distribution of information content across frequencies preserving information not even audible to humans. This allows the detection of temporal patterns in cardiovascular sounds that may persist even in noisy environments. Spectrograms of entropy emphasize areas of the signal with high information content. These visual representations are suitable inputs for convolutional neural networks and other machine-learning techniques based on images.

Entropygrams (i.e. Spectrograms of entropy) have not been used for the diagnosis of cardiovascular diseases to the best of our knowledge. When comparing the accuracy obtained using Entropygrams with that using energy spectrograms, MFCC spectrograms, and scalograms, which are the most widely used methods for converting audio signals to images for cardiovascular diagnosing, we found that entropygrams produce better results.

Diagnosing the specific murmur type for any valve failure remains a challenge and is ongoing work.

References

Alqudah, A. M., Alquran, H., & Qasmieh, I. A. (2020). Classification of heart sound short records using bispectrum analysis approach images and deep learning. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *9*(1), 66. https://doi.org/10.1007/s13721-020-00272-5

Banerjee, M., & Majhi, S. (2020). Multi-class heart sounds classification using 2D-convolutional neural network. In *2020 5th International conference on computing, communication and security (ICCCS)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCCS49678.2020.9277204

Bao, X., Xu, Y., Lam, H. K., Trabelsi, M., Chihi, I., Sidhom, L., & Kamavuako, E. N. (2023). Time-frequency distributions of heart sound signals: A comparative study using convolutional neural networks. *Biomedical Engineering Advances*, *5*, 100093. https://doi.org/10.1016/j.bea.2023.100093

Bentley, P., Nordehn, G., Coimbra, M., Mannor, S., & Getz, R. (2011). Classifying heart sounds challenge. *Retrieved from Classifying Heart Sounds Challenge:* http://www.peterjbentley.com/heartchallenge

Camarena-Ibarrola, A., Figueroa, K., & García, J. (2020). Speaker identification using entropygrams and convolutional neural networks. In *Advances in Soft Computing: 19th Mexican International Conference on Artificial Intelligence, MICAI 2020, Mexico City, Mexico, October 12–17, 2020, Proceedings, Part I 19* (pp. 23-34). Springer International Publishing. https://doi.org/10.1007/978-3-030-60884-2_2

Chakrabarti, T., Saha, S., Roy, S., & Chel, I. (2015). Phonocardiogram signal analysis-practices, trends and challenges: A critical review. In *2015 international conference and workshop on computing and communication (IEMCON)* (pp. 1-4). IEEE. https://doi.org/10.1109/IEMCON.2015.7344426

Chaudhuri, A., & Jayanthi, T. (2016). Diagnosis of cardiac abnormality using heart sound. *Biomedical Engineering: Applications, Basis and Communications*, *28*(05), 1650032. https://doi.org/10.4015/S1016237216500320

Chen, Y., Sun, Y., Lv, J., Jia, B., & Huang, X. (2021). End-to-end heart sound segmentation using deep convolutional recurrent network. *Complex & Intelligent Systems*, *7*, 2103-2117. https://doi.org/10.1007/s40747-021-00325-w

Chen, Y., Wei, S., & Zhang, Y. (2020). Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. *Medical & Biological Engineering & Computing*, *58*, 2039-2047. https://doi.org/10.1007/s11517-020-02218-5

Cheng, J., & Sun, K. (2023). Heart sound classification network based on convolution and transformer. *Sensors*, *23*(19), 8168. https://doi.org/10.3390/s23198168

Einthoven, W., & Geluk, M. A. J. (1894). Die Registrirung der Herztöne. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, *57*(12), 617-639. https://doi.org/10.1007/BF01662312

El-Hanjouri, M., Alkhaldi, W., Hamdy, N., & Alim, O. A. (2002). Heart diseases diagnosis using HMM. In *11th IEEE Mediterranean Electrotechnical Conference (IEEE Cat. No. 02CH37379)* (pp. 489-492). IEEE. https://doi.org/10.1109/MELECON.2002.1014641

Forbes, J., Laennec, R. T. H., Corvisart, J. N., Auenbrugger, L., & Collin, V. (1824). *Original Cases with Dissections and Observations Illustrating the Use of the Stethoscope and Percussion in the Diagnosis of the Chest: Also Commentaries on the Same Subjects Selected and Translated from Avenbrugger, Corvisart, Laennec and Others*. T. and G. Underwood.

