الجـــــــــــــــــــــمهـوريـــــة الجـــــــــــــزائـريـة الديمـــــــقـراطيـة

الـشـــــــــعبيـة

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

وزارة التـــــــــــــــــــــــــعليـم العـــــــالي و البـــــــــحث العـــــلـمـــي

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جـــــــــــــامعة أبي بكـــر بـلـقـايد – تـلمســـــــــــــان –

FACULTY OF TECHNOLOGY

DEPARTEMENT OF BIOMEDICAL ENGINEERING

DISSERTATION

Submitted for the **MASTER Degree's**

IN : Biomedical Engineering

Speciality : Electronics and Biomedical Maintenance

By : Elghribi Ibrahim and Hamlili Aymene

Theme

Parameters for Distinguishing Heart Sounds to Classify Normal and Pathological Cases

Publicaly supported on, 22 june 2025, at Tlemcen, Before the Jury Composed of :

LAZOUNI Sihem	MCB	University of Tlemcen	Chair
BENSMAIL Ilham	MCB	University of Tlemcen	Examiner
Baakek Yettou Nour El Houda	MCA	University of Tlemcen	Supervisor
Debbal Imane	PhD in biomedical engineering	University of Tlemcen	Co-Supervisor

Academic Year: 2024 / 2025

Acknowledgements

الحمد لله والصلاة السلام على رسول الله صلى الله عليه وسلم

First and foremost, we would like to express our deep gratitude to God Almighty, the Most Merciful, for granting us the strength and patience to bring this humble work to completion.

We extend our heartfelt thanks to our supervisor, **Dr. Baakek Yettou Nour El Houda**, Our professor throughout these five years, who guided us during the entire course of this project. We are sincerely grateful for her continuous support, for proposing this subject, and for her invaluable advice and encouragement throughout this work.

We also wish to express our sincere appreciation to our co-supervisor, **Dr. Debbal Imane**, For her exemplary mentorship during both the project and the writing of this thesis. Her unwavering availability, the time she devoted to us, and her insightful feedback played a crucial role in helping us reach our goals and significantly improve the quality of our research. For all of this and more, we are truly grateful.

To **Mrs. Lazzouni Sihem**, who kindly accepted to serve as president of our thesis jury we express our respectful appreciation and gratitude.

To Mrs. BENSMAIL Ilham, For accepting to be a member of the jury and for reviewing our work with attention and care we thank you warmly.

Finally, we extend our sincere thanks to all the professors and staff of our department who have contributed to our education and personal growth throughout these years.

Dedication

I dedicate this work to our dear parents, who have been my pillars and supported me at every stage of my life. Their unconditional love and unwavering support have been my constant source of inspiration.

To my brothers **Imed**, **Achref**, and **Yasser**,

To my beloved **Grandmother**, who has always been there for me with her prayers and encouragement,
To my dear uncle **AISSAM**, for all his support and care.

And to every member of my family, young and old, who contributed in their own way — this achievement is for you.

To my dear friend and project partner **Ibrahim**, and our close friend **Mohamed**,
To all my university friends with whom I shared joyful and unforgettable moments thank you from the bottom of my heart.

Aymene.

Dedication

I dedicate this work with great joy and an open heart to my dearest parents, who have always supported and encouraged me with their presence and active involvement throughout my studies. A special thank you to them.

To my dear sister **Hadjer** and my dear brother **Abdelkrim**, who are always there for me with moral support,

To my uncle Ismail, for his unconditional support, To my entire family I dedicate this final year project to all of you.

To my dear friend and project partner **Aymen**, and to my dear friend **Mohamed**, with whom I've shared unforgettable moments throughout these five years.

Finally, to everyone mentioned above and those I may not have named, To all who have helped me at any point during this journey a heartfelt thank you.

To all our professors who shared their knowledge with us over these years of study, And to anyone who supported us directly or indirectly please accept our deepest gratitude and appreciation.

Ibrahim.

Abstract:

The phonocardiogram (PCG) signal is a recording of heart sounds with useful and significant mechanical activity of the human heart information. In this work, we employed signal processing methods such as the Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), the Bispectral analysis, and MFCCs to handle these data. These methods are used on our already filtered and labelled dataset as either normal or abnormal. S1, S2, and murmurs were segmented using an automatic approach. From these segments, we extracted multiple parameters and combined them in one feature matrix including the maximum amplitude, dominant frequency, weighted centroid, energy, phase entropy, and more. These features served as training features for machine learning algorithms such as the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) for a binary classification of normal and pathological heart sounds. Furthermore, we produced 2D images from the analysing techniques to feed a CNN-SVM hybrid model, were we reached quite important accuracy results for different classification purposes.

Keywords: CNN-SVM, Binary classification, KNN, SVM, FFT, CWT, Bispectral, MFCC, Phonocardiogram signal.

Résumé :

Le signal du phonocardiogramme (PCG) est un enregistrement des bruits cardiaques avec des informations utiles et significatives sur l'activité mécanique du cœur humain. Dans ce travail, nous avons utilisé des méthodes de traitement du signal telles que la transformée de Fourier rapide (FFT), la transformée en ondelettes continue (CWT), l'analyse bispectrale et les MFCC pour traiter ces données. Ces méthodes sont utilisées sur notre ensemble de données déjà filtrées et étiquetées comme normales ou anormales. Les segments S1, S2 et les souffles ont été identifiés à l'aide d'une approche automatique. À partir de ces segments, nous avons extrait plusieurs paramètres et les avons combinés dans une matrice de caractéristiques comprenant l'amplitude maximale, la fréquence dominante, le centroïde pondéré, l'énergie, l'entropie de phase, etc. Ces caractéristiques ont servi de caractéristiques d'entraînement pour les algorithmes d'apprentissage automatique tels que le K-Nearest Neighbor (KNN) et le Support Vector Machine (SVM) pour une classification binaire des bruits cardiaques normaux et pathologiques. En outre, nous avons produit des images 2D à partir des techniques d'analyse pour alimenter un modèle hybride CNN-SVM, ce qui nous a permis d'obtenir des résultats de précision assez importants pour différents objectifs de classification.

Mots-clés : CNN-SVM, classification binaire, KNN, SVM, FFT, CWT, Bispectral, MFCC, signal phonocardiogramme.

ملخص:

إن إشارة المخطط الصوتي للقلب (PCG) عبارة عن تسجيل لأصوات القلب مع معلومات النشاط الميكانيكي المفيدة والمهمة لقلب الإنسان. في هذا العمل، استخدمنا طرق معالجة الإشارات مثل تحويل فورييه السريع (FFT)، وتحويل الموجيات المستمر (CWT)، والتحليل ثنائي الطيف والتحليل ثنائي الطيف الترددي (MFCCs) للتعامل مع هذه البيانات. يتم استخدام هذه الأساليب على مجموعة البيانات التي تمت تصفيتها وتصنيفها بالفعل على أنها طبيعية أو غير طبيعية. تم تجزئة S1 و S2 والنفخات باستخدام نهج تلقائي. ومن هذه المقاطع، استخرجنا معلمات متعددة ودمجناها في مصفوفة ميزات واحدة بما في ذلك السعة القصوى والتردد السائد والنقطة المركزية الموزونة والطاقة وإنتروبيا الطور وغيرها. كانت هذه الميزات بمثابة ميزات تدريب لخوارزميات التعلم الآلي مثل K-Nearest Neighbor (KNN)) وآلة دعم المتجهات (SVM) لتصنيف ثنائي لأصوات القلب الطبيعية والمرضية. وعلاوة على ذلك، أنتجنا صورًا ثنائية الأبعاد من تقنيات التحليل لتغذية نموذج هجين من CNN-SVM، حيث توصلنا إلى نتائج دقة مهمة جدًا لأغراض التصنيف المختلفة.

الكلمات المفتاحية: CNN-SVM، تصنيف ثنائي، KNN، SVM، FFT، CWT، CWT، Bispectral، MFCC، إشارة مخطط صوتيات القلب.

Table of Contents

Chaptre. II: Classification of Heart Sounds Using Signal Processing Techniques

Chaptre. III: HEART SOUNDS CLASSIFICATION USING NEURONAL NETWORKS AND SPECTRUMS

Table of Figures

Chaptre. I: THEORITICAL ASPECTS OF THE PHONOCARDIOGRAM

Chaptre. II: Classification of Heart Sounds Using Signal Processing Techniques

Chaptre. III: HEART SOUNDS CLASSIFICATION USING NEURONAL NETWORKS AND SPECTRUMS

List of Tables

Chaptre. I: THEORITICAL ASPECTS OF THE PHONOCARDIOGRAM

Chaptre. II: Classification of Heart Sounds Using Signal Processing Techniques

Chapter. III: HEART SOUNDS CLASSIFICATION USING NEURONAL NETWORKS AND SPECTRUMS

Key Words and Abbreviations

Abbreviation	Definition
PCG	Phonocardiogram
ECG	Electrocardiogram
S1	First heart sound
S2	Second heart sound
S3	Third heart sound
S4	Fourth heart sound
M1	Mitral valve closure component of S1
T1	Tricuspid valve closure component of S1
A2	Aortic valve closure component of S2
P2	Pulmonary valve closure component of S2
FFT	Fast Fourier Transform
CWT	Continuous Wavelet Transform
MFCC	Mel-Frequency Cepstral Coefficients
HSMM	Hidden Semi-Markov Model
CNN	Convolutional Neural Network
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
STFT	Short-Time Fourier Transform
DCT	Discrete Cosine Transform
QPC	Quadratic Phase Coupling
WC	Weighted Centroid
EntPh	Bispectral Phase Entropy
VSD	Ventricular Septal Defect
MR	Mitral Regurgitation
MS	Mitral Stenosis
TR	Tricuspid Regurgitation
AS	Aortic Stenosis
AR	Aortic Regurgitation
PAS	Pulmonary Artery Stenosis
MVP	Mitral Valve Prolapse
AFib	Atrial Fibrillation
LLSB	Lower Left Sternal Border
ICS	Intercostal Space

General Introduction

Cardiovascular diseases, including heart valve disorders, remain a leading global health challenge. Recent statistics show that about one-third of all deaths worldwide [1] are related to heart diseases. Often quiet until they become severe, valvular heart diseases in which one or more heart valves fail—cause a great clinical burden. Thus, efficient treatment depends on timely identification and monitoring of such diseases. Early diagnosis has always mostly relied on auscultation with a stethoscope as the main clinical technique. René Laënnec's development of the stethoscope in 1816 transformed cardiac evaluation by providing a practical means of acoustic assessment of heart sounds . Auscultation is naturally limited, though; heart sounds are transient and complicated, and the human ear even with stethoscope guidance may not be able to distinguish overlapping sounds [2] or detect faint murmurs. In response, more advanced imaging techniques such as cardiac catheterisation and Doppler echocardiography have been developed, although these call for specific equipment and knowledge.

Advancements in biomedical signal processing have led to phonocardiography (PCG) – the electronic documentation of cardiac sounds as a potent diagnostic supplement. Phonocardiography is a non-invasive technique that provides a graphic record of the timing and amplitude of heart sounds and murmurs throughout the cardiac cycle. By capturing this information, digital PCG complements traditional auscultation and can reveal features of cardiac pathology (e.g. valve stenosis or regurgitation) that may not be obvious to the unaided ear [3].

The development of artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), marks a paradigm change in the study of biological signals. Because they can model complicated, non-linear relationships and find hidden patterns in data, these technologies have become crucial instruments in the categorisation of phonocardiogram (PCG) signals.

In this study, we investigate advanced signal processing techniques and machine learning approaches comprising feature extraction methods like Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), bispectral analysis, and Mel-Frequency Cepstral Coefficients (MFCC). We generated large datasets for training conventional machine learning algorithms including K-Nearest Neighbours (KNN) and Support Vector Machines (SVM), hence proving strong classification capability using these approaches. We also investigated deep learning techniques using a CNN-SVM hybrid model, combining the robust classification strengths of SVM with the automated feature extraction powers of CNN, improving by that the diagnostic accuracy.

In Precedent works , A.Debbal & F.Berekesi Reguig in Frequency analysing of Heartbeat Sounds (2008) noted that the FFT analysis as like a basic methods for frequency analysis that helps clinics to extract features and measurement of PCG .Yet it limitation remains in time domain. F.Meziani,

in her PhD Thesis (2012) used the Wavelet analysis (WT) in his two specific methods CWT and DWT. As known, the WT can show the Frequency-time domains and permit a binary classification through efficient PCG features, To visualize PCG signals in 2D and perform spectro-temporal analysis, the bispectrum has revealed novel features that may be crucial for PCG diagnosis, as shown in studies like those of I. Debbal and Boudis. Moreover, heart sounds are audible signals, similar to human voice, which is why recent works — such as Liu et al. (2021) — have employed MFCCs to extract audio-based features from PCG signals

After feature extraction, machine learning models such as KNN and SVM have been used to classify PCG signals, achieving high accuracy in binary classification, as demonstrated by Langley et al. (2017). For image-based classification, deep learning approaches like Convolutional Neural Networks (CNNs) have also shown excellent performance, as reported by D. Mebarki (2023) .

Our work will be divided in three important parts by following:

Chapter I: Theoretical Features of the Phonocardiogram:

The theoretical background required for comprehending our work is established in this chapter. With a special emphasis on the cardiac cycle and the related heart sounds (S1, S2, S3, S4). The chapter gives a general review of heart anatomy and operation. An introduction of basic ideas about the PCG signal is given together with its temporal and frequency-domain properties as well as possible definitions of abnormal heart sounds.

Chapter II: Heart Sound Classification Using Various Analysis Techniques

This chapter explores both traditional and advanced signal processing methods for analysing PCG signals. We begin with pre-processing and database preparation then the segmentation, followed by the application of various feature extraction techniques:

- Fast Fourier Transform (FFT), used to analyse maximum amplitude and spectral bandwidth.
- Continuous Wavelet Transform (CWT), using two different methods to extract time–frequency characteristics.
- Bispectral analysis.
- Mel-Frequency Cepstral Coefficients (MFCC).

We then focus on extracting key parameters such as:

- Phase entropy, which quantifies the signal's complexity;
- Spectral centroid (absolute bispectrum value of Y axis), which characterizes the energy distribution of the signal.

- Max amplitude and frequency domain .
- Weighted Center

All these features are compiled into a matrix and used to train ML algorithms (KNN and SVM) for the classification of heart sounds. The results are discussed in detail, highlighting the strengths and limitations of each approach.

Chapter III: Heart Sound Classification Using Neural Networks and Spectral Analysis

In the final chapter, we shift our focus to modern approaches based on neural networks and spectral image analysis of PCG signals, particularly hybrid CNN-SVM models. We begin with a literature review of recent works employing these techniques and proceed with an in-depth description of neural network architectures and types. The obtained results are compared with those of traditional methods, followed by a detailed discussion evaluating their effectiveness in classifying pathological heart sounds.

Chapter I:

THEORITICAL ASPECTS OF THE PHONOCARDIOGRAM

Chapter I: THEORITICAL ASPECTS OF THE PHONOCARDIOGRAM

Introduction:

The heart is a complicated organ whose diagnosis depends on a thorough awareness of its mechanical, electrical, and acoustic activity. Every pulse creates a series of sounds that mirror blood flow dynamics and appropriate operation of heart valves. Our effort focusses on heart sounds to differentiate them into normal and aberrant cases, so helping to identify cardiovascular disorders.

First, one must grasp the anatomy and physiology of the heart before delving into engineering tools including signal processing and machine learning. Following a thorough review of cardiac structure and function in this chapter, heart sounds (S1, S2, S3, S4) and pathological instances are discussed in great detail here. We also present the Phonocardiogram (PCG), a non-invasive instrument for capturing and evaluating heart sounds, so clarifying the correspondence between each sound and a particular cardiac event.
The signal processing and classification techniques investigated in the next chapters will be built upon this theoretical background.

Heart Basics Anatomy:

The Heart:

Nestled between the two lungs, the sternum, and the spine in the centre portion of the thoracic cavity, the muscular organ known as the heart is It is rather bigger in men than in women. Comprising three separate layers—walls—from the inside to the outside—the endocardium, the myocardium or cardiac muscle, and the epicardium [4], the heart is a hollow organ.

Two distinct entities make up it: the left and the right hearts. Separated by the tricuspid valve for the right and the mitral valve for the left heart, every heart consists of a ventricle in the lower part and an atrium in the upper part. The wall separating the ventricles is the interventricular septum; the wall separating the atria is the interational septum. Between the atria and ventricles, there is no link of cardiac (muscular) tissue. There is thus no blood flow between the upper and lower sections. This passes via the heart valves; the flow is unidirectional, from the atrium to the ventricle.

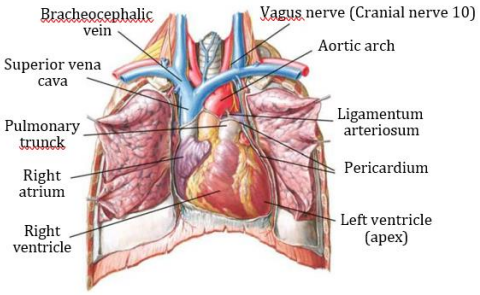


Figure I 1 Anterior view of the heart

Maintaining appropriate circulation of oxygenated blood throughout the body's vascular network is the job of the heart:

- The right atrium gets venous blood brought by the vena cava and propels it into the right ventricle, which, upon contraction, sends the blood to the lungs via the pulmonary artery (the only artery that carries oxygen-poor blood).
- The right side of the heart sends oxygen-poor blood to the lungs to remove carbon dioxide and reoxygenate the blood.
- Now oxygenated in the lungs, the blood returns via the four pulmonary veins to the left side of the heart at the left atrium.
- The blood then moves via the mitral valve, which regulates flow, into the left ventricle.
- Once contracted, the heart pushes the blood across the aortic valve and subsequently into the aorta, the biggest blood vessel in the body, so distributing it across the whole arterial network.

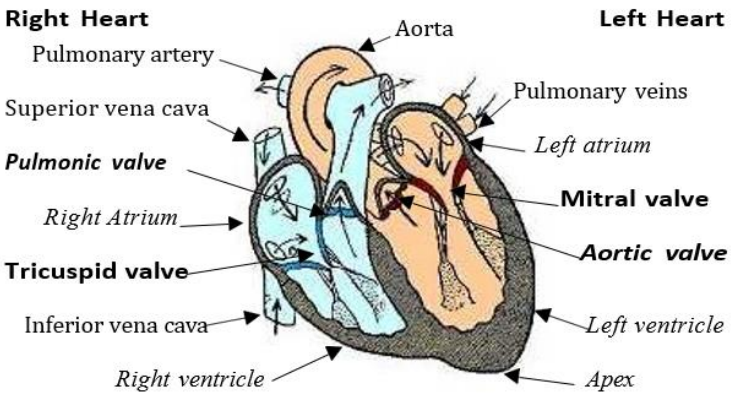


Figure I 2 Diagram of the heart [4]

Right Atrium :

Comprising the heart chamber in charge of receiving deoxygenated blood from the superior and inferior vena cava and the coronary sinus, the right atrium directs it to the right ventricle via the tricuspid valve [3]. Its role affects heart sounds during the cardiac cycle, especially S1, which results from the tricuspid and mitral valves closing at the start of ventricular systole [4]. While tricuspid valve diseases can affect S1 intensity and timing, S2 caused by aortic and pulmonary valve closure is more linked to ventricular activity [5]. These changes could point to disorders like tricuspid stenosis or regurgitation, so influencing auscultation results and heart haemodynamic. Dosing cardiovascular diseases depends on an awareness of the function of the right atrium in heart sounds.

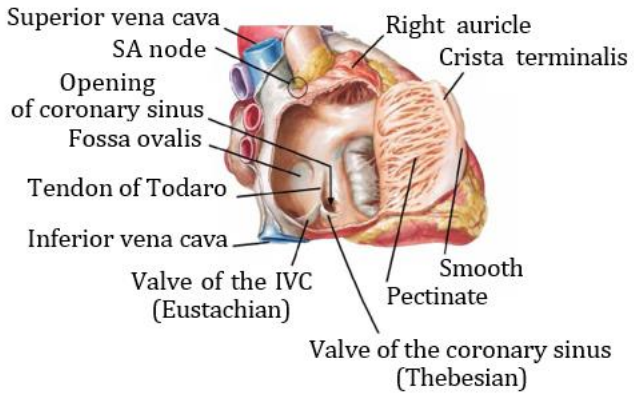


Figure I 3 Right atrium [4]

Tricuspid Valve:

The tricuspid valve is a vital component of the heart, located between the right atrium and the right ventricle. It operates like a one-way valve, stopping the return of blood during ventricular contraction. Anatomically, the tricuspid valve is made up of three leaflets: anterior, posterior, and septal which are connected to the right ventricular wall by chordae tendinous linked to papillary muscles. This design guarantees that the valve shuts tightly during systole, stopping regurgitation into the right atrium. The correct operation of the tricuspid valve aids in S1 [5] . This sound is produced by the closure of both the tricuspid and mitral valves at the beginning of ventricular systole. Irregularities in the tricuspid valve, like tricuspid regurgitation, can result in modified heart sounds and might signify underlying heart issues.[6].

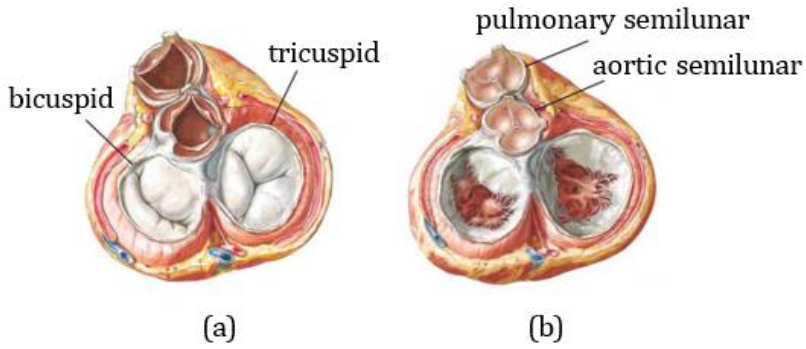


Figure I 4 Intracardiac valves: (a) systole and (b) diastole [4]

Right ventricle:

The right ventricle pumps blood from the right atrium to the lungs via the pulmonary trunk. Its wall has trabeculae carnae, and it's separated from the left ventricle by the infundibular septum. The conus arteriosus lies below the semilunar valves.

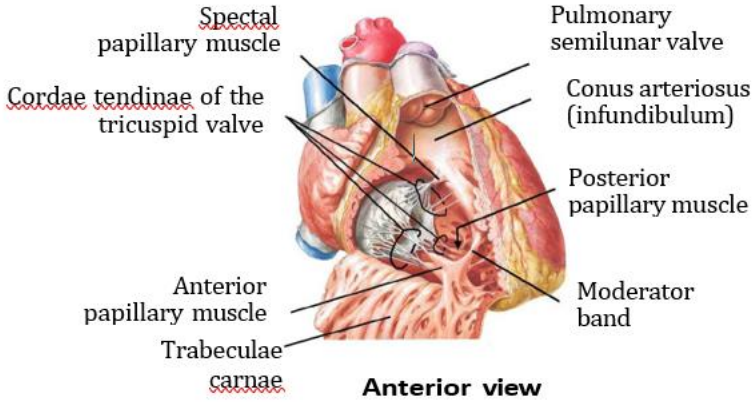


Figure I 5 Right ventricle

The Pulmonary Valve :

Comprising the right ventricle and the pulmonary trunk, the pulmonary semilunar valve is situated By avoiding regurgitation of blood back into the right ventricle after systole, it is absolutely essential in guaranteeing unidirectional blood flow. Three cusps—anterior, left, and right—anchored to a fibrous annulus at the junction of the pulmonary trunk and the right ventricular infundibulum define structurally the valve. The second heart sound (S2) reflects the end of right ventricular ejection [5][6], so proper closure of this valve is related with S2.

. Left Atrium:

Oxygenated blood returning from the lungs by the pulmonary veins finds a reserve in the left atrium. Except for the pectinate muscles of the atrial appendage, its inner wall is mostly smooth. Remarkably similar to the embryonic foramen ovale, the atrial septum comprises the fossa ovalis. In pathological conditions, any aberrant pressure shift or flow turbulence across this area might produce extra heart sounds including the third (S3) or fourth (S4). [5]

. Left Ventricle:

Receiving oxygen-rich blood from the left atrium, the left ventricle is in charge of pushing it via the aortic valve into the systemic circulation. The dense myocardium of this chamber distinguishes it anatomically since it is necessary to provide the high pressures required for systemic perfusion. Comparatively to the right ventricle, the left ventricle's interior surface consists of trabeculae carneae, finer and more homogeneous muscle ridges. The interventricular septum, which bends towards the right ventricle, affects its unique barrel form. The initial heart sound (S1) is produced in great part by the closing of the mitral valve and consequent contraction of the left ventricle [5][6]

Cardiac revolution:

cardiac cycle and heart sounds Usually separated into three main phases—atrial systole, ventricular systole, and diastole—the cardiac cycle is The atria squeeze in atrial systole to drive blood into the relaxed ventricles. Following atrial ejection, the closing of the atrioventricular valves (mitral and tricuspid) stops backflow and generates the first heart sound (S1), therefore initiating ventricular systole. The phase when the ventricles contract to drive blood into the aorta and pulmonary artery is known as ventricular systole. The semilunar valves—pulmonary and aortic—close to stop retrograde flow once blood is evacuated. The second cardiac sound (S2) produced by closing this valve indicates diastole's start. The phase of total relaxation is diastole, in which case the ventricles fill passively with blood coming from the atria. Examining the sound events and timing of S1 and S2 divides the cycle into systolic (from S1 to S2) and diastolic (from S2 to the next S1) phases, therefore offering vital diagnostic information on valve performance and cardiac haemodynamics [5][6].

Theoretical notions about the phonocardiogram:

Phonocardiogram (PCG) Signal:

The phonocardiogram (PCG) is a graphical representation of the acoustic signals generated by the mechanical activity of the heart over time. It captures and visualizes heart sounds such as S1, S2, and possible murmurs, making it possible to localize the events within the cardiac cycle, identify the

number of sound components, characterize their frequency content, and assess the severity of potential pathological murmurs.

Phonocardiography is based on recording the vibrations or oscillations of varying frequencies—whether audible or not—originating from normal or abnormal heart sounds. When recorded and plotted in waveform format, these sounds form what is called a phonocardiogram. The device used to capture and record these signals is known as a phonocardiograph.

Analyzing PCG signals serves as a complementary tool to auscultation, refining and confirming the auditory findings of a physician. It provides enhanced insight into the timing and nature of cardiac events, particularly when diagnosing pathological conditions such as valve dysfunctions or heart murmurs. The PCG thus helps correlate pathological features with the normal heart sound cycle [7].

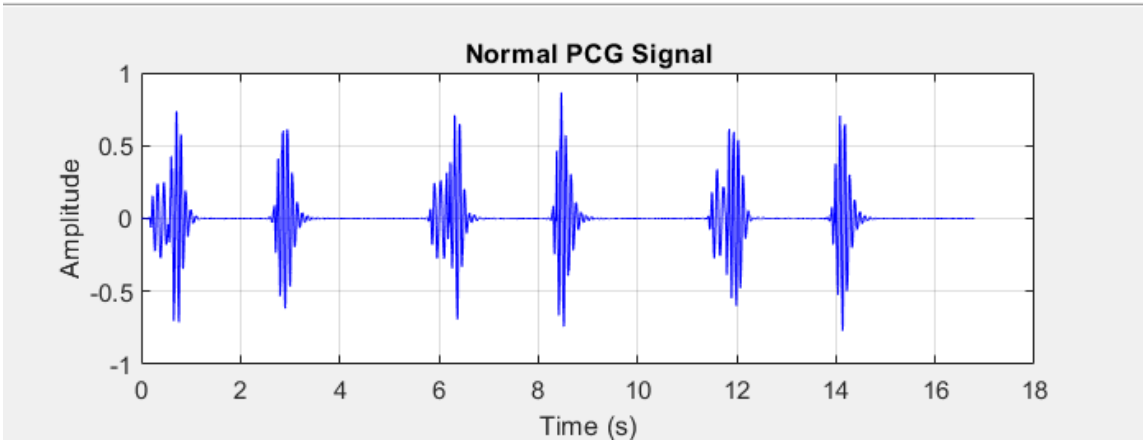


Figure I 6 Normalized PCG signal for a healthy subject .

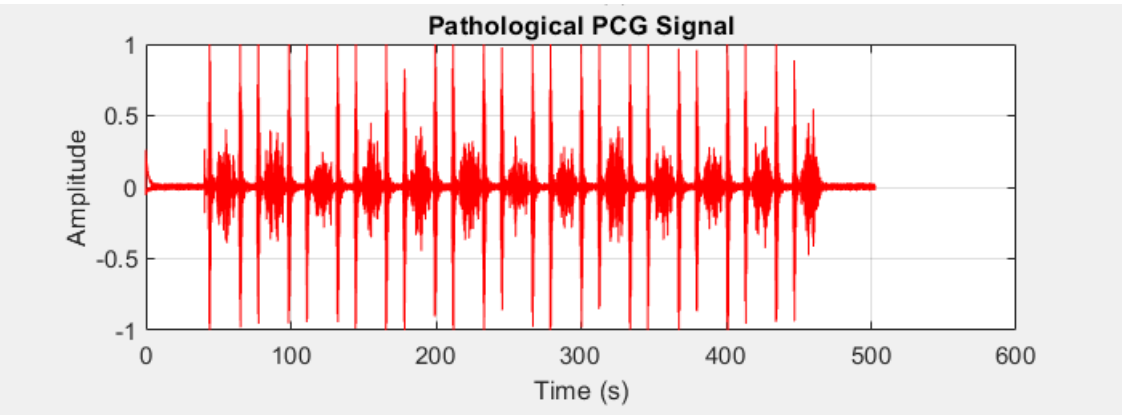


Figure I 7 Pathological PCG Signal For Murmur

Heart Sounds Description:

In a healthy case, the heart produces four heart sounds, S1 and S2, which occur during each cardiac cycle, and S3 and S4, with significantly lower amplitudes than the first two sounds [4/7]. The latter two are sometimes early signs of pathology or are related to age (childhood or old age).

Audible Sounds:

Normally, only the first and second sounds (S1 and S2) are clearly audible with a stethoscope. These sounds resemble the onomatopoeia "tic tac.". (lub-dub).

First heart sound S1:

Origins of S1:

The first heart sound (S1) is due to the closure of the mitral (M1) and tricuspid (T1) valves, along with blood flow through their leaflets. S1 marks the start of systole, followed by aortic valve opening. In regular auscultation, it is the loudest at the apex due to the valves' movement and sudden stopping of blood creating a shock wave during the ventricular contraction.

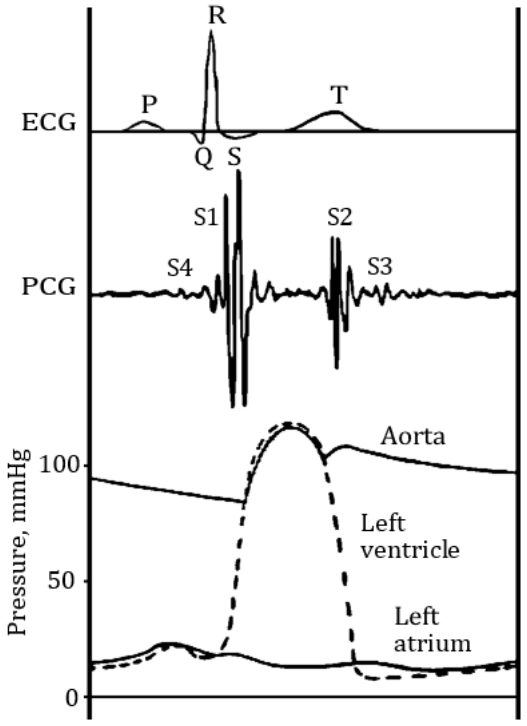


Figure I 8 Temporal relationship between the Electrocardiogram (ECG), the Phonocardiogram (PCG) and left atrial, left ventricular, and aortic pressures. [8]

The T1 component (tricuspid valve closure) is softer than M1 (mitral valve closure) and is best heard at the lower left sternal border. S1 is louder than S2 at the apex, influenced by ventricular pressure, valve structure, and position. S1 intensity varies in conditions like hyperkinetic states, mitral stenosis, or heart failure, and can split in rare cases due to delayed tricuspid closure , S1 duration is ~100 ms.

Split S1 heart sound :

S1 splitting is rare in normal subjects due to the softer tricuspid closure (T1) compared to mitral closure (M1). It can occur with a ~30 ms delay, heard at the lower left sternal border. Pathological S1 splitting involves delayed tricuspid closure.

Second Heart Sound S2:

Origins of S2 heart sound :

The second heart sound (S2) is caused by the closure of the aortic (A2) and pulmonary (P2) valves. A2 is louder than P2 due to higher left ventricular pressure and is heard at the right upper sternal border, while P2 is localized at the left upper sternal border. S2 marks the start of diastole when ventricular pressure drops below aortic/pulmonary pressure. It is best heart at the heart's base.

In normal case S2 is higher in frequency and shorter in duration than S1. At the cardiac base, it is louder than S1. A2 (aortic valve closure) is louder and heard widely, while P2 (pulmonary valve closure) is softer and localized. Increased P2 intensity (e.g., in pulmonary hypertension) makes S2 louder than S1 at the apex. Both A2 and P2 last <60 ms.

Split S2 heart sound :

The S2 heart sound can be separated into two components in relation to aortic and pulmonary valves closures. This phenomenon can be observed in individuals less than 50 years old rather than older subjects. A split S2 can be best heard at pulmonic auscultation area (left upper sternal border). it can be persistent (widened), fixed, or paradoxical (reversed).

Paradoxical splitting of S2 heart sound

A paradoxical split of S2 occurs when A2 (aortic closure) is delayed, merging with P2 (pulmonary closure) during expiration. It's caused by conditions like aortic stenosis, hypertrophic cardiomyopathy (HOCM), or left bundle branch block (LBBB), where A2 follows P2, narrowing with inspiration and widening with expiration (Add graphic representation of each pathology).

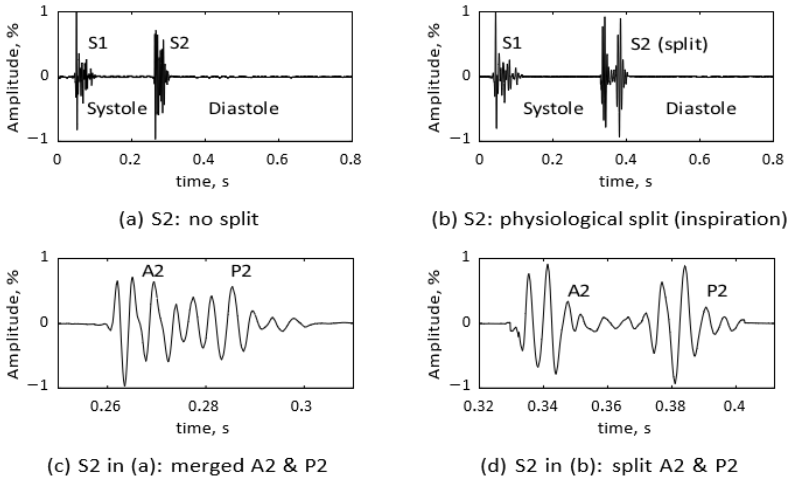


Figure I 9 Physiological split of S2. (a): unsplit S2 during normal respiration, (b): split S2 during inspiration, (c): S2 in (a); A2 & P2 are merged, (d): S2 in (b); P2 is delayed from A2 [9]

Pathological split of S2 heart sound:

Wide fixed splitting of S2 is constant during respiration and linked to atrial septal defects. Increased right ventricular output and reduced pulmonary impedance cause a prolonged right ventricular ejection, creating a wide split. Blood shunting through the septal defect maintains equilibrium, minimizing respiratory effects.

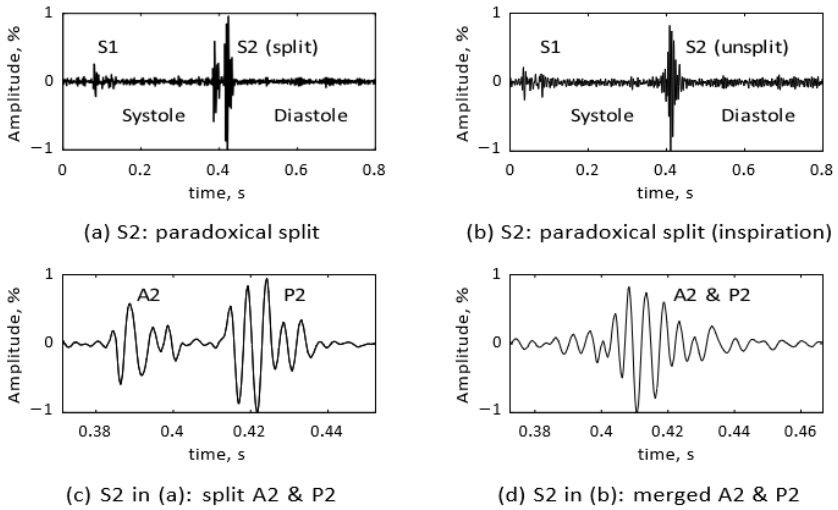


Figure I 10 Paradoxical split of S2. (a): paradoxical split of S2 during normal respiration, (b): S2 merged during inspiration, (c): S2 in (a), (d): S2 in (b) [1].

NAUDIBLE SOUNDS:

Third heart sound S3:

The third heart sound (S3), or ventricular gallop, occurs during early diastole due to rapid ventricular filling. It's low-frequency and best heard at the apex in left lateral decubitus. S3 is normal in children, athletes, and pregnant women but pathological in adults over 40, indicating conditions like heart failure or valvular regurgitation.

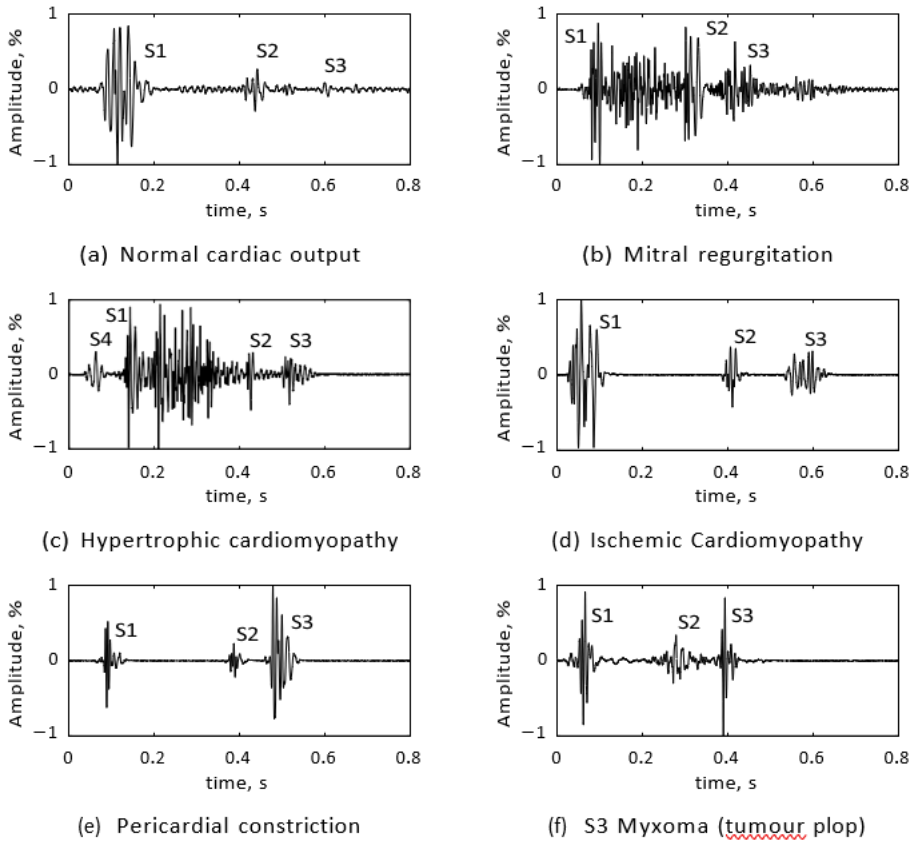


Figure I 11 Third heart sound (S3) [9]

Mitral Regurgitation:Third Sound and Diastolic Rumble Hypertrophic Cardiomyopathy:

Under severe systolic backflow, a considerable regurgitated blood inflow returns rapidly to the ventricle. Therefore, the mitral regurgitation causes a holosystolic murmur as well as a mid–diastolic rumble and an S3 heart sound (Figure I.20(b)).

Hypertrophic Cardiomyopathy:

The rapidly ejected left ventricular blood causes mitral regurgitation perceived as a systolic murmur. An S3 heart sound is produced under mitral regurgitated diastolic inflow going back to the ventricle which returns to its passive capacity. An atrial contraction augments the ventricular volume and produces a presystolic S4 heart sound (Figure I.20(c)).

Ischemic Cardiomyopathy:

An S3 heart sound is produced by the resistance of the dilated ventricle to passive filling. This ventricular dysfunction, which is similar to dilated cardiomyopathy, causes lower ejection (Figure I.20(d)).

Pericardial Constriction: Pericardial Knock:

passive filling of a restricted ventricle by endomyocardial disease or constrictive pericarditis generates an S3 heart sound. In constrictive pericarditis, the S3 heart sound is sharper than usual thud sound of a full ventricle and thus is also described as pericardial knock. A pericardial knock produced by the right ventricle usually rises with inspiration (Fig- ure I.20(e)).

Myxoma: Tumour Plop:

A left atrial myxoma20 produces an S3 heart sound. The left atrial myxoma moves to and from the atrium to the ventricle through the mitral valve. Movement of this tumor produces an S3 heart sound which is known as "tumor plop" (Figure I.20(f)).

Fourth heart sound S4:

The fourth heart sound (S4), or atrial gallop, occurs 40–120 ms before S1 due to forceful atrial contraction into a stiff ventricle. It's low-pitched, heard at the apex (left ventricle) or 4th left intercostal space (right ventricle), and indicates pathological ventricular stiffness. S4 disappears with standing and is absent in atrial fibrillation.

Extra–systolic sounds:

Extra-systolic sounds include ejection sounds (early systole, caused by aortic/pulmonic stenosis) and mid-systolic clicks (late systole, caused by mitral/tricuspid valve prolapse). Both are high-pitched and result from abrupt valvular stretching at maximum distensibility.