Geddes, L. A. (2005). Birth of the stethoscope. *IEEE Engineering in Medicine and Biology Magazine*, *24*(1), 84-86. https://doi.org/10.1109/MEMB.2005.1384105

Groch, M. W., Domnanovich, J. R., & Erwin, W. D. (1992). A new heart-sounds gating device for medical imaging. *IEEE Transactions on Biomedical Engineering*, *39*(3), 307-310. https://doi.org/10.1109/10.125016

Han, W., Yang, Z., Lu, J., & Xie, S. (2018). Supervised threshold-based heart sound classification algorithm. *Physiological Measurement*, *39*(11). https://doi.org/10.1088/1361-6579/aae7fa

Hanna, I. R., & Silverman, M. E. (2002). A history of cardiac auscultation and some of its contributors. *The American journal of cardiology*, *90*(3), 259-267. https://doi.org/10.1016/S0002-9149(02)02465-7

Hooke, R. (1705). The posthumous works of Dr. robert hooke. *Journal of Signal and Information Processing*, 4:140–144.

Kumar, A. K., & Saha, G. (2018). Improved computerized cardiac auscultation by discarding artifact contaminated PCG signal sub-sequence. *Biomedical Signal Processing and Control*, *41*, 48-62. https://doi.org/10.1016/j.bspc.2017.11.001

Kumar, D., Carvalho, P., Antunes, M., Henriques, J., Eugénio, L., Schmidt, R., & Habetha, J. (2006). Detection of S1 and S2 heart sounds by high frequency signatures. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1410-1416). IEEE. https://doi.org/10.1109/IEMBS.2006.260735

Laennec, R. T. (1819). *De l'auscultation médiate: ou traité du diagnostic des maladies des poumons et du coeur* (Vol. 2).

Leatham, A. (1949). Phonocardiography. *Postgraduate Medical Journal*, *25*(289), 568. https://doi.org/10.1093/oxfordjournals.bmb.a074199

Lewis, T. (1920). *The mechanism and graphic registration of the heart beat*. Shaw.

Liang, H., Lukkarinen, S., & Hartimo, I. (1997). Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in Cardiology 1997* (pp. 105-108). IEEE. https://doi.org/10.1109/CIC.1997.647841

Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., Castells, F. Roig, J. M., Silva, I., ... & Clifford, G. D. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological measurement*, *37*(12), 2181. https://doi.org/10.1088/0967-3334/37/12/2181

Low, J. X., & Choo, K. W. (2018). Classification of heart sounds using softmax regression and convolutional neural network. In *Proceedings of the 2018 International Conference on Communication Engineering and Technology* (pp. 18-21). https://doi.org/10.1145/3194244.3194255

McKusick, V. A., Massengale, O. N., Wigod, M., & Webb, G. N. (1956). Spectral phonocardiographic studies in congenital heart disease. *British Heart Journal*, *18*(3), 403. https://doi.org/10.1136/hrt.18.3.403

McKusick, V. A., Reagan, W. P., Santos, G. W., & Webb, G. N. (1955). The splitting of heart sounds: A spectral phonocardiographic evaluation of clinical significance. *The American Journal of Medicine*, *19*(6), 849-861. https://doi.org/10.1016/0002-9343(55)90152-2

Meintjes, A., Lowe, A., & Legget, M. (2018, July). 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. 409-412). IEEE. https://doi.org/10.1109/EMBC.2018.8512284