Ejection sounds:

Ejection sounds occur 40–60 ms after S1, are high-pitched, and short in duration. They are vascular (blood ejection into vessels) or valvular (aortic/pulmonary valve stretching). In aortic/pulmonic stenosis, they're heard at specific auscultation areas and often indicate pathology when accompanied by a murmur.

Bicuspid aortic valve:

In aortic or pulmonic stenosis, blood ejection creates a sharp ejection sound just after S1, making S1 appear split. A mid-systolic murmur results from turbulent blood flow across the stenosed valve.

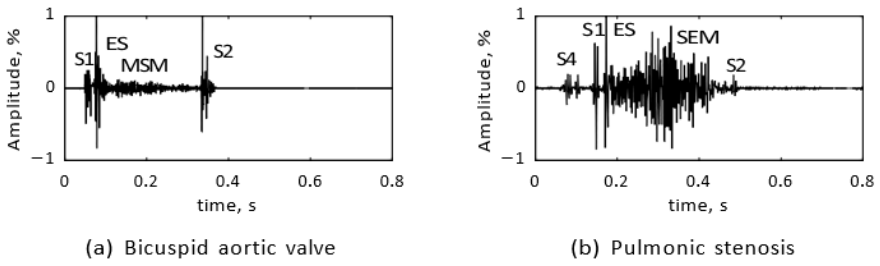


Figure I 12 Ejection sounds: (a) Bicuspid aortic valve, (b) Pulmonic stenosis [9].

Pulmonic stenosis:

In conditions like right ventricular hypertrophy, the ejection sound merges with S1 during inspiration and separates during expiration. It's accompanied by an S4, a mid-systolic murmur, and a delayed/soft pulmonary component (P2).

Aortic regurgitation:

In aortic regurgitation, the ejection sound results from aortic valve opening and root stretching during ventricular ejection, similar to aortic stenosis. It's accompanied by an early diastolic murmur, loudest at the left sternal border, and a split S1 heard best at the apex.

Mid–systolic clicks:

The mid-systolic click (MSC) is a high-pitched sound caused by chordae tendineae stretching in mitral or tricuspid valve prolapse.

Mobile Click: Standing and Squatting:

A systolic click results from increased blood volume on prolapsed valves. Unlike ejection sounds, its timing changes with posture: standing brings it earlier in systole, while squatting delays it due to increased ventricular volume and valve expansion.

Mitral prolapse:

The non-ejection click occurs in mid/late systole due to mitral valve prolapse into the left atrium. Unlike ejection sounds, its timing varies with ventricular filling changes (e.g., posture or respiration). A late murmur follows in severe mitral valve prolapse, indicating valve incompetence.

Opening sounds:

The opening snap (OS) is a high-pitched sound in early diastole, caused by mitral or tricuspid valve opening in stenosis. It's heard at specific auscultation areas and differs from S2 splitting by being louder at the apex than the 2nd left intercostal space.

Timing of murmurs:

Heart murmurs vary in shape and timing, occurring in systole (e.g., aortic stenosis, mitral regurgitation), diastole (e.g., aortic regurgitation, mitral stenosis), or as continuous murmurs (e.g., patent ductus arteriosus).

Systolic murmurs:

Systolic murmurs vary in timing and location, indicating specific pathologies. Early/holosystolic murmurs suggest mitral/tricuspid regurgitation or ventricular septal defect, while mid-systolic (ejection) murmurs indicate outflow obstruction. Late systolic murmurs point to mitral prolapse or hypertrophic cardiomyopathy.

Early systolic murmurs:

Holosystolic (pansystolic) murmurs start with S1 and continue through S2, caused by retrograde blood flow from high to low-pressure chambers (e.g., mitral/tricuspid regurgitation or ventricular septal defect). They indicate backward flow through atrioventricular valves.

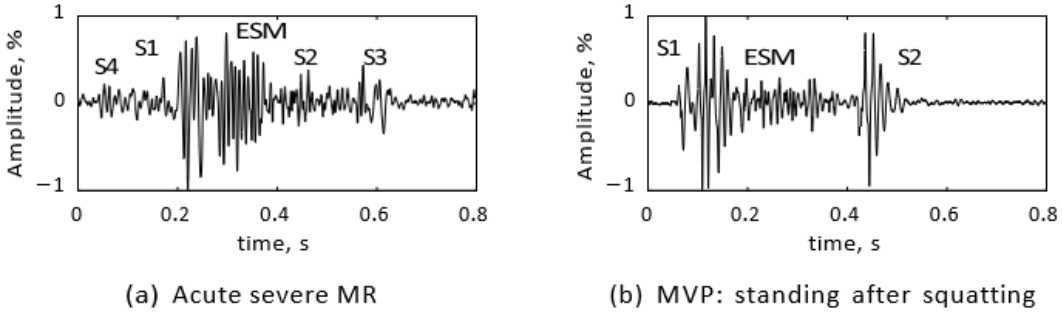


Figure I 13 Systolic murmurs: early systolic murmurs (ESM). (a): Acute severe mitral regurgitation (apex), (b) MVP: standing after squatting (apex) [9]

Holosystolic murmurs:

Holosystolic murmurs in mitral regurgitation are plateau-shaped, loudest at the apex with radiation to the axilla, and may include a wide S2 split and S3 gallop. In tricuspid regurgitation, the murmur is loudest at the left sternal border, increases with inspiration (Carvallo's sign), and lacks axillary radiation. VSD murmurs are harsh, heard at the left sternal border, and show a fixed/wide S2 split without axillary radiation or inspiratory increase.

Mid systolic murmurs:

Midsystolic (ejection) murmurs are diamond-shaped, occurring between S1 and S2. They result from increased blood flow (e.g., pregnancy, anemia), dilated aorta, or aortic/pulmonic stenosis. Aortic stenosis murmurs are harsh, heard at the right sternal border/apex, and decrease with inspiration. Pulmonic stenosis murmurs are louder at the left sternal border and increase with inspiration.

Late systolic murmurs:

Late systolic murmurs occur in mitral/tricuspid valve prolapse, often with a mid-systolic click. They have a crescendo shape toward S2 and are transient, appearing occasionally in prolapse cases.

Diastolic murmurs:

Diastolic murmurs include filling murmurs (forward flow through stenotic AV valves) and regurgitant murmurs (backward flow through semilunar valves). Mid-diastolic/presystolic murmurs (e.g., Austin-Flint murmur) occur in mitral/tricuspid stenosis or aortic regurgitation, often with an S3 rumble.

Early diastolic murmurs:

Early diastolic murmurs start right after S2 and fade before S1, caused by regurgitant flow through aortic or pulmonic valves. They indicate aortic or pulmonic regurgitation.

Aortic insufficiency:

Aortic regurgitation causes a high-pitched, decrescendo early diastolic murmur, heard at the right/left sternal border and apex. It's accompanied by an S3, a midsystolic ejection murmur, and an Austin-Flint murmur (low-pitched rumble at the apex), indicating severe regurgitation.

Pulmonic regurgitation:

Pulmonary regurgitation murmurs increase with inspiration and lack the peripheral signs of aortic regurgitation (e.g., wide pulse pressure, bounding pulses). They are distinct from aortic regurgitation murmurs.

Mitral stenosis:

Mitral stenosis produces a low-pitched, rumbling diastolic murmur at the apex, starting after an opening snap and showing presystolic accentuation. An increased S1 is a key sign of mitral stenosis.

Tricuspid stenosis:

Tricuspid stenosis murmurs are loudest at the 4th left intercostal space and increase with inspiration, distinguishing them from mitral stenosis murmurs.

Mid–diastolic and presystolic murmurs:

Mid-diastolic and presystolic murmurs result from turbulent flow through AV valves, like in mitral stenosis. The murmur is low-pitched, rumbling, and best heard at the apex in left lateral decubitus, often starting after an opening snap.

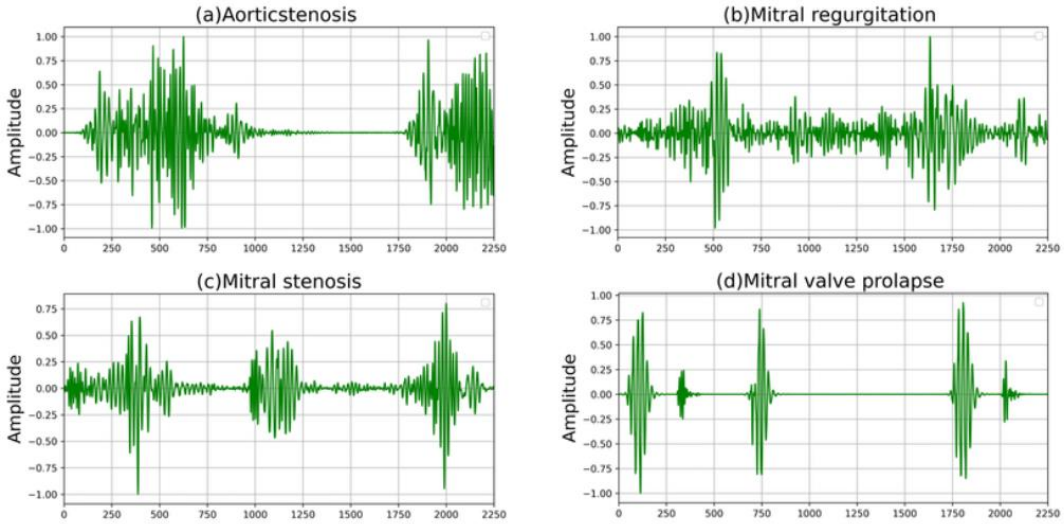


Figure I 14 Phonocardiogram (PCG) signals for various pathological cardiac conditions. [7]

Table 1 Summary of Heart Sounds (S1–S4) and Murmurs: Timing, Frequency Characteristics, and Anatomical Origins .

Sound / Murmur	**Audibility**	**Frequency (Hz)**	**Duration (ms)**	**Timing**	**Auscultation Site**	**Anatomical / Pathological Origin**
S1 (First Sound)	Audible	30 – 100	100 – 150	Onset of systole	Apex	Closure of **mitral (M1)** and **tricuspid (T1)** valves
S2 (Second Sound)	Audible	50 – 150	<60	End of systole (start of diastole)	Base (A2: R 2nd ICS, P2: L 2nd ICS)	Closure of **aortic (A2)** and **pulmonary (P2)** valves
S3 (Third Sound)	Low (inaudible in adults)	~30 – 50	~100	Early diastole	Apex / LLSB	Rapid ventricular filling (ventricular gallop); common in heart failure

Sound / Murmur	Audibility	Frequency (Hz)	Duration (ms)	Timing	Auscultation Site	Anatomical / Pathological Origin
S4 (Fourth Sound)	Low (absent in AFib)	~30 – 50	~80 – 100	Late diastole (pre-S1)	Apex / L 4th ICS	**Atrial contraction against stiff ventricle (LVH, ischemia)**
Systolic Murmur	Often loud	100 – 400+	Variable	Between S1 and S2	Apex (MR), LSB (TR), Base (AS/PS)	**Mitral/tricuspid regurgitation**, **aortic/pulmonary stenosis**, **VSD**
Diastolic Murmur	Soft / rumbling / high	50 – 200	Variable	After S2 (diastole)	Apex (MS), LSB (TS), L 3rd ICS	**Mitral/tricuspid stenosis**,
Continuous Murmur	Constant	100 – 400	Continuous	Systole + diastole	Left infraclavicular or upper back	**Patent ductus arteriosus (PDA)**, **arteriovenous fistulas**, **coarctation**

Phonocardiograms Signal used :

In this work the database we worked with has 65 normal signal and same number for the patholigicals siganls Although the data were originally housed in an open-access GitHub repository curated by researchers in the field (Yaseen Khan, 2021) [4] they are based on clinically acquired signals similar to those available on credible biomedical databases such PhysioNet, which offers validated datasets for cardiac auscultation research.

Every signal was manually examined and processed to represent structure and quality equivalent to datasets from PhysioNet/CinC Challenge 2016 or similar clinical sources, therefore guaranteeing compatibility and uniformity.

Table 2 PCG Signals Used and database sources

Name	Site	Last Visit	Signal used	Samples Frequency
Yassen2018	**https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features-**	**2023**	**65N (New_N_001 to New_N_065)**	**1024 Hz**
Pathological from Public Datasets	**http://www.egeneralmedical.com/ http://www.cardiosource.com/heartsounds/ http://www.dundee.ac.uk/medther/Cardiology/hsmur.html**	**2009**	**65 Pathological: 4 AR, 20 AS, 20 MR, 7 MS, 5 MVP, 4 TR, 2 PAS, 3 AI**	**8000–11000 Hz**

Conclusion :

In this chapter we covered essential ideas on the anatomy of the cardiovascular system, common connected diseases, phonocardiography principles, and the several heart sounds (S1 to S4). By giving a better knowledge of the degree of anomalies and their haemodynamic impact, the phonocardiogram (PCG) has proved to be a useful and non-invasive diagnostic tool for valvular heart disorders.

Beyond its therapeutic importance, the PCG signal has significant promise in developing disciplines including telemedicine and biometric identification. Nevertheless, the interpretation of PCG signals remains a difficult process requiring rigorous research and knowledge due to its complicated character and the existence of several noise sources.

Signal processing methods are therefore indispensable for extracting pertinent components (like S1, S2, murmurs, etc.), so improving signal quality, and for noise filtering. Before any effort at automation of classification or application of machine learning techniques, such preparation is absolutely vital.

This first chapter so lays the groundwork for the upcoming stage of our effort. Aiming to build a reliable and intelligent diagnosis support system, we will apply advanced signal processing techniques (such as FFT, wavelet transform, bispectrum analysis) in combination with machine learning algorithms in the following chapter to automatically classify heart sounds into normal and pathological categories.

Chapter II:

Classification of Heart Sounds Using Signal Processing Techniques

Chapter II: Classification of Heart Sounds Using Signal Processing Techniques

Introduction:

The phonocardiogram (PCG) signal represents cardiac activity, encapsulating temporal and spectral information. Nonetheless, its non-stationary characteristics and nuanced differences between normal and pathological instances pose considerable challenges for visual interpretation. Therefore, establishing a reliable binary classification system necessitates the utilization of advanced signal processing techniques to emphasize the diagnostic features of heart sounds.

This chapter initiates with standard time-frequency analysis methods, particularly the Fast Fourier Transform (FFT), which facilitates the extraction of global spectral parameters, including maximum amplitude and frequency range as discussed by Bereksi-Reguig and Debbal[11]. To overcome the limitations of stationary assumptions in biological signals, we utilize the Continuous Wavelet Transform (CWT), which provides localized time-frequency representations that are particularly effective for analyzing transient features in PCG signals [12]. We also address higher-order spectral techniques, particularly bispectral analysis, which enables the extraction of nonlinear features including phase entropy and the energy-weighted centroid (squared magnitude of the bispectrum). These parameters define the complexity and energy distribution of the signal, effectively differentiating between normal and pathological sounds. Shahbakhti et al., studied the efficiency of this technique for biomedical signals classification, particularly in heart sound analysis [13].

The dataset comprises 130 PCG signals, equally distributed between normal and pathological instances. After a pre-processing phase that encompasses resampling, normalization, and manual segmentation into S1, S2, and murmur phases, each signal will undergo analysis using the previously mentioned techniques.

The extracted features will then be used to train three classifier types: a conventional K-Nearest Neighbors (KNN) model, a Support Vector Machine (SVM), and a deep Convolutional Neural Network (CNN). Lastly, a comparative evaluation of the two methods will be performed to elucidate their performance, limitations,and interpretability.

This chapter delineates the complete analytical pipeline from preprocessing to classification, highlighting the significance of each technique in biomedical signal analysis.

Methodology:

Phonocardiogram Signals: Pre-processing :

This stage is crucial in the study of phonocardiogram (PCG) signals. Usually, pre-processing consists in resampling, normalizing, and filtering. Particularly sensitive to noise and interference, PCG signals generally show low amplitudes and limited bandwidth [11].

All cardiac sound signals in this work are recorded in WAV format and resampled at 8000 Hz. Since heart sounds typically lie within the 25 Hz to 750 Hz range, this sampling frequency is widely recognized as sufficient to capture the fundamental frequency components of PCG signals. This decision also fulfils the Nyquist-Shannon sampling theorem, which requires that the sample frequency have to be at least twice the maximum frequency of interest:

$$fe \geq 2 \times fmax \quad \text{(II.1)}$$

(Shannon,1949)

Normalizing is usually done before any signal processing is used. This stage preserves the dynamic range of the audio signal by varying its amplitude using a specified gain, therefore ensuring that the loudness is constant over all recordings. This stage's primary goal is to scale the signal in relation to its peak value, therefore reducing sample variability and preventing distortions brought about by amplitude differences.

Let X(t) be the raw PCG signal. The normalized signal Xnorm(t) is defined as:

$$Xnorm(t) = \frac{X(t)}{\max(|X(t)|)} \quad \text{(II.2)}$$

Segmentation:

Segmentation of the phonocardiogram (PCG) signal is the first and important step in the analysis process, playing a key role in automatically diagnosing heart sounds. Many segmentation methods

conventionally use the electrocardiogram (ECG) as a temporal reference for identifying key cardiac phases.

This study employs an alternative approach: segmentation is conducted exclusively on the PCG signal, independent of any ECG trace reference. This method improves the system's practicality and portability, particularly in low-cost or real-time diagnostic environments.

In accordance with the methodology established by Mouloud Guermoui.[15], the PCG signal is segmented into brief, clinically significant intervals through a four-step algorithmic process aimed at identifying S1, S2, and transitional phases including systole and diastole. Our method solely exploits the PCG signal itself, enabling a fully self-contained segmentation process without auxiliary sensors.

In order to accomplish this, we implemented an automatic segmentation strategy based on the approach established by Springer et al.[14], who created a logistic regression Hidden Semi-Markov Model (HSMM) that has been extensively utilized in the PhysioNet/CinC 2016 Challenge. This model employs existing knowledge of the timing and amplitude patterns of the cardiac cycle to segment the PCG signal into separate components:

S1: Closure of the mitral and tricuspid valves.

Systole: Ejection of blood from the ventricles.

S2: Closure of the aortic and pulmonary valves.

Diastole: Passive ventricular filling.

The segmentation annotations were provided in `.tsv` files containing time intervals and associated numerical labels:

Table II 1 Time and Order of the segments used in PCG signals

Label	Meaning
0	Silence
1	S1
2	Systole

Label	Meaning
3	S2
4	Diastole

$$A(n) = \sum_{m=0}^{L} s^2(n) \cdot win(n)$$

$$win(n) = \begin{cases} 1 - \frac{2|n|}{L} & \text{if } -L/2 \leq n \leq L/2 \\ 0 & \text{otherwise} \end{cases}$$

(II.3)

Each signal was automatically parsed in MATLAB to extract and store the time intervals corresponding to these phases.

Automatic Segmentation :

Our automatic segmentation process draws inspiration from the envelope-based strategy outlined in Zhang's foundational study (Analysis-Synthesis and Time-Frequency Scaling of the Phonocardiogram, 1996). This approach utilizes an energy envelope to identify the boundaries of heart sounds according to various criteria:

- It is essential for each segment to exceed 100 ms in duration in order to maintain the transient characteristics.
- There should be no decrease in energy during the segment.
- It is essential for the beginning and conclusion of each segment to uphold near-zero continuity in order to prevent the occurrence of reconstruction artifacts.
- It is essential that acoustic events (S1, S2, murmurs) are clearly distinguishable from background noise.

The selection of the energy distribution method for envelope calculation was made due to its enhanced capability in isolating pertinent components with greater precision. The energy envelope A(n) is calculated as follows:

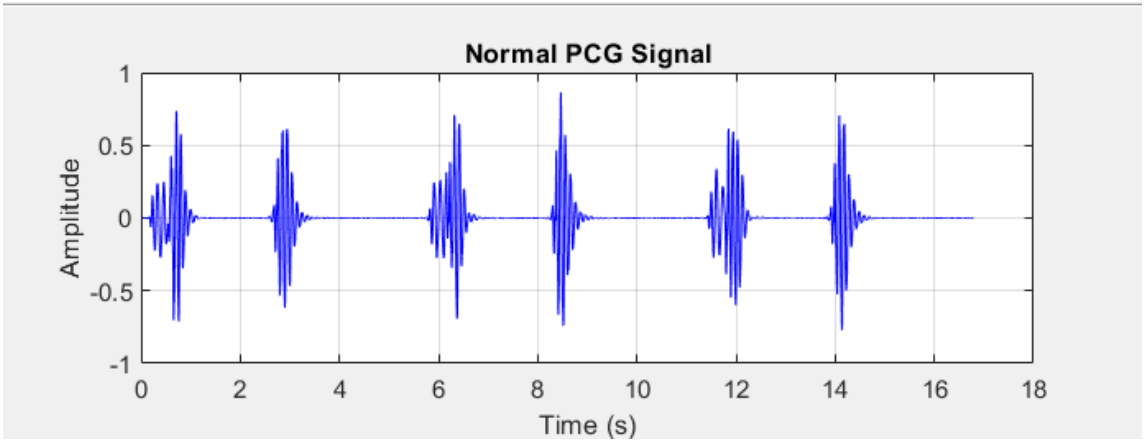


Figure II 1 Normal Original Signal

The segmentation algorithm relied on manually defined intervals derived from .tsv annotations, which were previously labelled following the method outlined by Springer et al [14] , using a logistic regression-hidden semi-Markov model (HSMM). The application of these labels facilitated the automatic extraction of temporal indices related to:

Label 1 corresponds to S1

Label 3 is associated with S2.

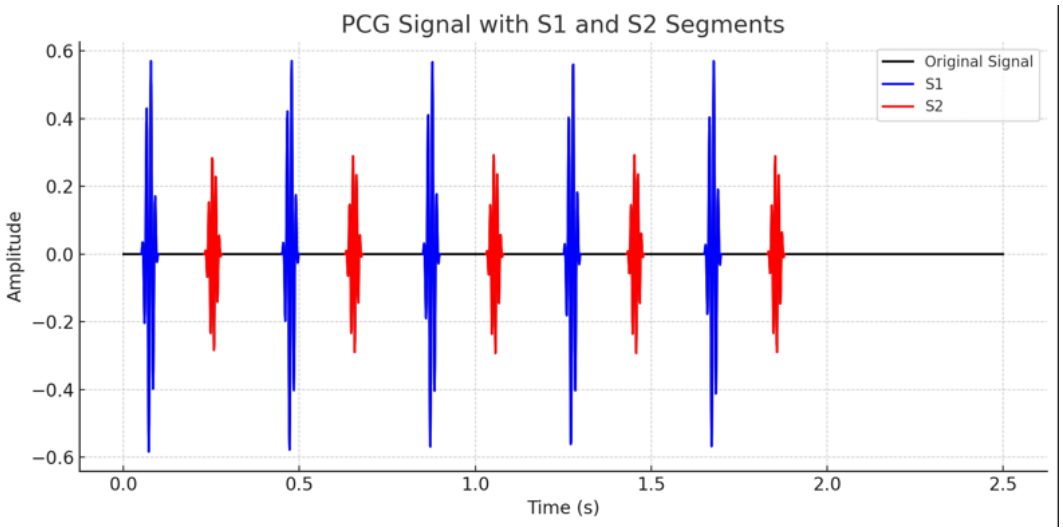


Figure II 2 Segmented PCG Signal (S1 & S2 Highlighted)

Implementation of Automatic Segmentation:

Figure II.3 demonstrates the automatic segmentation results on a phonocardiogram (PCG) signal through MATLAB. The objective of this process is to identify and categorize the key components of the heart:

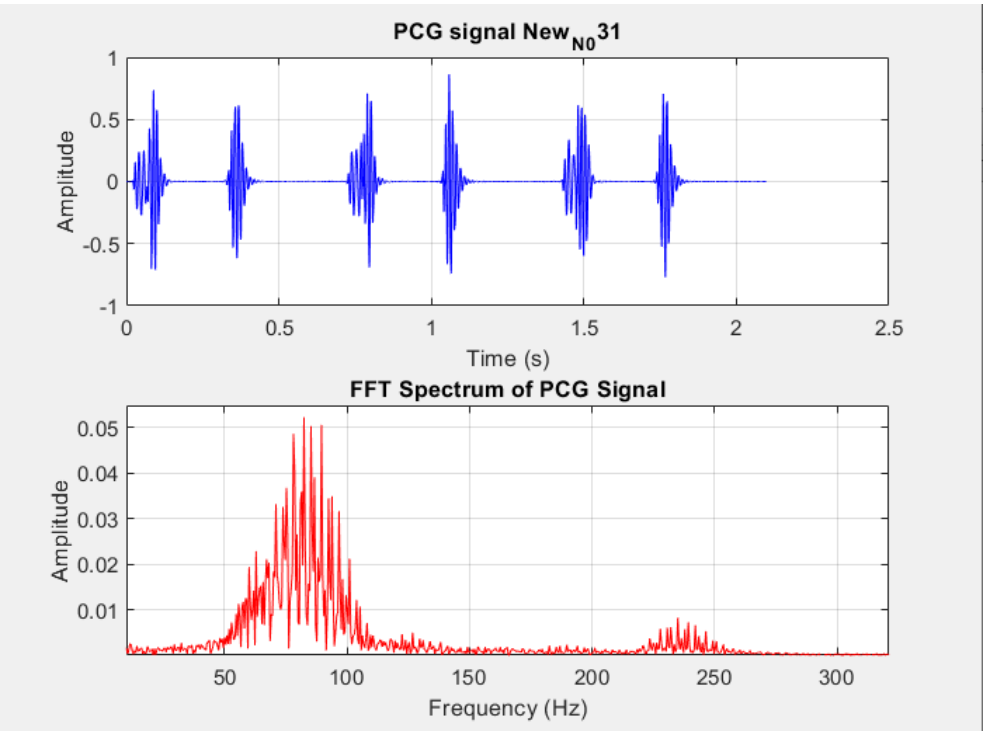


Figure II 3 FFT analysis of a normal split PCG signal: time-domain waveform (top) and corresponding frequency spectrum (bottom).

Theoretical notions about the employed techniques for features extraction :

Following the segmentation of the PCG signals into their essential components (S1, S2, and murmurs), we move forward with the extraction of significant features from each segment. These characteristics act as distinguishing factors that inform classification algorithms and assist in differentiating between normal and pathological heart sounds.

We employed various time-frequency and higher-order spectral methods to obtain both temporal and spectral information

Each of these methods is utilized independently for the segments associated with S1, S2, and murmurs (pertaining to pathological signals). The features obtained are assembled into a dataset that will be utilized subsequently for classification through machine learning techniques.

Fourier Analysis :

Fourier Series: Analysing the Breakdown of Periodic Signals, A Fourier series expresses a periodic signal as a combination of sinusoids, specifically sine and cosine terms, that are related by harmonic frequencies. Essentially, any continuous-time periodic function with a specific period can be represented as a linear combination of sinusoids whose frequencies are integer multiples of the fundamental frequency. The Fourier series expansion can be expressed mathematically as follows:

$$x(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)) \quad \text{(II.4)}$$

Where:

$\omega = \frac{2\pi}{T}$ is the **fundamental angular frequency**.

a_0, a_n $a_b$are the Fourier coefficients, computed over one period TTT of the signal.

This expansion suggests that even intricate periodic waveforms can be understood through their simpler sinusoidal components, shedding light on the frequency content of the signal. [16]

Fourier Transform for Aperiodic Signal Analysis :

An Extension for Aperiodic Signals Although the Fourier series is designed for periodic signals, it is important to note that many real-world signals, particularly in the biomedical field, are aperiodic or non-repeating. By conceptually extending the period T to infinity, the discrete sum of the Fourier series transforms into an integral, resulting in the Fourier Transform (FT) for non-periodic signals.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)\, e^{-j\omega t}\, dt \quad \text{(II.5)}$$

The inverse transform reconstructs the original signal:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega)\, e^{j\omega t}\, d\omega \quad \text{(II.6)}$$

In practice, PCG signals are digitised. Therefore, the Discrete Fourier Transform (DFT) is used:

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1 \tag{II.7}$$

The **Fast Fourier Transform (FFT)** is an optimized algorithm that computes the DFT in O (N log N) time (Cooley & Tukey, 1965), making spectral analysis efficient for long signals [17][18]

The efficiency of the FFT has made frequency-domain analysis widely accessible, allowing for real-time processing even with lengthy signals. The FFT efficiently provides the spectrum of a digital signal, serving as a fundamental element in contemporary spectral analysis methods.

Utilising FFT for the Analysis of PCG Signals Phonocardiogram (PCG) signals, which are recordings of heart sounds, are commonly analysed through Fast Fourier Transform (FFT) techniques. This method helps in extracting overall frequency-domain features that define the characteristics of heart sounds.

Utilising the FFT on a PCG segment produces its power spectrum, allowing for the extraction of essential features that can be used for diagnostic or classification objectives.

Two significant characteristics derived from the FFT of a heart sound are:

- The dominant frequency, which is the frequency at which the spectral amplitude reaches its peak.
- The frequency band, or bandwidth, that encompasses the significant signal energy.

In clinical discussions, the dominant frequency represents the most prominent tone in the heart sound, whereas the bandwidth signifies the spectrum of frequencies present, spanning from the lowest to the highest significant frequency content.

A significant limitation of the Fourier Transform, and consequently the FFT, lies in its assumption of signal stationarity, meaning that it presumes the frequency content of the signal remains constant over time. Heart sounds are naturally non-stationary signals, as their spectral characteristics change throughout the cardiac cycle. For instance, brief high-frequency components may emerge during a murmur or a valve click, but they do not persist throughout the entire cycle. A standard FFT applied to a complete PCG cycle provides insight into the overall frequency content; however, it does not indicate the specific moments in time when those frequencies manifest. This phenomenon is referred to as the time-localization problem. The Fourier Transform, while powerful, sacrifices time resolution as it integrates across the entire duration of the signal. As a result, significant transient or time-varying characteristics, like the timing of a murmur during the heartbeat, cannot be directly discerned from the standard FFT spectrum

Time–Frequency Analysis :

Many biomedical signals, including the phonocardiogram (PCG), are non-stationary, meaning their frequency content changes over time. Traditional Fourier analysis offers a global frequency spectrum; however, it falls short for such signals as it lacks the ability to maintain time localisation.

For instance, a heart sound signal encompasses transient events such as S1 and S2 sounds, along

with murmurs, which exhibit changing spectral characteristics throughout each cardiac cycle.

In these instances, it is essential to utilise a time-frequency representation to understand the variations in spectral energy over time.

This encourages the application of sophisticated time-frequency methods, particularly the wavelet transform, which can adjust its resolution according to the signal's content and more effectively highlight transient features in PCG recordings. Modern digital signal processing methods, such as wavelet analysis, have demonstrated their ability to extract both qualitative and quantitative information from PCG signals, offering insights that traditional auscultation methods may overlook.[19]

Wavelet Transform :

The Wavelet Transform (WT) serves as a valuable technique for analysing signals in both time and frequency domains, allowing for a detailed multiresolution decomposition. The wavelet transform employs wavelets that can be scaled (dilated or compressed) and translated,[12] serving as effective functions for analysis.

A wavelet is a brief oscillatory waveform that has a zero mean and is confined to specific time and frequency intervals. By stretching or compressing a prototype “mother” wavelet in time, the wavelet transform adjusts its ability to resolve time and frequency. Short, narrow wavelets are effective at capturing high-frequency components with precise temporal detail, while stretched, wide wavelets are better suited for capturing low-frequency components with improved frequency resolution. Essentially, the wavelet transform conducts a series of band-pass analyses, with each scale concentrating on a distinct frequency band. This approach results in a logarithmic tiling of the time-frequency plane, where higher frequency bands are examined using shorter windows, and lower frequency bands are analysed with longer windows. This produces a combined time-frequency representation, commonly depicted as a scalogram, which can uncover the fleeting and spectral characteristics of PCG signals within a single visual display.[12][23].

Mallat’s seminal 1989 work formalized the multiresolution framework of wavelets, and Daubechies (1988) developed orthonormal wavelet bases with compact support, enabling efficient

implementations of the discrete wavelet transform. Since then, wavelet techniques have been widely adopted in biomedical signal processing due to their flexibility and robustness in handling non-stationary data .[24]

Continuous Wavelet Transform (CWT) :

Mathematically, the **Continuous Wavelet Transform (CWT)** of a signal X(t) is defined as the inner product of X(t) with a scaled and shifted version of a mother wavelet Ψ(t). Specifically,

$$W_x(a,b) = \langle x(t), \Psi_{a,b}(t) \rangle = \int_{-\infty}^{+\infty} x(t)\, \Psi^*_{a,b}(t)\, dt, \quad \text{(II.8)}$$

Where :

$\Psi a,b(t)= \frac{1}{\sqrt{|a|}} \Psi(\frac{t-b}{a})$, $\begin{cases} a \text{ is the scale factor }, a > 0 \\ b \text{ is the time} \end{cases}$

For the continuous wavelet transform to be considered valid and invertible, the mother wavelet Ψ(t) must meet specific admissibility conditions. Primarily, it should be a function that has a mean of zero and finite energy.

where Ψ∗ is the complex conjugate of the mother wavelet Ψ. The coefficients Wx(a,b) represent the signal's localized energy around time b and scale a, forming a time-scale representation of the signal [12].

The relationship between scale a and frequency f is inverse:

$$f \approx \frac{f_c}{a} \quad \text{(II.9)}$$

where fcis the central frequency of the mother wavelet. Thus, a smaller scale corresponds to a higher frequency, and vice versa [12].

To ensure that the CWT is invertible and well-defined, the mother wavelet Ψ(t)\Psi(t)Ψ(t) must satisfy certain mathematical conditions, particularly the **admissibility condition [12]:**

- Zero mean:

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0 \quad \text{and equivalently} \quad \Psi(0) = 0$$

- Finite energy:

$$E = \int_{-\infty}^{+\infty} |\Psi(t)|^2 dt < \infty \tag{II.10}$$

- Finite admissibility constant:

$$C_\Psi = \int_0^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty$$

These conditions ensure no DC component (zero-frequency) and enable accurate reconstruction of the original signal from wavelet coefficients .

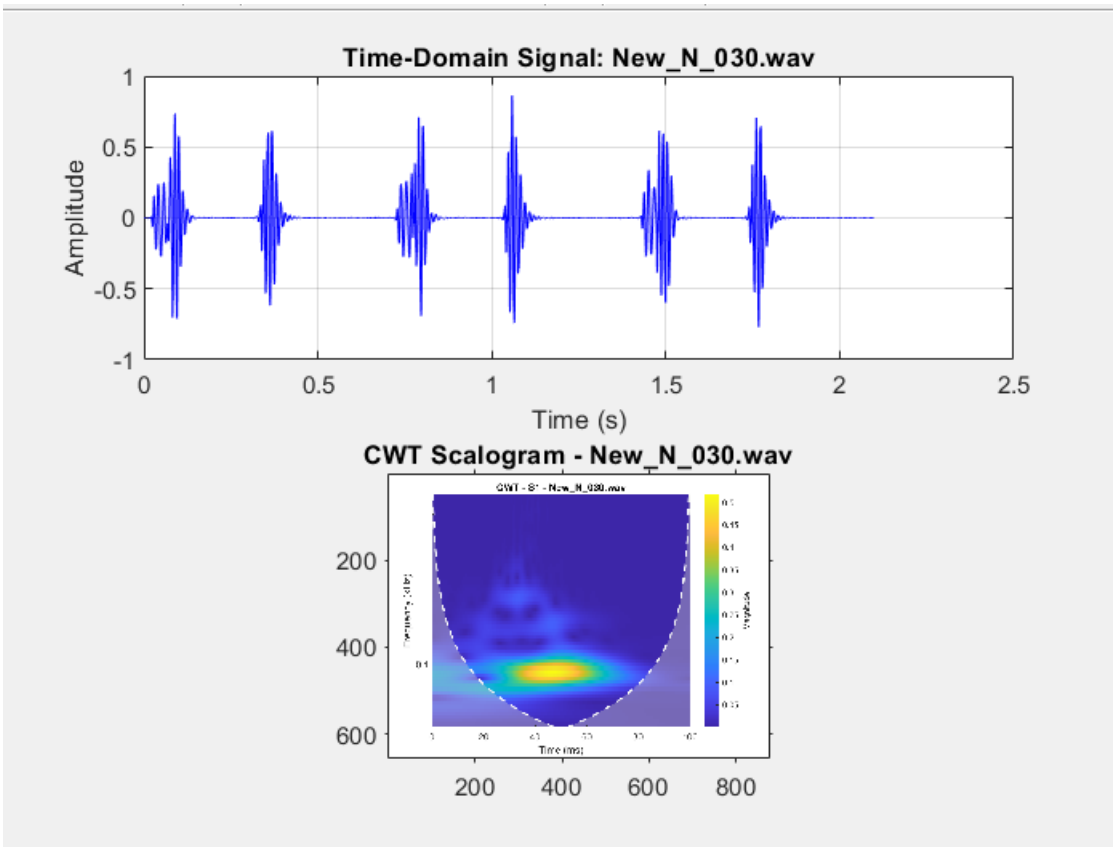


Figure II 4 Original Signal of Normal PCG and CWT

The Continuous Wavelet Transform has a notable limitation in that it exhibits computational redundancy, given that it calculates a continuous array of coefficients over both time and scale. The presence of this redundancy may lead to elevated computational expenses and a greater need for data

storage resources [2]. Moreover, the interpretation of wavelet scales in relation to precise frequencies, often referred to as pseudo-frequencies, can occasionally be counterintuitive and is heavily influenced by the selection of the mother wavelet [2][15].

Though it has certain advantages, bispectral analysis has certain clear limits: Comparatively to second-order techniques like FFT, computational complexity requires more resources. As a 2D complex function, the bispectrum can be more difficult to understand without careful feature engineering.

Accuracy in estimation of noise depends on enough averaging and strong pre-processing. Still, it is among the most exciting instruments available for nonlinear signal analysis, particularly in cases when more traditional methods fall short in capturing minor deviations.

Bi-spectrum Analysis :

The bispectrum is a higher-order spectral analysis technique that extends traditional Fourier methods by capturing phase relationships between frequency components.

It is defined as the Fourier transform of the third-order cumulate of a signal :

$$B(f_1, f_2) = E\left[X(f_1)X(f_2)X^*(f_1 + f_2)\right] \tag{II.11}$$

where:

- X(f) is the Fourier transform of x(t)$x(t)$,
- X∗(f) denotes the complex conjugate,
- E[·]represents the expectation operator.

This technique is very important for the detection of nonlinear interactions in non-Gaussian signals such as diseased heart sounds [15], the bispectrum estimates quadratic phase coupling (QPC) between frequencies f1, f2, and f1+f2 .

The bispectral analysis offers valuable insights into the complexities of nonlinear and phase-coupled phenomena in PCGs, allowing for effective differentiation of murmurs and valvular pathologies. Although it requires significant computational resources, its ability to withstand noise and sensitivity to phase variations render it a valuable complement to FFT and wavelet techniques.

Symmetry and Periodicity: The bispectrum exhibits unique symmetry properties, as its correlation function is even, resulting in spectral symmetry in the Fourier transform. Additionally, since

the bispectrum is derived from the tricorrelation function, it demonstrates a periodic repetition four times beyond the Nyquist frequency [17]. Therefore, we consider only a non-redundant triangular region (Ω)

The analysis of (Ω) is necessary. This area is characterised by:

$$f_2 \geq 0,\ f_1 \geq f_2,\ \text{and}\ f_1 + f_2 \leq 0.5 \qquad \text{(II.12)}$$

The bispectrum satisfies the following symmetries:

$$S3(f1,f2)=S3*(-f2,f1)=S3*(-f1,-f2)=S3(-f1-f2,f2)=S3(f1,-f1-f2)=S3(-f1-f2,f1)=S3(f2,-f1-f2).S_3(f_1,f_2)=S_{3*}(-f_2,f_1)=S_{3*}(-f_1,-f_2)=S_3(-f_1-f_2,f_2)=S_3(f_1,-f_1-f_2)=S_3(-f_1-f_2,f_1)=S_3(f_2,-f_1-f_2). \qquad \text{(II.13)}$$

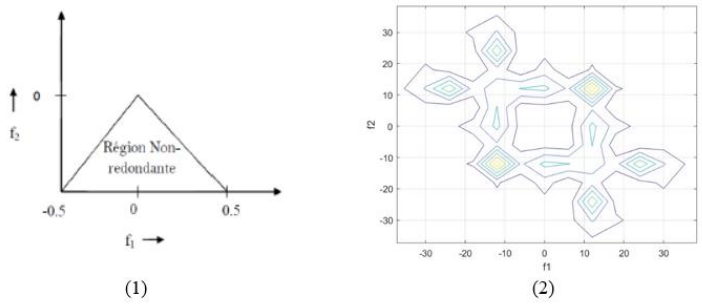


Figure II 5 The non-redundant region in a bispectrum analysis [15] (1) An 2D image of spectrum analysis for a S1 of normal PCG Signal (2).

Mel-Frequency Cepstral Coefficients (MFCCs) :

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used features in audio signal processing that provide a compact representation of the short-term power spectrum of a signal. The MFCC pipeline models human auditory perception by warping the frequency axis to the mel scale, which reflects how humans perceive pitch [20].

The typical MFCC extraction process consists of several stages [21]:

- **Pre-emphasis**, using a high-pass filter , to amplify high-frequency components.
- **Framing and windowing**, where the signal is divided into short overlapping frames (20–40 ms) and multiplied by a Hamming window to minimize edge effects.

- **Short-Time Fourier Transform (STFT)**, applied to each windowed frame x[n] produces a frequency spectrum:

$$X(k) = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi kn/N}, \quad k = 0, 1, ..., N-1$$

(II.14)

Mel-filterbank processing, where the magnitude spectrum is passed through a series of triangular filters spaced on the mel scale. The mel scale approximates human pitch perception and is computed as:

$$m(f) = 2595 \cdot \log_{10}(1 + \frac{f}{700})$$

(II.15)

Each filter sums the spectral energy in a specific frequency band:

$$E_i = \sum_{k=0}^{N-1} |X(k)|^2 \cdot \phi_i(k), \quad i = 1, ..., M$$

(2.16)

Logarithmic compression, which mimics the nonlinear response of human loudness perception:

$$\mathrm{Si} = \ln(\mathrm{Ei}) \qquad \text{(II.17)}$$

Discrete Cosine Transform (DCT), which decorrelates the log filterbank energies and compresses them into a small set of coefficients:

$$c_n = \sum_{i=1}^{M} \ln(E_i) \cdot \cos\left[n\pi \frac{i - 0.5}{M}\right], \quad n = 0, ..., N_c - 1$$

(II.18)

Typically, only the first 12–13 cepstral coefficients are retained, as they effectively describe the spectral envelope, while discarding fine spectral variations. These coefficients are often augmented with first and second-order derivatives (delta and delta-delta) to capture temporal dynamics [22].