Noman, F., Ting, C. M., Salleh, S. H., & Ombao, H. (2019). Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1318-1322). IEEE. https://doi.org/10.1109/ICASSP.2019.8682668

Oliveira, J., Renna, F., Costa, P. D., Nogueira, M., Oliveira, C., Ferreira, C., ... & Coimbra, M. T. (2021). The CirCor DigiScope dataset: from murmur detection to murmur classification. IEEE Journal of Biomedical and Health Informatics, 26(6), 2524-2535. https://doi.org/10.1109/JBHI.2021.3137048

Panah, D. S., Hines, A., & McKeever, S. (2023). Exploring the impact of noise and degradations on heart sound classification models. *Biomedical Signal Processing and Control*, *85*, 104932. https://doi.org/10.1016/j.bspc.2023.104932

Ren, Z., Chang, Y., Nguyen, T. T., Tan, Y., Qian, K., & Schuller, B. W. (2024). A comprehensive survey on heart sound analysis in the deep learning era. *IEEE Computational Intelligence Magazine*, *19*(3), 42-57. https://doi.org/10.1109/MCI.2024.3401309

Ren, Z., Cummins, N., Pandit, V., Han, J., Qian, K., & Schuller, B. (2018). Learning image-based representations for heart sound classification. In *Proceedings of the 2018 international conference on digital health* (pp. 143-147). https://doi.org/10.1145/3194658.3194671

Renna, F., Oliveira, J., & Coimbra, M. T. (2019). Deep convolutional neural networks for heart sound segmentation. *IEEE journal of biomedical and health informatics*, *23*(6), 2435-2445. https://doi.org/10.1109/JBHI.2019.2894222

Schmidt, S. E., Toft, E., Holst-Hansen, C., Graff, C., & Struijk, J. J. (2008). Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model. In *2008 Computers in Cardiology* (pp. 345-348). IEEE. https://doi.org/10.1109/CIC.2008.4749049

Sprague, H. B. (1962). Dr. George E. Fahr and His Era. *Circulation*, *26*(3), 439-444. https://doi.org/10.1161/01.CIR.26.3.439

Springer, D. B., Tarassenko, L., & Clifford, G. D. (2015). Logistic regression-HSMM-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, *63*(4), 822-832. https://doi.org/10.1109/TBME.2015.2475278

Subasi, A. (2019). *Practical guide for biomedical signals analysis using machine learning techniques: A MATLAB based approach*. Academic Press.

Takezaki, S., & Kishida, K. (2021). Construction of cnns for abnormal heart sound detection using data augmentation. In *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists* (pp. 20-22).

Torre-Cruz, J., Canadas-Quesada, F., Ruiz-Reyes, N., Vera-Candeas, P., Garcia-Galan, S., Carabias-Orti, J., & Ranilla, J. (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. https://doi.org/10.1016/j.jbi.2023.104475

Wells, F. (2014). *The Heart of Leonardo: Foreword by HRH Prince Charles, the Prince of Wales*. Springer Science & Business Media.

Wibawa, M. S., Maysanjaya, I. M. D., Novianti, N. K. D. P., & Crisnapati, P. N. (2018). Abnormal heart rhythm detection based on spectrogram of heart sound using convolutional neural network. In *2018 6th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE. https://doi.org/10.1109/CITSM.2018.8674341

World Health Organization 2024. Cardiovascular Diseases. Retrieved May 12, 2024 from https://www.who.int/health-topics/cardiovascular-diseases

Yaseen, Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences*, *8*(12), 2344. https://doi.org/10.3390/app8122344

Yildirim, M. (2022). Diagnosis of Heart Diseases Using Heart Sound Signals with the Developed Interpolation, CNN, and Relief Based Model. *Traitement du Signal*, *39*(3). https://doi.org/10.18280/ts.390316

Zhong, L., Wan, J., Huang, Z., Cao, G., & Xiao, B. (2013). Heart murmur recognition based on hidden Markov model. http://doi.org/10.4236/jsip.2013.42020