The key advantage of MFCCs lies in their alignment with the human auditory system. The mel scale emphasizes low-frequency bands—where human hearing is more sensitive—and compresses high-frequency regions [23]. This perceptual modeling improves robustness in tasks like speech and speaker recognition compared to using raw spectrograms [24].

MFCCs have become standard features in a wide range of applications, including speech recognition, speaker identification, and audio classification, due to their ability to preserve phonetic information in a compact and discriminative form [25].

Features Extraction :

FFT Application for Spectral Feature Extraction :

After obtaining the heart sound segments, the Fast Fourier Transform (FFT) is utilised to transform each segment from the time domain into the frequency domain. In our implementation, we utilised standard libraries, specifically MATLAB's fft, to process each segment of the waveform.

The results of the FFT provides a collection of frequency components that illustrate the distribution of the signal's energy across various frequencies. The frequency resolution is influenced by both the segment length and the sampling rate. For instance, with a sampling rate of 1000 Hz and 1024 points, the spectrum spans from 0 to 500 Hz, achieving a resolution of approximately 0.98 Hz.

We analysed the amplitude spectrum of each heart sound to gather quantitative features that could define the segment. Namely, the maximum spectral amplitude and the dominant frequency.

The maximum spectral amplitude (MaxAmp)

It indicates the peak magnitude of the FFT spectrum, highlighting the most prominent frequency component found in the sound.

The dominant frequency (F_Dom)

It refers to the frequency where the highest spectral amplitude is observed, essentially representing the frequency component that holds the most energy within the segment.

These two features effectively highlight the most prominent frequency component of the heart sound and its significance.

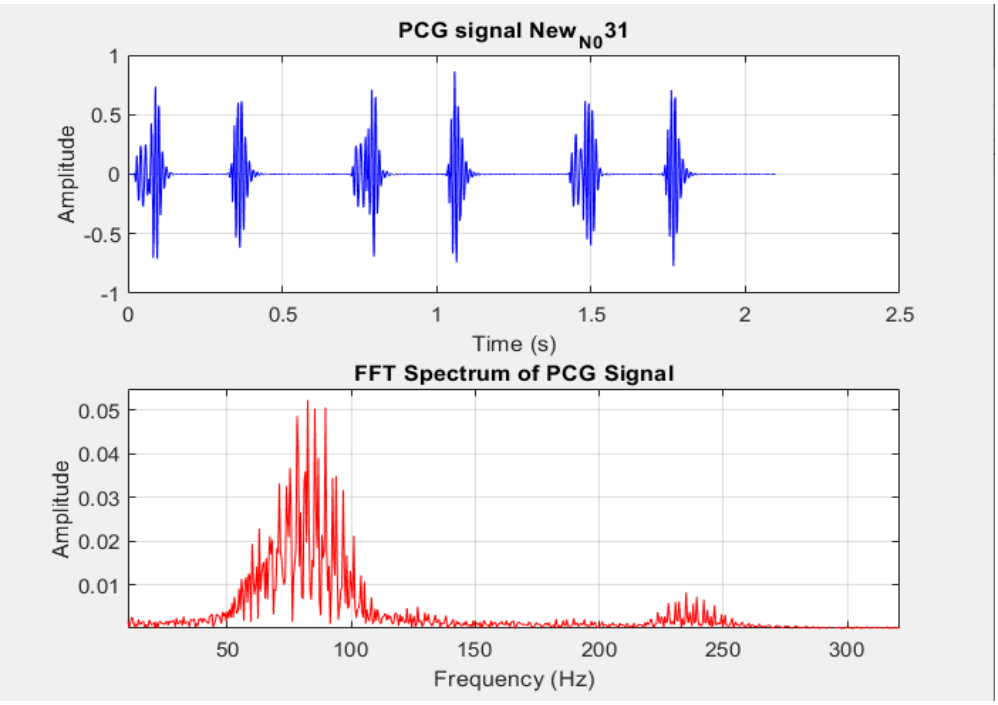


Figure II 6 Normal PCG Signal with the Spectre of the FFT (Max Amp = 0.14 f-Dom~75 Hz)

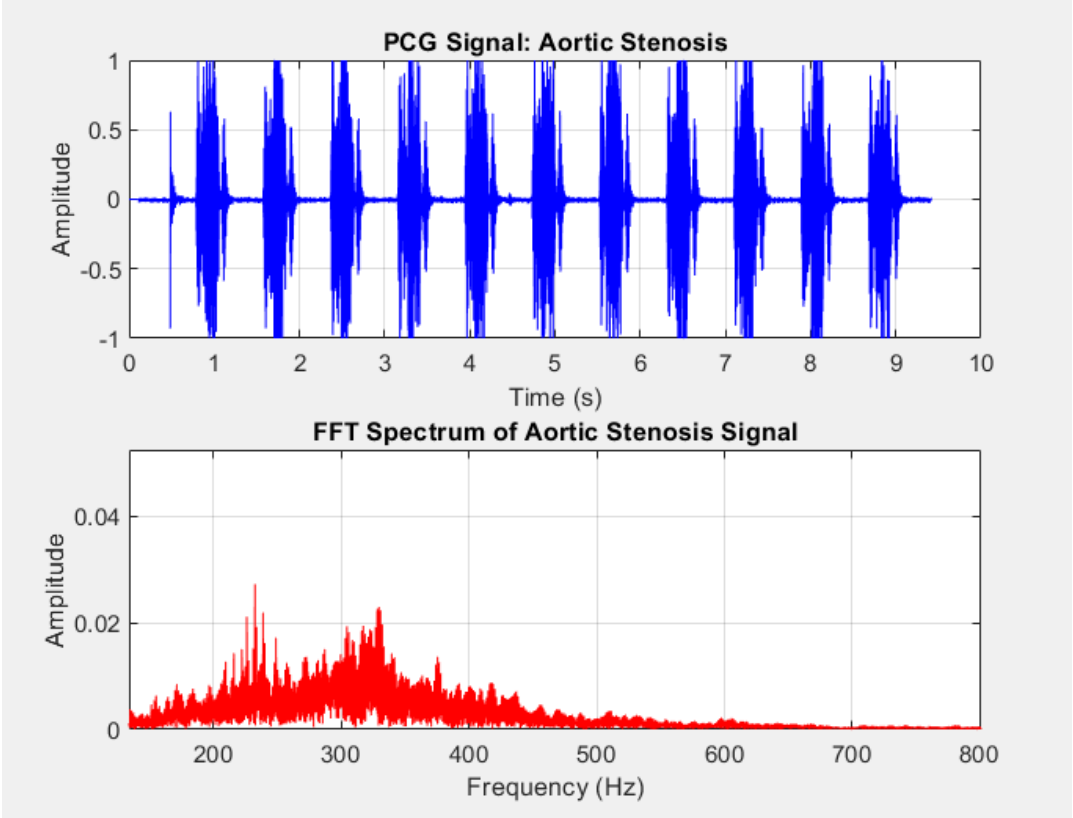


Figure II 7 Patholigical PCG signal with FFT spectrum to extract Max Amp and F Dom

While carrying out the FFT analysis mentioned above, we made sure to compute and store the features of each segmented sound in an organised manner. For every PCG recording, we gather a collection of features, such as MaxAmp_S1, F_Dom_S1, MaxAmp_S2, F_Dom_S2, MaxAmp_Murmur and F_Dom_Murmur. A feature matrix or table stores these values, with each row representing a single heartbeat recording.

CWT Application for Time-Frequency Features Extraction:

The Continuous Wavelet Transform (CWT) is utilised to examine heart sound (PCG) signals within the time-frequency domain. In contrast to the FFT, which offers a global frequency spectrum while overlooking temporal details, the CWT generates a scalogram that illustrates the evolution of signal energy across both time and frequency.The Continuous Wavelet Transform (CWT) utilises wavelets that expand and compress over time, enabling a multi-resolution analysis. This approach provides enhanced time resolution for high frequencies while ensuring improved frequency resolution for low frequencies. This indicates that the CWT is capable of effectively capturing both the short, high-frequency elements of valve closure sounds and the longer, low-frequency aspects of murmurs within a cohesive framework.[21] .

We extract descriptive information from each segmented cardiac sound event (S1, S2, and murmur segments) using the CWT. We use MATLAB's cwtfilterbank to calculate continuous wavelet transform coefficients from 25 Hz to the signal's maximum frequency. Focussing on 25 Hz and above reduces low-frequency noise and baseline drift, directing analysis to the relevant heart sound bandwidth [12]. In our recordings, the upper limit is established at the Nyquist frequency (~1 kHz), which contains the highest significant frequencies in murmurs. The segment's scalogram (time–frequency energy distribution) is the magnitude-squared of the CWT's time–frequency coefficient matrix. This scalogram depicts when and at what frequencies the heart sound's energy is focused, describing each component in detail.

We calculate quantitative features from each segment's scalogram to characterise its time-frequency content. We extract two characteristics that capture the segment's energy intensity and spectral distribution:

Time-frequency mean _Energy

The CWT scalogram's average energy, calculated as the squared wavelet coefficients' time and frequency mean. It shows the segment's signal power after filtering to the chosen band. S1 and S2 segments have more energy than quieter murmur segments if normalised by duration.

$$E_{\text{mean}} = \frac{1}{N_a \cdot N_b} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} |W(a_i, b_j)|^2 \tag{II.19}$$

Spectral Entropy:

Shannon entropy of scalogram energy distribution across frequencies. The scalogram's energy is first summed throughout time at each frequency to get a frequency energy marginal, which is normalised to a probability distribution. - This feature reflects the frequency complexity or dispersion in the segment [23] pmc.ncbi.nlm.nih.gov. For clear, low-frequency S1/S2 sounds, the energy is concentrated in a small frequency band, while murmurs and noise have a wider, more uniform energy distribution. Research suggests wavelet-based entropy estimates can distinguish between normal and diseased heart sounds, as murmurs typically have higher-frequency components and a less concentrated spectrum .

$$\text{Hnorm} = \text{H}/\log 2(\text{N}) \qquad \text{(II.20)}$$

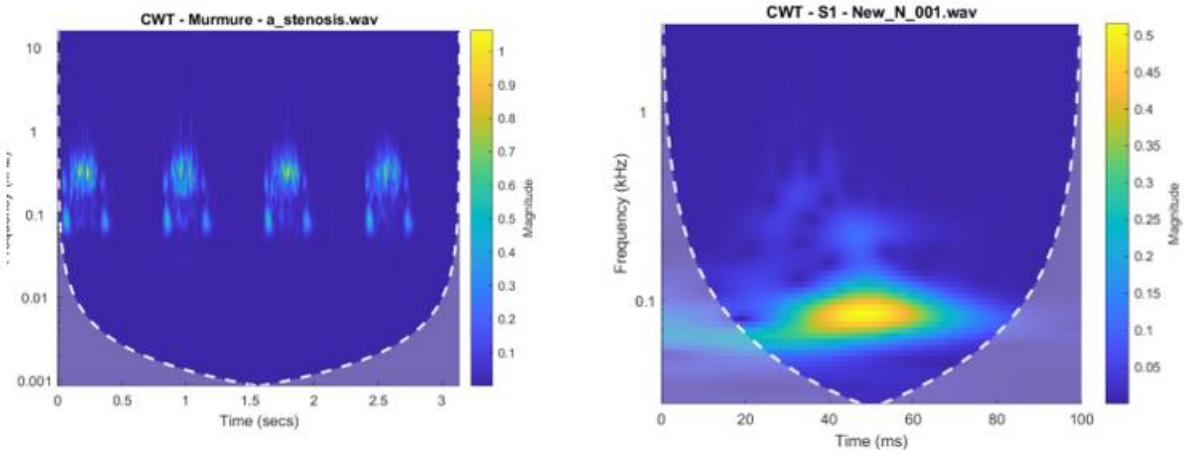


Figure II 8 Illustrate the scalogram of two different cardiac cases: a) Normal S1 segment , (b) Aortic stenosis murmur segment.wav

One can notice from this figure that for a normal S1 segment, the energy is localized between ~30–60 ms and concentrated around **100 Hz,** which reflects clean, short valve closure (S1) with low entropy and high energy. **Thus, for the murmur segment, we observe** repetitive energy bands post-S1 across ~0.2–0.4 s with frequency spreading up to **300–400 Hz**, which is characteristic of systolic murmur. Broad spectral distribution leads to higher entropy, multiple bursts of energy.

Bispectral Analysis for High-Order Feature Extraction :

as previously stated Higher-order spectral techniques like bispectrum analysis capture nonlinear signal properties. By retaining phase interactions, the bispectrum (a third-order statistic) can disclose quadratic phase coupling (QPC) between frequency components, unlike the typical power spectrum (second-order). This makes it useful for analysing small nonlinearities in phonocardiogram (PCG) signals. We isolate the first heart sound (S1), second heart sound (S2), and any murmur segments from

each PCG cycle and compute their bispectrum using MATLAB's Higher-Order Spectral Analysis (HOSA) toolbox. The HOSA toolset has strong signal data bispectrum estimation techniques.

Two key features were extracted from the obtained bispectrum:

Bispectral Phase Entropy (EntPh):

This feature measures the randomness or disorder in the phase angles of the bispectrum. We first treat the normalized bispectral phase distribution as a probability distribution and compute its entropy (in practice, dividing the bispectrum into bins of phase values and applying the Shannon entropy formula). **Phase Entropy is low when the bispectrum's phase angles are consistently aligned (indicating strong phase coupling), and high when the phase is uniformly random**. In other words, a highly structured, phase-coupled signal (with nonlinear interactions) yields lower entropy, whereas a noise-like or uncoupled signal yields near-maximal entropy.

$$\mathrm{EntPh} = -\sum_{k=1}^{K} p_k \log_2(p_k)$$

$$\mathrm{EntPh}_{\mathrm{norm}} = \frac{\mathrm{EntPh}}{\log_2(K)} \qquad \text{(II.21)}$$

In our context, we expect normal heart sounds (S1, S2) to exhibit lower phase entropy (more phase-coupling structure), while murmurs (which are often turbulent and noise-like) produce higher phase entropy due to their random phase structure.

Bispectral Weighted Centroid (WC):

This feature captures the "center of mass" of the bispectrum's magnitude distribution in the frequency domain. It is analogous to the spectral centroid used in audio analysis but here computed over the two-dimensional bispectral plane. We calculate the weighted centroid by averaging the frequency values in the bispectrum, weighted by their magnitude. In practice, this can be implemented by summing f1+f2f_1+f_2f1+f2 (or another representative frequency metric) over all bispectral components, each weighted by |B(f_1,f_2)|, and normalizing by the total bispectral energy.

$$\mathrm{WC} = \frac{\sum_{i,j}(f_{1i} + f_{2j}) \cdot |B_{i,j}|}{\sum_{i,j} |B_{i,j}|} \qquad \text{(II.22)}$$

- f1i and f2j are the discrete frequency indices
- |Bi,j| is the bispectral magnitude at each frequency pair

The WC indicates where the bispectral energy is concentrated: a lower WC means the energy of nonlinear interactions is concentrated at lower frequencies, while a higher WC suggests significant coupling energy at higher frequency combinations. Thus, WC provides a summary of the **energy distribution in the bispectrum** – for example, an S1 segment (which is typically a low-frequency thump) might have a lower bispectral centroid than a high-pitched murmur that spreads energy across a broader band.

Mel-Transform for Audio/ Perception Feature Extraction :

MFCCs are popular audio features created for speech recognition, designed to capture the way we perceive sound's spectral qualities. An MFCC shows the short-term spectral shape of a signal on a mel scale, which reflects how humans hear sounds. MFCCs give a concise description that highlights frequency bands similar to how our ears perceive them, capturing the sound's timbral qualities. These coefficients don't really relate to each other much, but they are great at distinguishing different audio types.

MFCC features are great for analysing speech and music because they effectively summarise important spectral information. They've also been used to analyse heart sounds (PCG) to distinguish between normal and abnormal beats. Using a perceptual frequency scale (mel) helps MFCCs emphasise frequency content like how doctors listen for specific murmurs or heart sound issues based on their pitch and tone. Putting it into action: In this study, we use MFCC feature extraction on each part of the heart sound, including the S1, S2, and murmur segments. We break down the PCG into its basic sounds and look at each part separately. This way, we can focus on the unique spectral details of S1, S2, and murmurs without any overlap from other phases, similar to what previous studies have done.

We use a standard MFCC computation, like the one in MATLAB's MFCC toolbox, to turn each segment's waveform into a series of cepstral coefficient vectors. This process includes breaking the segment into short windows, calculating the Fourier spectrum for each frame, mapping the spectral magnitudes to a mel-frequency filter bank, taking the logarithm of each band's energy, and then applying a discrete cosine transform (DCT) to get the cepstral coefficients. Usually, we get 12 to 13 MFCCs for each frame, which capture the segment's spectral shape in a simpler way. This analysis results in an MFCC coefficient matrix for a heart sound segment, showing how the segment's frequency content changes over time, with time frames on one axis and cepstral coefficient indices on the other. Global Cepstral Features Instead of using all the frame-by-frame MFCC values, we simplify the cepstral matrix for each segment into a few key descriptors that capture the segment's overall cepstral features.

Specifically, we get three features from the MFCC matrix of each S1, S2, or murmur segment: **cepstral energy entropy, cepstral phase entropy, and MFCC magnitude centroid.**

First, the **cepstral energy entropy** measures how evenly the spectral energy is distributed across the MFCC coefficients within a segment. It is computed from the normalized energy of each coefficient Ek , where

$$H = -\sum_{k=1}^{K} p_k \log_2(p_k) \tag{II.21}$$

Higher phase entropy indicates turbulent or noisy segments, as found in pathological murmurs, while lower values suggest periodic and stable patterns characteristic of normal heart sounds [30].

Lastly, the **MFCC magnitude centroid** quantifies the spectral sharpness or brightness of the segment. It computes the center-of-mass of the MFCC magnitude spectrum as:

$$WC = \frac{\sum_{k=1}^{K} k \cdot |M_k|}{\sum_{k=1}^{K} |M_k|} \tag{II.22}$$

where **the numerator** is the magnitude of the **the denominator** is the MFCC. A higher centroid implies greater emphasis on higher-frequency components (e.g., sharp, high-pitched murmurs), whereas a lower value suggests a duller, low-frequency profile often associated with normal heart tones [29].

Together, these features provide a compact yet expressive summary of each heart sound segment's spectral structure, supporting effective classification between normal and pathological conditions.

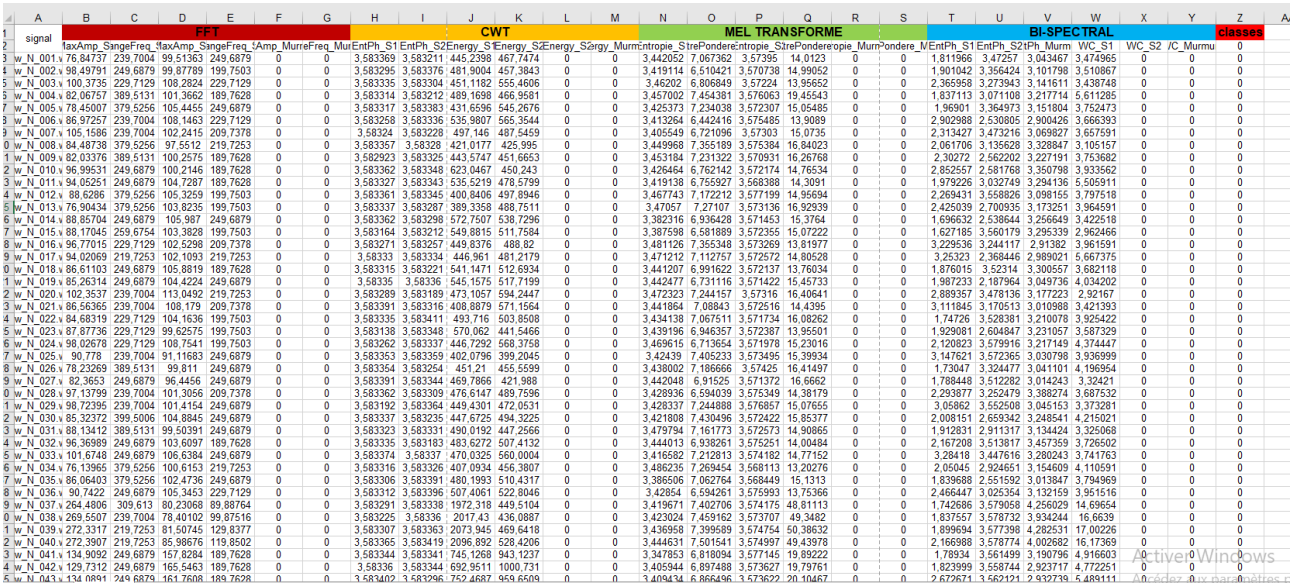
Extracted Features Resume :

Table II 2 Features Extraction with all the techniques (FFT , CWT , Bispectrum , MFCC's)

Technique	Feature	Definition	Interpretation in PCG Analysis
MFCC (Segment-based)	**Cepstral Energy Entropy**	Measures how evenly the spectral energy is distributed across MFCCs in a heart sound segment.	High → broadband/complex spectrum (e.g., murmur). Low → concentrated, tonal spectrum (e.g., clean S1/S2).

Technique	Feature	Definition	Interpretation in PCG Analysis
	Cepstral Phase Entropy	Quantifies irregularity in phase content from the complex cepstrum of the segment.	High → irregular or noisy structure (e.g., pathological murmur). Low → periodic, stable beat.
Bispectral Analysis	**Phase Entropy (EntPh)**	Measures randomness of phase coupling in the bispectrum.	Low → structured signal (e.g., normal beat). High → random coupling (e.g., noisy murmur).
	Weighted Centroid (WC)	Measures the average frequency location of significant bispectral components.	Low → energy concentrated at low frequencies (e.g., S1). High → broad coupling (e.g., high-frequency murmur).
CWT (Scalogram)	**Mean Energy**	Reflects the average power distribution across time-frequency regions in the signal.	High → strong heart sound intensity (often in murmur regions). Low → lower energy segments.
	Spectral Entropy	Indicates how spread or concentrated the energy is across frequencies in the CWT.	High → complex frequency content (e.g., murmurs). Low → clear, focused frequency bands (e.g., S1/S2).
FFT (Spectrum)	**Maximum Amplitude**	Captures the peak value in the frequency domain of the heart sound.	High → dominant tonal peak (may indicate strong S1/S2). Low → weak or diffuse spectral content.
	Frequency Range / Bandwidth	Describes the span of frequency content in the signal spectrum.	Wide range → turbulent or noisy signal (e.g., murmur). Narrow range → pure tones (e.g., S1/S2).

In our project, we built a matrix where each column corresponds to a specific feature, and there's an extra column that contains the class label (0 for normal, and 1 for pathological). For instance, the entry corresponding to a typical heart sound recording will display specific values for the S1 and S2 features, while the murmur features will be zero (indicating the absence of a murmur), with the class labelled as 0. On the other hand, in a pathological case where a murmur is present, we can expect to see elevated values in the murmur feature columns, indicating a classification of 1. Organising features in a clear format, like a spreadsheet or CSV file, makes it easier to analyse and train classifiers later on the next Figure II 9:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	signal	FFT						CWT						MEL TRANSFORME						BI-SPECTRAL						classes
2		faxAmp_S	angeFreq_S	faxAmp_S	angeFreq_S	Amp_Murre	Freq_Murr	EntPh_S1	EntPh_S2	Energy_S1	Energy_S2	Energy_S2	argy_Murm	ntropie_S1	rePondere	ntropie_S2	rePondere	opie_Murm	ondere_M	EntPh_S1	EntPh_S2	tPh_Murm	WC_S1	WC_S2	/C_Murmu	0
3	w_N_001.w	76,84737	239,7004	99,51363	249,6879	0	0	3,583369	3,583211	445,2398	467,7474	0	0	3,442052	7,067362	3,57395	14,0123	0	0	1,811966	3,47257	3,043467	3,474965	0	0	0
4	w_N_002.w	98,49791	249,6879	99,87789	199,7503	0	0	3,583295	3,583376	481,9004	457,3843	0	0	3,419114	6,510421	3,570738	14,99052	0	0	1,901042	3,356424	3,101798	3,510867	0	0	0
5	w_N_003.w	100,3735	229,7129	108,2824	229,7129	0	0	3,583335	3,583304	451,1182	555,4606	0	0	3,46202	6,806849	3,57224	13,95652	0	0	2,365958	3,273943	3,141611	3,438748	0	0	0
6	w_N_004.w	82,06757	389,5131	101,3662	189,7628	0	0	3,583314	3,583212	489,1698	466,9581	0	0	3,457002	7,454381	3,576063	19,45543	0	0	1,837113	3,071108	3,217714	5,611285	0	0	0
7	w_N_005.w	78,45007	379,5256	105,4455	249,6879	0	0	3,583317	3,583363	431,6596	545,2676	0	0	3,425373	7,234038	3,572307	15,05485	0	0	1,96901	3,364973	3,151804	3,752473	0	0	0
8	w_N_006.w	86,97257	239,7004	108,1463	229,7129	0	0	3,583258	3,583336	535,9807	565,3544	0	0	3,413264	6,442416	3,575485	13,9089	0	0	2,902988	2,530805	2,900426	3,666393	0	0	0
9	w_N_007.w	105,1586	239,7004	102,2415	209,7378	0	0	3,58324	3,583228	497,146	487,5459	0	0	3,405549	6,721096	3,57303	15,0735	0	0	2,313427	3,473216	3,069827	3,657591	0	0	0
10	w_N_008.w	84,48738	379,5256	97,5512	219,7253	0	0	3,583357	3,58328	421,0177	425,995	0	0	3,449968	7,365189	3,575384	16,84023	0	0	2,061706	3,135628	3,328847	3,105157	0	0	0
11	w_N_009.w	82,03376	389,5131	100,2575	189,7628	0	0	3,582923	3,583325	443,5747	451,6653	0	0	3,453184	7,231322	3,570931	16,26768	0	0	2,30272	2,562202	3,227191	3,753682	0	0	0
12	w_N_010.w	96,99531	249,6879	100,2146	189,7628	0	0	3,583362	3,583348	623,0467	450,243	0	0	3,426464	6,762142	3,572174	14,76534	0	0	2,852557	2,581768	3,350798	3,933562	0	0	0
13	w_N_011.w	94,05251	249,6879	104,7287	189,7628	0	0	3,583327	3,583343	535,5219	478,5799	0	0	3,419138	6,755927	3,568388	14,3091	0	0	1,979226	3,032749	3,294136	5,505911	0	0	0
14	w_N_012.w	88,6286	379,5256	105,3259	199,7503	0	0	3,583361	3,583345	400,8406	497,8946	0	0	3,467743	7,172212	3,577199	14,95694	0	0	2,269431	3,558826	3,098155	3,797518	0	0	0
15	w_N_013.w	76,90434	379,5256	103,8235	199,7503	0	0	3,583337	3,583287	389,3358	488,7511	0	0	3,47057	7,27107	3,573136	16,92939	0	0	2,425039	2,700935	3,173251	3,964591	0	0	0
16	w_N_014.w	88,85704	249,6879	105,987	249,6879	0	0	3,583362	3,583298	572,7607	538,7296	0	0	3,382316	6,936428	3,571453	15,3764	0	0	1,696632	2,538644	3,256649	3,422618	0	0	0
17	w_N_015.w	88,17045	259,6754	103,3828	199,7503	0	0	3,583164	3,583212	549,8815	511,7584	0	0	3,387598	6,581889	3,572355	15,07222	0	0	1,627185	3,560179	3,295339	2,962466	0	0	0
18	w_N_016.w	96,77015	229,7129	102,5298	209,7378	0	0	3,583271	3,583257	449,8376	488,82	0	0	3,481126	7,355348	3,573269	13,81977	0	0	3,229536	3,244117	2,91382	3,961591	0	0	0
19	w_N_017.w	94,02069	219,7253	102,1093	219,7253	0	0	3,58333	3,583334	446,961	481,2179	0	0	3,471212	7,112757	3,572572	14,80528	0	0	3,25323	2,368446	2,989021	5,667375	0	0	0
20	w_N_018.w	86,61103	249,6879	105,8819	189,7628	0	0	3,583315	3,583221	541,1471	512,6934	0	0	3,441207	6,991622	3,572137	13,76034	0	0	1,876015	3,52314	3,300557	3,682118	0	0	0
21	w_N_019.w	85,26314	249,6879	104,4224	249,6879	0	0	3,58335	3,58336	545,1575	517,7199	0	0	3,442477	6,731116	3,571422	15,45733	0	0	1,987233	2,187964	3,049736	4,034202	0	0	0
22	w_N_020.w	102,3537	239,7004	113,0492	219,7253	0	0	3,583289	3,583189	473,1057	594,2447	0	0	3,472323	7,244157	3,57316	16,40641	0	0	2,889357	3,478136	3,177223	2,92167	0	0	0
23	w_N_021.w	86,56365	239,7004	108,179	209,7378	0	0	3,583391	3,583316	408,8879	571,1564	0	0	3,441864	7,08843	3,572516	14,4395	0	0	3,111845	3,170513	3,010988	3,421393	0	0	0
24	w_N_022.w	84,68319	229,7129	104,1636	199,7503	0	0	3,583335	3,583411	493,716	503,8508	0	0	3,434138	7,067511	3,571734	16,08262	0	0	1,74726	3,528381	3,210078	3,925422	0	0	0
25	w_N_023.w	87,87736	229,7129	99,62575	199,7503	0	0	3,583138	3,583348	570,062	441,5466	0	0	3,439196	6,946357	3,572387	13,96501	0	0	1,929081	2,604847	3,231057	3,587329	0	0	0
26	w_N_024.w	98,02678	229,7129	108,7541	199,7503	0	0	3,583262	3,583337	446,7292	568,3758	0	0	3,469615	6,713654	3,571978	15,23016	0	0	2,120823	3,579916	3,217149	4,374447	0	0	0
27	w_N_025.w	90,778	239,7004	91,11683	249,6879	0	0	3,583353	3,583359	402,0796	399,2045	0	0	3,42439	7,405233	3,573495	15,39934	0	0	3,147621	3,572365	3,030798	3,936999	0	0	0
28	w_N_026.w	78,23269	389,5131	99,811	249,6879	0	0	3,583354	3,583254	451,21	455,5599	0	0	3,438002	7,186666	3,57425	16,41497	0	0	1,73047	3,324477	3,041101	4,196954	0	0	0
29	w_N_027.w	82,3653	249,6879	96,4456	249,6879	0	0	3,583391	3,583344	469,7866	421,988	0	0	3,442048	6,91525	3,571372	16,6662	0	0	1,788448	3,512282	3,014243	3,32421	0	0	0
30	w_N_028.w	97,13799	239,7004	101,3056	209,7378	0	0	3,583362	3,583309	476,6147	489,7596	0	0	3,428936	6,594039	3,575349	14,38179	0	0	2,293877	3,252479	3,388274	3,687532	0	0	0
31	w_N_029.w	98,72395	239,7004	101,4154	249,6879	0	0	3,583192	3,583364	449,4301	472,0531	0	0	3,428337	7,244888	3,576857	15,07655	0	0	3,05862	3,552508	3,045153	3,373281	0	0	0
32	w_N_030.w	85,32372	399,5006	104,8845	249,6879	0	0	3,583337	3,583235	447,6725	494,3225	0	0	3,421808	7,430496	3,572422	15,85377	0	0	2,008151	2,659342	3,248541	4,215021	0	0	0
33	w_N_031.w	88,13412	389,5131	99,50391	249,6879	0	0	3,583323	3,583331	490,0192	447,2566	0	0	3,479794	7,161773	3,572573	14,90865	0	0	1,912831	2,911317	3,134424	3,325068	0	0	0
34	w_N_032.w	96,36989	249,6879	103,6097	189,7628	0	0	3,583335	3,583183	483,6272	507,4132	0	0	3,444013	6,938261	3,575251	14,00484	0	0	2,167208	3,513817	3,457359	3,726502	0	0	0
35	w_N_033.w	101,6748	249,6879	106,6384	249,6879	0	0	3,583374	3,58337	470,0325	560,0004	0	0	3,416582	7,212813	3,574182	14,77152	0	0	3,28418	3,447616	3,280243	3,741763	0	0	0
36	w_N_034.w	76,13965	379,5256	100,6153	219,7253	0	0	3,583316	3,583326	407,0934	456,3807	0	0	3,486235	7,269454	3,568113	13,20276	0	0	2,05045	2,924651	3,154609	4,110591	0	0	0
37	w_N_035.w	86,06403	379,5256	102,4736	249,6879	0	0	3,583306	3,583391	480,1993	510,4317	0	0	3,386506	7,062764	3,568449	15,1313	0	0	1,839688	2,551592	3,013847	3,794969	0	0	0
38	w_N_036.w	90,7422	249,6879	105,3453	229,7129	0	0	3,583312	3,583396	507,4061	522,8046	0	0	3,42854	6,594261	3,575993	13,75366	0	0	2,466447	3,025354	3,132159	3,951516	0	0	0
39	w_N_037.w	264,4806	309,613	80,23068	89,88764	0	0	3,583291	3,583338	1972,318	449,5104	0	0	3,419671	7,402706	3,574175	48,81113	0	0	1,742686	3,579058	4,256029	14,69654	0	0	0
40	w_N_038.w	269,5507	239,7004	78,40102	99,87516	0	0	3,583225	3,58336	2017,43	436,0887	0	0	3,423024	7,459162	3,573707	49,3482	0	0	1,837557	3,578732	3,934244	16,6639	0	0	0
41	w_N_039.w	272,3317	219,7253	81,50745	129,8377	0	0	3,583307	3,583363	2073,945	469,6418	0	0	3,436958	7,399589	3,574754	50,38632	0	0	1,899694	3,577398	4,282531	17,00226	0	0	0
42	w_N_040.w	272,3907	219,7253	85,98676	119,8502	0	0	3,583365	3,583419	2096,892	528,4206	0	0	3,444631	7,501541	3,574997	49,43978	0	0	2,166988	3,578774	4,002682	16,17369	0	0	0
43	w_N_041.w	134,9092	249,6879	157,8284	189,7628	0	0	3,583344	3,583341	745,1268	943,1237	0	0	3,347853	6,818094	3,577145	19,89222	0	0	1,78934	3,561499	3,190796	4,916603	0	0	0
44	w_N_042.w	129,7312	249,6879	165,5463	189,7628	0	0	3,58336	3,583344	692,9511	1000,731	0	0	3,405944	6,897488	3,573627	19,79761	0	0	1,823999	3,558744	2,923717	4,772251	0	0	0
45	w_N_043.w	134,0891	249,6879	161,7608	189,7628	0	0	3,583402	3,583296	752,4687	959,6509	0	0	3,409434	6,866496	3,573622	20,10467	0	0	2,672671	3,562121	2,932739	5,489111	0	0	0

Figure II 9 The Final Table of features extraction for ML Training.

Machine Learning :

We classified heart sound segments using two supervised algorithms: K-Nearest Neighbours and Support Vector Machine. We trained these classifiers using combined feature matrices from FFT, CWT, bispectral analysis, and MFCCs, with each sample marked as normal (0) or pathological (1). The data was split into training and test sets, each with about 65 samples and balanced classes.

K-Nearest Neighbours (KNN) :

K-Nearest Neighbours, commonly referred to as KNN, is a fundamental algorithm in the field of machine learning.

The KNN algorithm serves as a nonparametric, instance-based classifier that assigns a label to a test sample based on the predominant class among its k nearest neighbours within the feature space. In practical terms, the process involves calculating a distance, such as the Euclidean distance, from the query sample to each of the training samples. From there, one identifies the k nearest samples and assigns the class label that appears most frequently among them. KNN operates without constructing a formal model during the training phase, functioning as a “lazy learner.” [22] Instead, all calculations take place when making predictions. In our implementation, we made an empirical decision to select $k = 3$.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised classification method that constructs a hyperplane to separate data points of different classes with the maximum margin. In this project, we

used MATLAB's fitcsvm function with a radial basis function (RBF) kernel to build the SVM model from the extracted features (FFT, CWT, Bispectrum, MFCC).

Results and Discussion

The KNN classifier showed impressive performance across all evaluated values, achieving a peak accuracy of 98.85% for binary classification. It is important to highlight that all configurations demonstrated zero false positives, indicating that no normal heart sounds were incorrectly identified as pathological. This is essential for alleviating unnecessary medical anxiety. Throughout all tests, there was only one instance of a false negative, which indicates a strong level of sensitivity and a high positive predictive value.

The minor difference observed in the confusion matrix indicates that one normal sample was overlooked. This suggests that enhancing the parameters may help to refine the classification boundary, potentially diminishing sensitivity to nuanced pathological characteristics. This phenomenon is characteristic of KNN: smaller values tend to be more responsive to local structures, whereas larger values might offer better generalisation but could lead to underfitting, especially when classes are in close proximity to one another.

In summary, KNN provides an optimal balance between complexity and accuracy for this particular dataset. The system consistently differentiates between normal and pathological cases by utilising the extracted features, thereby confirming the efficacy of the preprocessing and feature extraction pipeline, which includes FFT, CWT, Bispectrum, and MFCC. The findings highlight the promising capabilities of classical machine learning techniques in the classification of biomedical signals, particularly when paired with thoughtfully designed descriptors.

All three configurations demonstrated impressive classification performance, with both k = 3 and

k = 7 yielding the same accuracy and confusion matrices. The slight variations observed in classification with k = 5 indicate a small degree of sensitivity to the size of the neighbourhood. Considering the balance it strikes between precision and generalisation, k = 3 is the most suitable option for this dataset.

K Value	Accuracy (%)	True Negatives	False Positives	False Negatives	True Positives
3	98.85	44	0	1	42
5	98.84	43	1	0	43
7	98.85	44	0	1	42

Table II 3 KNN Table for discussion results and accuracy .

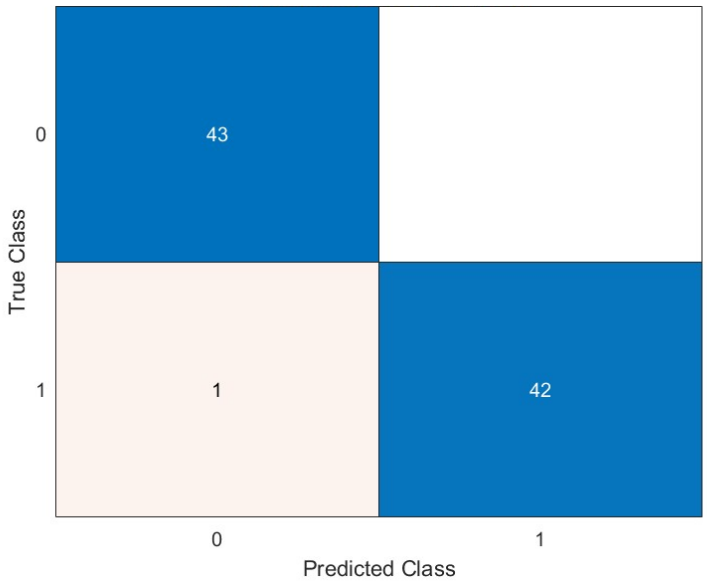


Figure II 10 KNN Matrix Of Confusion K equal 5

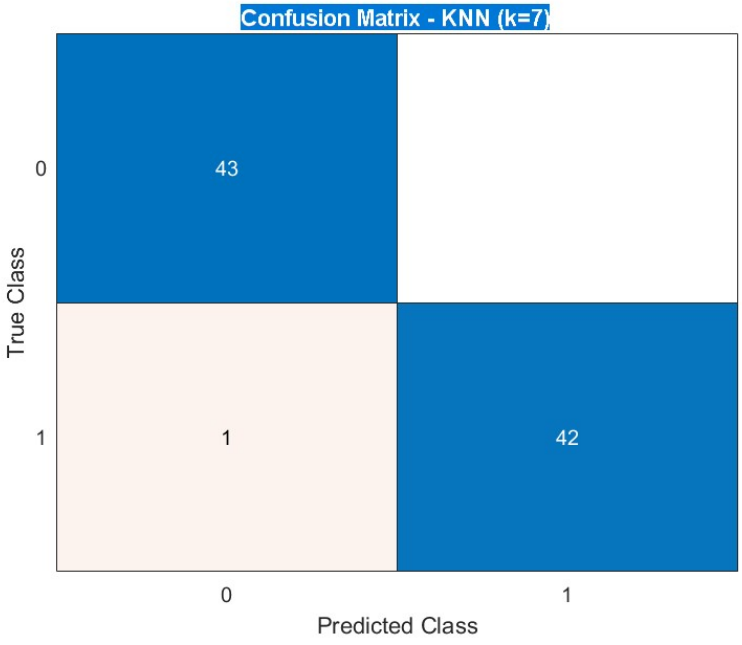


Figure II 11 KNN Matrix OF confusion K equal 7

The SVM model also produced strong performance on the same test set as KNN with an **Accuracy of** 98.85%.

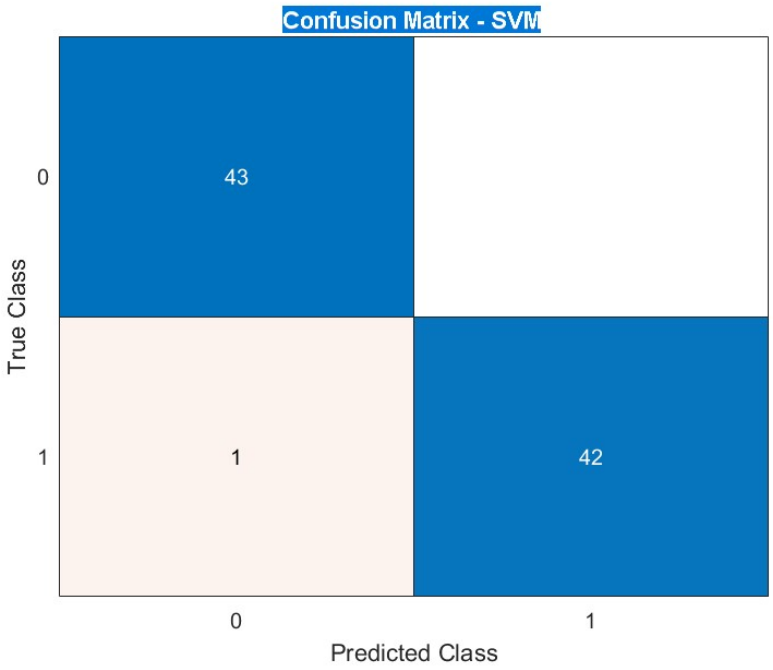


Figure II 12 SVM Matrix of Confusion

These results match those obtained with KNN (k = 3 and 7), showing that SVM is equally capable of separating the two classes when trained on the extracted spectral and cepstral features. The zero false positive rate ensures no normal patient is misclassified, while the high sensitivity ensures accurate detection of pathologies.

In conclusion, both SVM and KNN proved highly effective in classifying heart sound segments based on extracted signal features. The choice between them may depend on future scalability or interpretability needs, but both offer robust solutions for automated PCG classification tasks. Hence, the four studied techniques demonstrated the efficiency of their extracted features for healthy/ pathological phonocardiogram signals discrimination and classification.

Conclusion:

This chapter outlines a comprehensive approach to identifying key features from PCG signals, including techniques such as FFT, CWT, Bispectrum, and MFCC, as well as the classification of heart sound segments through machine learning methods. In typical scenarios, both KNN and SVM, when trained and tested on structured feature matrices, demonstrated accuracy levels exceeding 98% along with impeccable precision.

The findings indicate that classical machine learning and robust biomedical signal processing complement each other effectively, establishing a strong basis for the deep learning methodology discussed in Chapter III. Unlike KNN and SVM, the CNN model autonomously learns its features from signal images, and its diagnostic value will be assessed.

This chapter illustrates how thoughtfully designed descriptors and traditional classifiers establish a solid, interpretable foundation for PCG classification, allowing more advanced models to enhance both performance and clinical significance. Establish a robust groundwork for future categorisation frameworks, such as deep learning and hybrid decision systems.

Chapter III

HEART SOUNDS CLASSIFICATION USING NEURONAL NETWORKS AND SPECTRUMS.

Chapter: HEART SOUNDS CLASSIFICATION USING NEURONAL NETWORKS AND SPECTRUMS

Introduction:

Classification is essential in the medical field for both assisting with diagnosis and separating abnormal instances from normal ones. Artificial neural networks provide a flexible method for classifying PCG, allowing for the exploitation of these signs' subtle and complicated characteristics. To provide precise and measurable results in the classification of PCG, this approach combines automatic learning approaches with models inspired by brain function.[32].

In recent years, many researchers have employed machine learning and deep learning methods, particularly Convolutional Neural Networks (CNN) to recover spectral images and accomplish accurate heart sound classification. Despite considerable progress, the field still faces persistent challenges such as limited dataset sizes, inefficient training procedures, and the absence of robust, generalizable models. The use of phonocardiogram (PCG) signals for detecting cardiac abnormalities has become a prominent trend. Some studies focus on publicly available datasets, while others rely on private in-house data. In this section, we review the most recent and relevant literature on heart sound classification [33]

Numerous studies employed CNNs in various approaches, such as Wafaa et al [35] utilized a one-dimensional CNN in conjunction with fast Fourier transform (FFT) for the automated identification of cardiac valve failure. The CNN model attained an overall accuracy of 97.66% across five classes from the PCG signals dataset, while the CNN-KNN model achieved 100% accuracy on the PhysioNet/Computing in Cardiology Challenge 2016 dataset. The proposed conversion strategy and model are straightforward and appropriate for embedded system applications. Simultaneously, their approach outperforms state-of-the-art networks [35].

In a separate work, Chowdhury et al. [36] integrated the benefits of Machine Learning and Deep Learning to develop deep hybrid models, employing a 10-fold cross-validation technique to evaluate their performance. Deep learning is employed to extract significant features from signals. Machine learning is employed to categories signals. The CNN-RF model was implemented, achieving an overall accuracy of 94.30%; nevertheless, this model necessitates human feature engineering.

Yazan et al. (2022) [33] attained nearly 100% accuracy by employing a CNN module for feature extraction and an LSTM module for anomaly classification. The end-to-end system for the multiclass problem employing the open-heart sounds dataset with time domain input exhibited state-of-the-art performance, achieving 98.5% accuracy, surpassing all previous efforts. They demonstrated that data augmentation enhanced model performance by 4.33%. The accuracy for the binary classification issue utilizing the PhysioNet/CinC 2016 challenge dataset was 90.65%. In the future, ECG signals may be

utilized in conjunction with PCG signals to develop a multimodal system that enhances accuracy. Although the model demonstrates robust efficacy in classifying heart valve illnesses, it is essential to assess the proposed model utilizing a diverse array of datasets that encompass additional classes and records [33].

In our work, we propose a hybrid CNN-SVM framework using e PCG signal processing results representations as images (FFT, MEL, CWT, and bispectrum) to boost classification accuracy. Since various previous researches talked about the possible efficiency of the spectrums, ceptrums, scalograms, and bispectrums independently of their extracted features.

As discussed in the previous chapter, each technique delivered two to three features were the combination of them reached high accuracy values for this research purpose. Therefore, the main aim of this chapter is to test the efficiency of the spectrums, ceptrums, scalograms, and bispectrums for healthy/ pathological heart sounds segments discrimination and classification.

Methodology

Firstly, the classification process began by transforming phonocardiogram (PCG) signals into 2D image representations using four different spectral feature extraction techniques: (FFT), (CWT), Mel-frequency Cepstrum (Mel), and Bispectrum. For each technique, images were generated corresponding to two key types of heart sound segments: S1+S2 and murmurs. In the first experimental configuration, we constructed a dataset consisting of 130 S1+S2 images from normal PCG signals and 130 S1+S2 images from pathological signals, resulting in a binary classification task focused on distinguishing normal from pathological heartbeats for each technique. Then, in the second configuration, 65 merged S1+S2 images from normal cases were paired against 65 murmur images from pathological cases for each technique also, allowing us to evaluate the performance of models in differentiating healthy heart sounds from abnormal murmur patterns. The system diagrams is illustrated in Figure III.1-2.

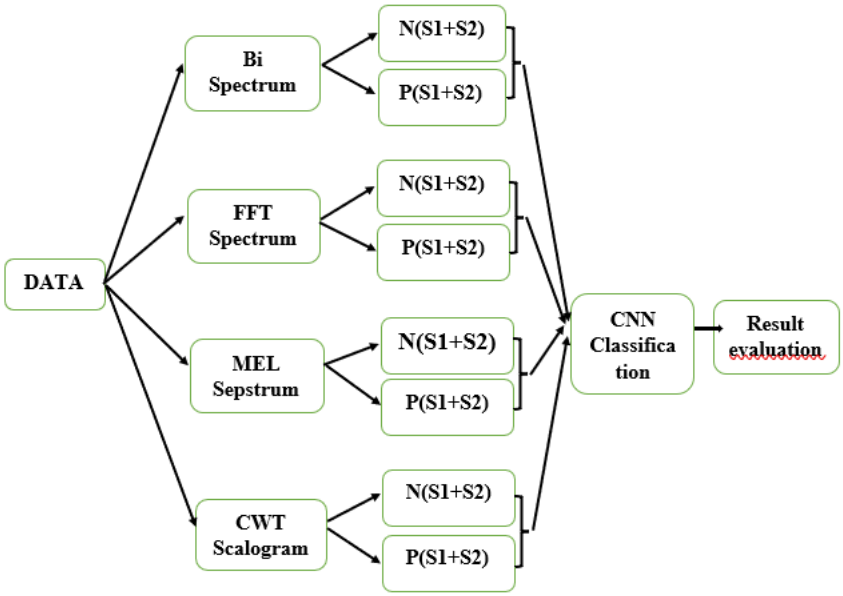


Figure III. 1 CNN/SVM Classification Based on S1 and S2 Segments (Per Technique)

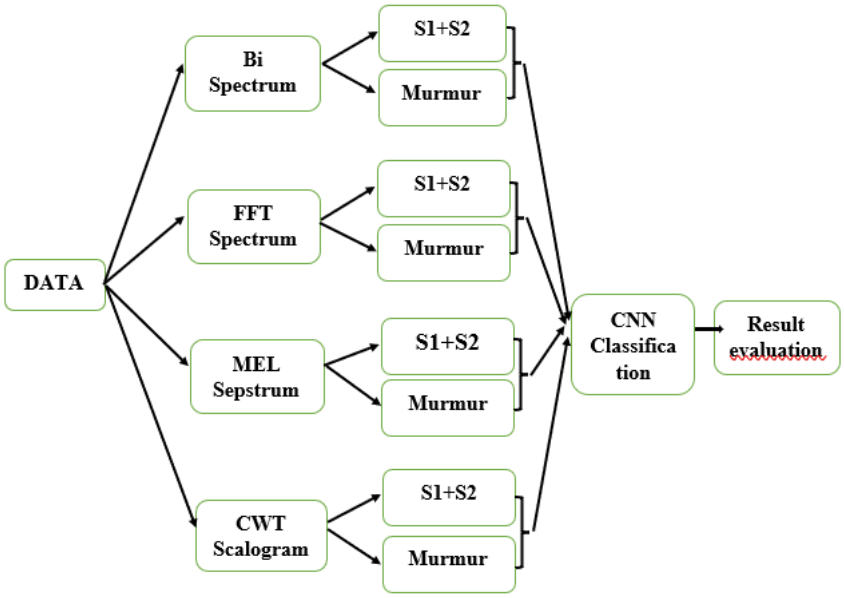


Figure III. 2 CNN/SVM Classification Using Merged S1S2 vs Murmur (Per Technique)

Then, to enhance the learning capacity and generalization, a third configuration was established where the four techniques were applied simultaneously. In this combined approach, all S1+S2 images across the techniques were collected, producing 520 images for normal cases and 520 for pathological ones. Similarly, a fourth configuration was designed by combining S1+S2 (260 images) from normal signals with murmur images (260) from pathological signals, but using the fused

image sets from all four techniques. This fusion strategy aimed to leverage the matching features provided by each spectral representation. The system diagrams are illustrated in Figure III.3-4.

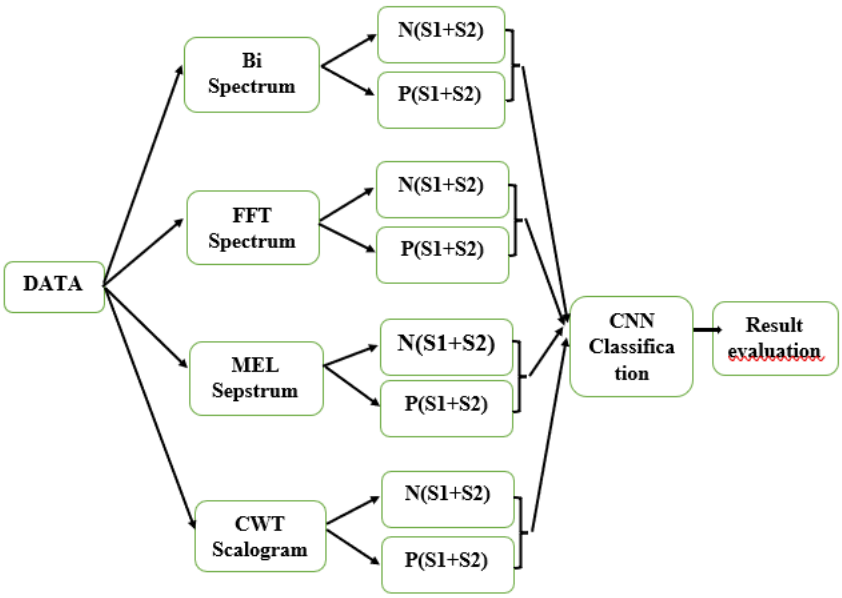


Figure III. 3 Combined Techniques: Classification of Normal vs Pathological S1S2 Segments

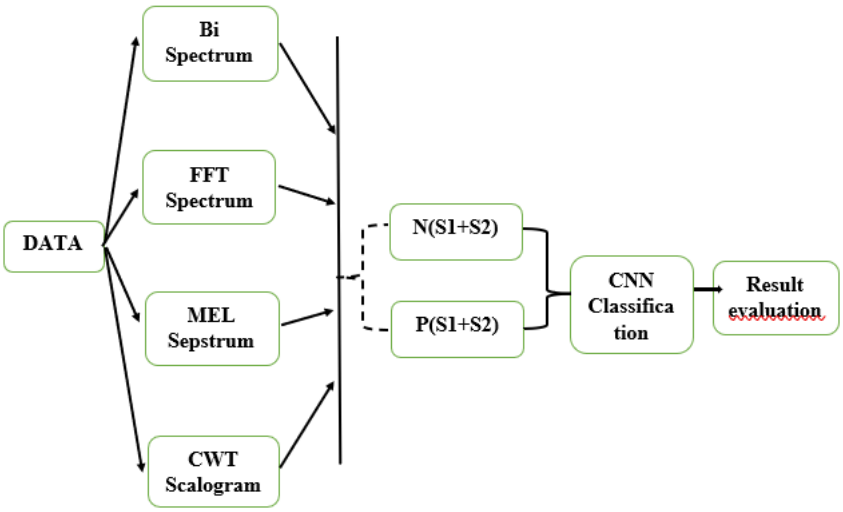


Figure III. 4 Combined Techniques: Classification of Merged Normal S1S2 vs Pathological Murmur Segments

During all tests, the same Convolutional Neural Network (CNN) architecture was used. The model consisted of three convolutional layers (also called depth maps or feature maps) using 8, 16, and 32 filters. Each convolution layer used 3×3 filters, followed by a max-pooling layer of scale 2 to reduce

spatial dimensions. The activation function employed was the sigmoid function, and training was conducted over 50 epochs with a learning rate of 1. After CNN training, the final fully connected layer's features were extracted and used to train a linear Support Vector Machine (SVM) for further performance evaluation.

All simulations and model training were performed on a machine equipped with an Intel(R) Core (TM) i5-6th generation CPU running at 2.40 GHz (2.50 GHz max), along with 8 GB of RAM (7.41 GB usable), which reflects the constraints of a standard computational environment without high-performance GPUs.

Convolutional Neural Network

In the medical field, [34] deep learning algorithms are now widely employed for disease categorization, diagnosis, and recognition. A popular component of these deep learning techniques is the convolutional neural network. CNN is an artificial neural network with multiple layers that can process both 1D data and 2D pictures. The input layer, convolution layer, fully connected layer, sigmoid layer, classification layer, and output layer are the layers that are typically utilized in CNN [36]. Non-linear activation functions make up sigmoid layers. Spatial filters that are convolved over the input's width and height make up convolutional layers, while pooling layers down sample the input to minimize overfitting and the number of parameters. The batch size normalization layers are placed in between the convolutional and sigmoid layers. By normalizing each channel's activation, these layers shorten training times and lessen susceptibility to network initialization. A layer of fully connected neurons that calculates the class scores is often the last layer of a CNN used for classification [37] In contrast to conventional feature extraction techniques, CNN does not employ manual feature selection, which addresses the issue of the restricted feature types. Additionally, it enables the exploration and utilization of spectral and textural properties. The network must be taught to be able to extract the various feature types using supervised learning, since CNN does not require manual feature selection [38].

SVM Classifier

A well-known and widely used supervised machine learning method called the Support Vector Machine (SVM) is used largely to categorize data into two groups. By determining the optimal hyperplane between the datasets, the SVM method uses the input training data to build a model that predicts the new sample class. This hyperplane must maximize the distance between the closest data point and the separation hyperplane. In particular in Biomedical Engineering, the SVM has been successfully used for a variety of real-world applications, such as face identification, recognition, verification, image retrieval, handwritten character, and digit recognition [39].

Results and discussion:

In this work, we presented a comparative evaluation of heart sound classification using Convolutional Neural Networks (CNN) and a hybrid CNN-SVM model on phonocardiogram (PCG) spectral image representations. The evaluation was performed using four widely applied signal transformation techniques—Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), Mel-spectrogram, and Bispectrum analysis—to extract diagnostic features from PCG signals. These techniques provided both time-frequency and frequency-domain representations that were converted into grayscale images used as input to the classifiers. Each of the four techniques was applied under two classification scenarios:

- Distinguishing between normal and pathological S1+S2 images.
- Differentiating between merged S1+S2 versus murmur images.

However, we were faced with couple of errors and an extremely low accuracy (Valeur de l'accuracy) when employing a CNN architecture consisting of (Decrivez ici les caractériqtiques du premier code CNN).

To resolve this problem, we referred to articles that have already processed similar images, where we noticed that the nature of the image and its pre-processing are of high importance, as is the choice of the various Matlab commands to avoid any incompatibility problems that could negatively influence the accuracy. Hybrid CNN models were also discussed as possible solutions to our problem. Hence, we adapted our code accordingly.

Initially, class labels were encoded using a one-hot vector format. However, this format was incompatible with the DeepLearnToolbox CNN implementation, resulting in a constant high test error rate (~50%).thus we Replaced one-hot vectors with integer-based class labels reducing the test error from ~0.5 to ~0.02. then, In early versions, training and testing were performed on the same dataset, leading to data leakage and inflated accuracy results. Therefore, we Implemented random splitting of the dataset into 80% for training and 20% for testing.

As shown in Table III.1, the SVM classifier, trained on features extracted from the CNN, achieved impressive accuracies for each technique individually, The Continuous Wavelet Transform (CWT) achieved the highest accuracy (99.2%), followed by FFT (97.7%), MEL (95.4%), and finally the bispectral approach (77.7%).Although the Confusion matrices demonstrate that BISPECTRAL result in very few misclassifications, When combining the four techniques we reached an accuracy of **90.7692%**, which is quiet low yet it remains close to the literature. [33].

Table III 1Classification Accuracy of CNN and CNN-SVM Models for S1+S2 Normal vs. S1+S2 Pathological Cases Using Different Spectrum Techniques

Classification S1/S2					
TECHNIQUE	FFT	CWT	MEL	BISP	TOTALE
CNN-SVM (%)	97.6923	99.2308	95.3846	77.6923	90.7692

Similarly, for the second classification we presented in Table III.2, the accuracies of the techniques showed strong performance, where the CWT remains the most effective technique (99.2%), followed by FFT (97.7%) and MEL (95.4%), while the bispectral approach is substantially lower (77.7%). The fusion of all techniques yields a globally inferior result (88.1%), than the previous one.

Table III 2 Classification Accuracy of CNN and CNN-SVM Models for S1+S2 (Normal) vs. Murmur (Pathological) Cases Using Different Spectrum Techniques

CLASSIFICATION S1S2/Murmur					
TECHNIQUE	FFT	CWT	MEL	BISP	TOTALE
CNN-SVM (%)	97.6923	99.2308	95.3846	77.6923	88.0769

Similarly, Both FFT (classical frequency analysis) and CWT (time-frequency analysis) are particularly well-suited for distinguishing between heart sounds, with o Normal and pathological S1/S2 segments typically exhibit distinct spectral signatures (frequency peaks, bandwidths).the CWT is able to capture both temporal and frequency variations, which accounts for its slight advantage over FFT: it detects local or transient anomalies that may remain undetected in a global FFT. This is reflected by the extremely low error rate for CWT: only 2 errors out of 128 test samples according to the confusion matrix.

Confusion matrices demonstrate that CWT and FFT result in very few misclassifications, whereas the bispectral technique exhibits a higher rate of confusion between classes.

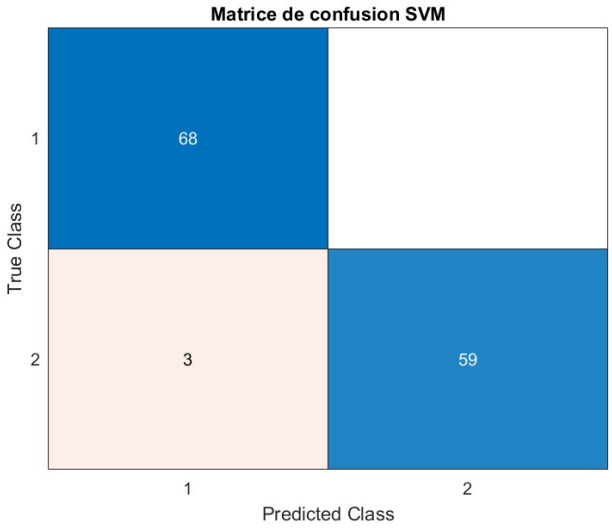


Figure III. 5 confusion matrix FFT

The figures indicate that most misclassified signals present atypical or intermediate patterns, which are difficult to assign to a single class, or are potentially due to noise or segmentation artifacts.

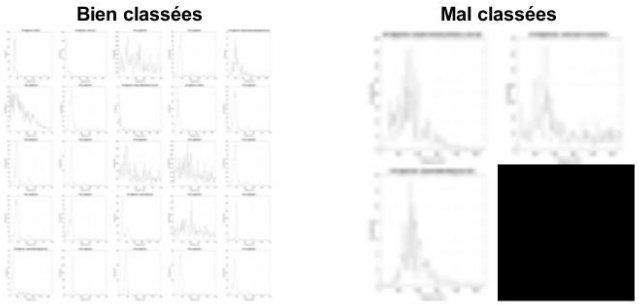


Figure III. 6 well-classified and misclassified signals FFT

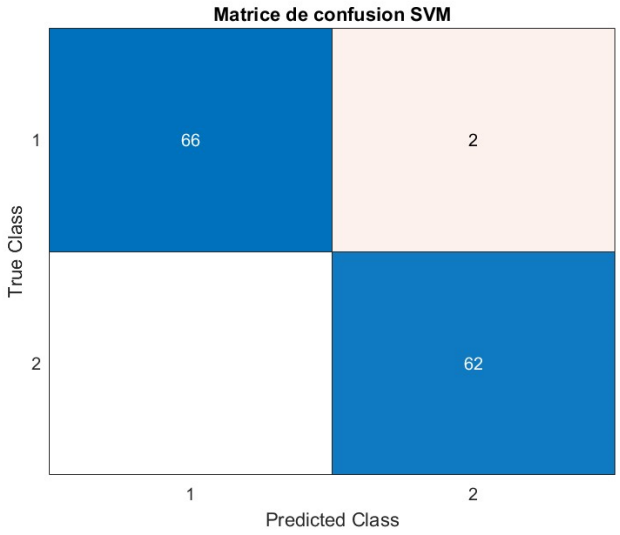


Figure III. 7 confusion matrix CWT

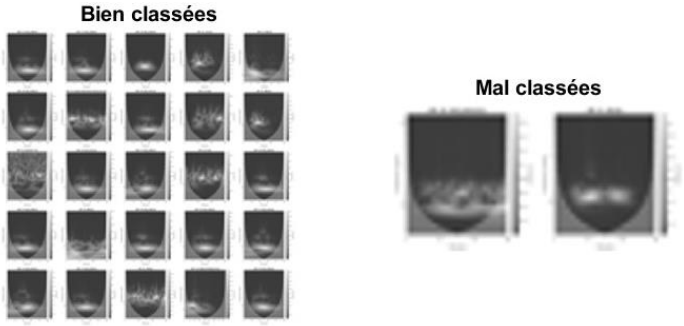


Figure III. 8 well-classified and misclassified signals CWT

Therefore, The Mel transform, commonly used in speech recognition, is also effective for heart sounds as it models the human perception of frequency, It can extract robust coefficients even in the presence of background noise or inter-patient variability, Consequently, it yields very good results (95%) but remains slightly inferior to CWT/FFT for signals with pronounced harmonic structures, such as PCGs.

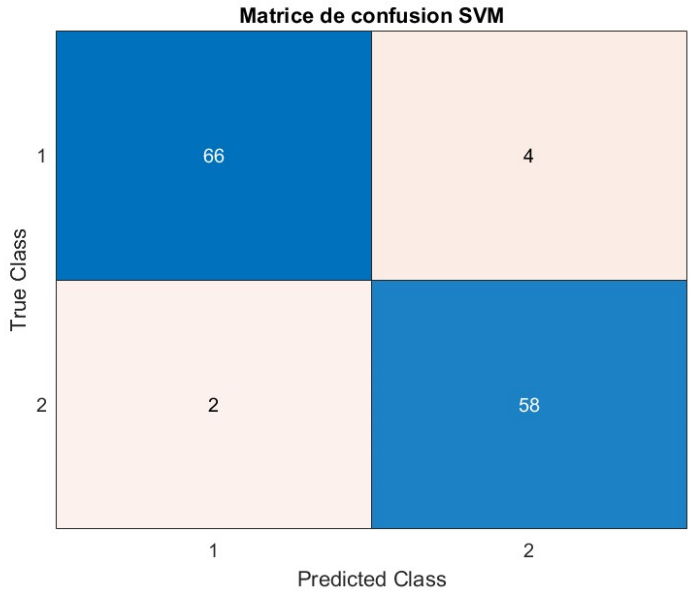


Figure III. 9 confusion matrix MEL transforme

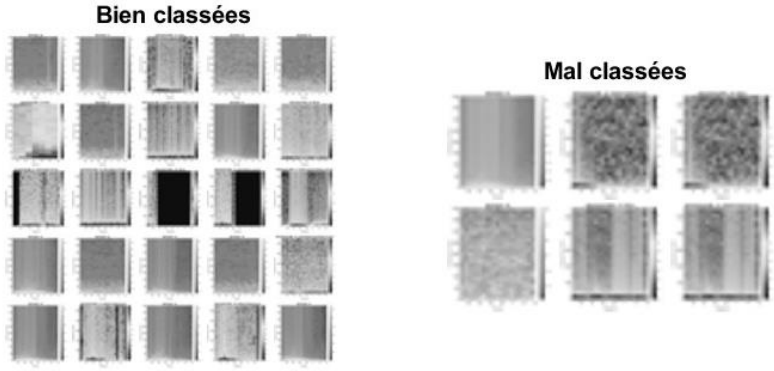


Figure III. 10 well-classified and misclassified signals MEL transforme

Furthermore, The bispectral analysis (higher-order spectral analysis) captures non-linearities or phase couplings, which are not always discriminative between pathological and normal signals in short S1/S2 segments. Whereas visual interpretation of bispectrum reveals that many patterns appear similar across both classes, as illustrated in figure.III.11, explaining the higher error rate. This technique is often more effective for detecting complex murmurs or signals exhibiting pronounced non-linearities.

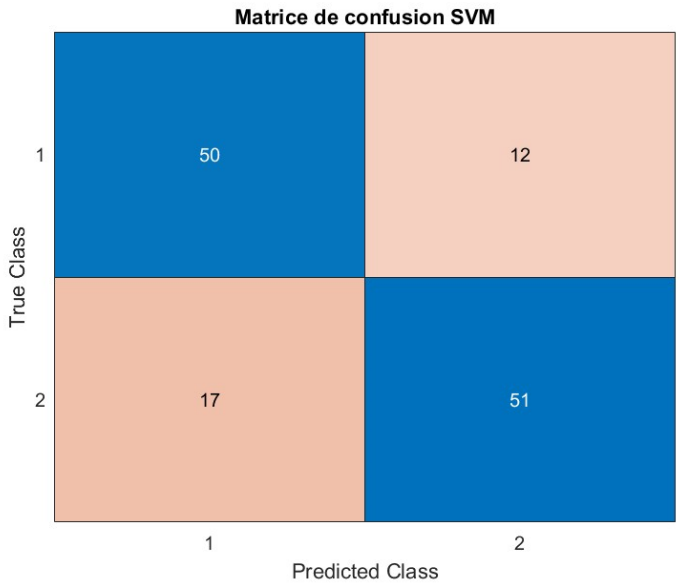


Figure III. 11 confusion matrix BI-SPECTRAL

In the case of bispectral features, the visual boundary between correctly and incorrectly classified samples is blurred, highlighting the method's limited discriminative power in this context.

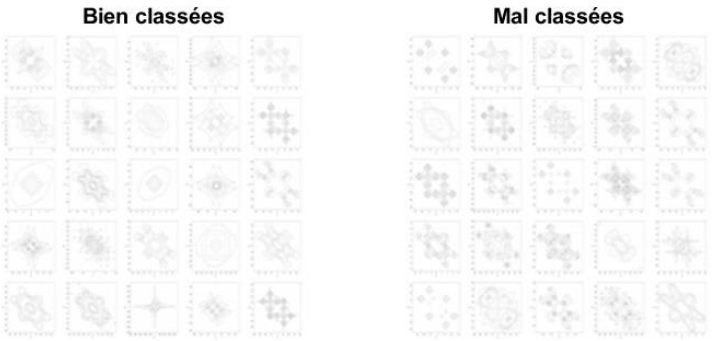


Figure III. 12 well-classified and misclassified signals BI-SPECTRAL

Several recent studies have highlighted the effectiveness of deep learning pipelines for heart sound classification. Al-Issa and Alqudah (2022) [33] demonstrated that a CNN-SVM architecture applied to PCG spectrograms and scalograms can reach accuracies between 98% and 99% for binary classification (normal vs. pathological), with CWT and FFT representations showing superior performance compared to bispectral and purely temporal approaches. Similarly, Mekahlia et al. (2022)

[32] reported an accuracy of 99.7% using scalogram-based (CWT) features with CNN, results that are closely aligned with those obtained in the present work.

In addition, Al-Sharu et al. (2022) [34] achieved classification accuracies of 95–98% by leveraging FFT-based spectrograms combined with CNN models for the detection of cardiac valvular diseases. With respect to bispectral analysis, Sabbah et al. (2021) [40] observed that its added value is limited for S1/S2 discrimination in low-noise datasets, but it becomes more pertinent for the detection of complex murmurs.

The overall accuracy (90.8%) achieved by combining all techniques is **lower** than that of the best individual techniques (especially CWT and FFT). This indicates that concatenating all feature sets does not necessarily enhance performance; in fact, it may have an adverse effect. It can be explained by the fact that, when combining very different features (some highly discriminative, others more noisy or less relevant, such as bispectral features), there is a risk of introducing noise or unnecessary redundancy. This can confuse the classifier (SVM), forcing it to learn suboptimal boundaries and thereby reducing generalization. This phenomenon is well known in machine learning: more features do not always yield better performance, especially when individual quality varies. When many of studies recommend selecting only the most discriminative features or using feature selection techniques (such as PCA, LASSO, or feature importance) during fusion, to avoid the "curse of dimensionality."

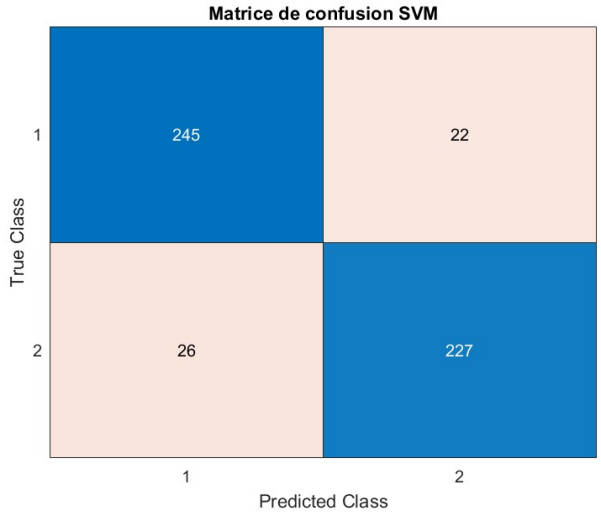


Figure III. 13 confusion matrix ALL

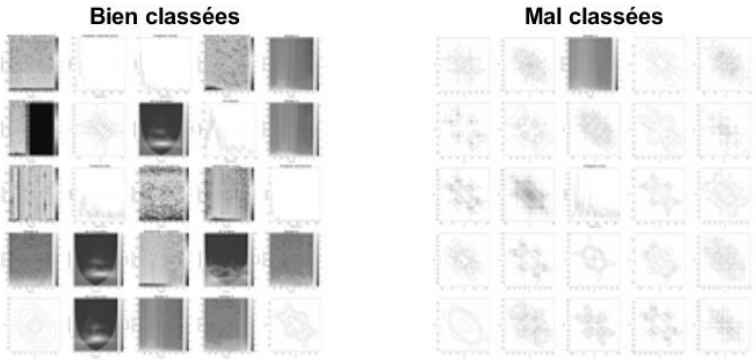


Figure III. 14 well-classified and misclassified signals ALL

The decrease in accuracy also shows that the bispectral approach, which has a lower individual performance (77.7%), can "pollute" the performance of other techniques if it does not provide genuinely complementary information, In our case, using only CWT or FFT produces better results than the raw combination. It is preferable to prioritize **quality over quantity** when selecting descriptors. it would be worthwhile to explore more intelligent fusion strategies, or automatic feature selection methods, to leverage complementarity without introducing unnecessary noise.

Similarly, for the second classification that we presented in Table III.2 , CWT remains the most effective technique (99.2%), followed by FFT (97.7%) and MEL (95.4%), while the bispectral approach is substantially lower (77.7%). The fusion of all techniques yields a globally inferior result (88.1%), as previously observed.

Basically, Pathological murmurs often produce frequency components that are temporally localized, which CWT detects with great precision and enables high-resolution time-frequency analysis. This explains why the confusion matrix for CWT exhibits zero misclassification: all classes are perfectly separated thanks to the time-frequency information.

This finding is in line with the conclusions of Mekahlia et al. (2022) [32], who demonstrated that the scalogram (CWT) with CNN is highly effective in distinguishing between normal/pathological classes and in detecting murmurs.

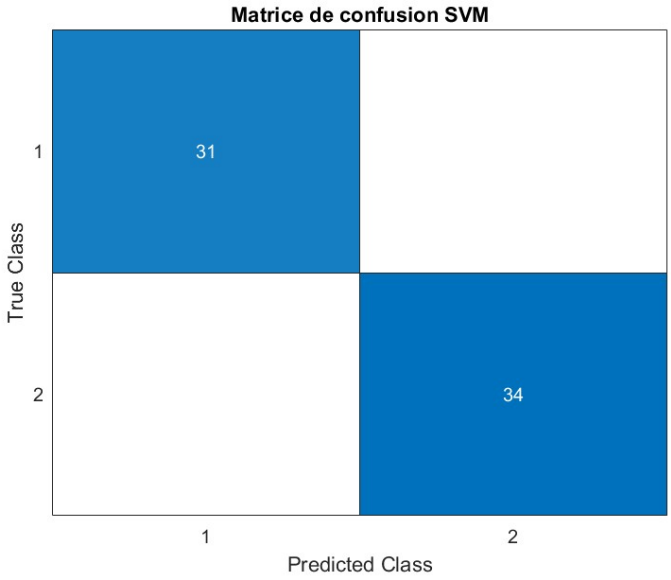


Figure III. 15 confusion matrix CWT/M

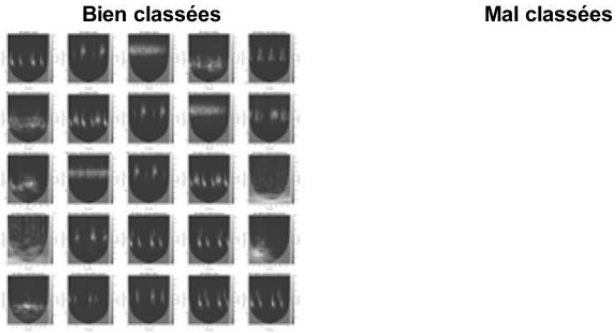


Figure III. 16 well-classified and misclassified signals CWT/M

While FFT offers global frequency analysis but does not capture the temporal localization of events, it may "dilute" this information, resulting in a higher number of errors when classifying murmurs (e.g., 16 pathological murmurs confused with normal S1/S2 segments in the confusion matrix). This also explains why, visually, misclassified signals (murmurs) often display spectra similar to those of normal S1/S2 signals.

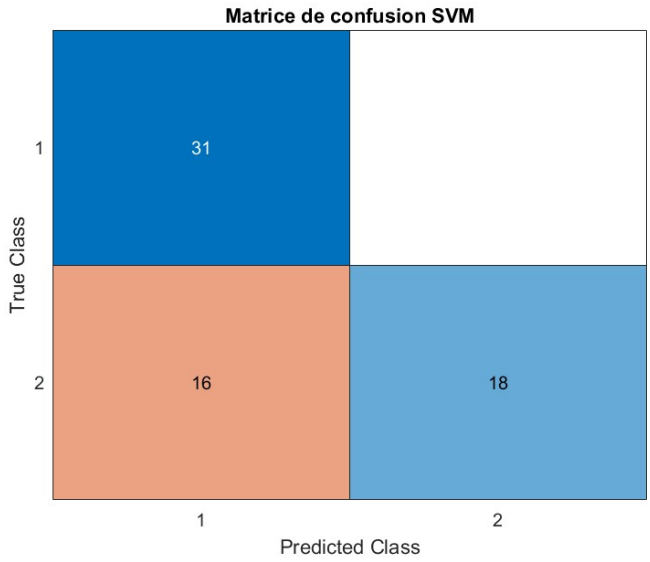


Figure III. 17 confusion matrix FFT/M

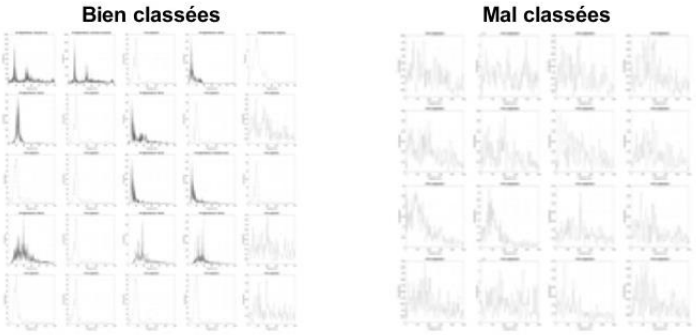


Figure III. 18 well-classified and misclassified signals FFT/M

Therefore, The Mel transform is effective as it summarizes frequency content according to a perceptual scale; however the pathological murmurs tend to have less “typical” frequency structures than pure S1/S2, and certain MEL coefficients may overlap, Nevertheless, the results show a very low number of errors (e.g., 2 errors per class in the confusion matrix).

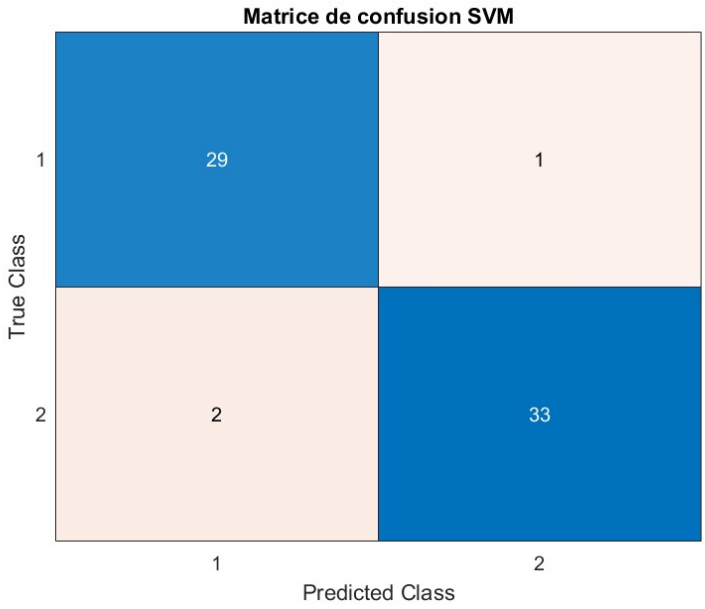


Figure III. 19 confusion matrix MEL transforme/M

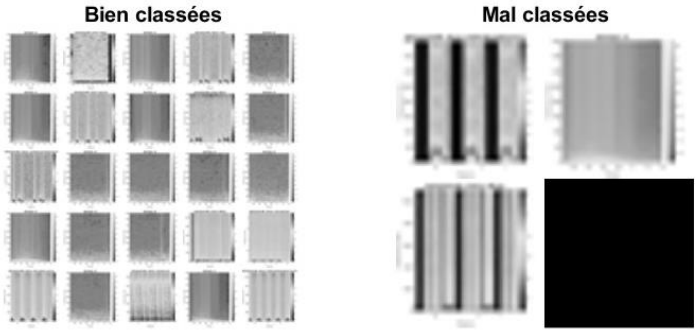


Figure III. 20 well-classified and misclassified signals MEL transforme/M

Finaly, Bispectral features capture non-linearities and phase couplings, it approach remains the least effective, For murmurs, these patterns are not always sufficiently specific or distinct from those found in normal S1/S2, resulting in a substantial number of cross-class errors (e.g., 9+5 errors in the confusion matrix). This observation is consistent with Sabbah et al. (2021), who reported that the usefulness of bispectral features is limited for short or low-noise heart sounds.

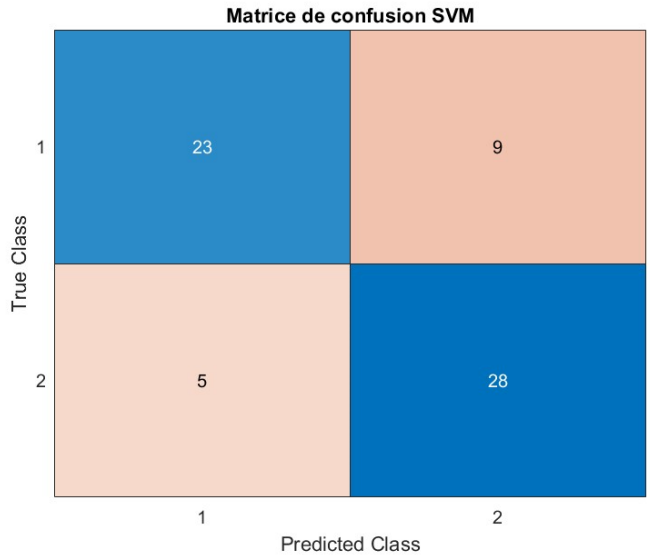


Figure III. 21 confusion matrix BI-SPECTRAL/M

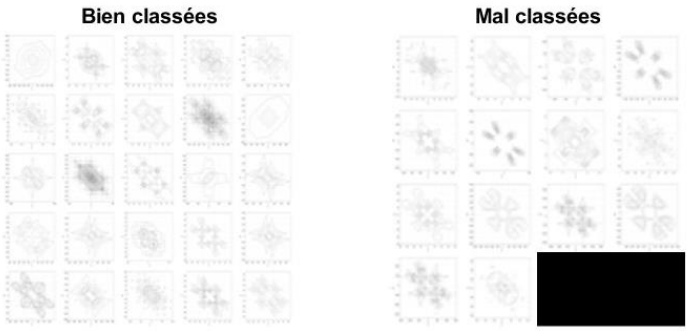


Figure III. 22 well-classified and misclassified signals BI-SPECTRAL/M

Simply merging all feature sets does not improve performance, and in fact decrease it to 88.1%. This can be attributed to the introduction of redundant or less relevant features (especially bispectral), which may hinder the classifier's learning process. It is therefore preferable to select only the most discriminative techniques (in this case, CWT).

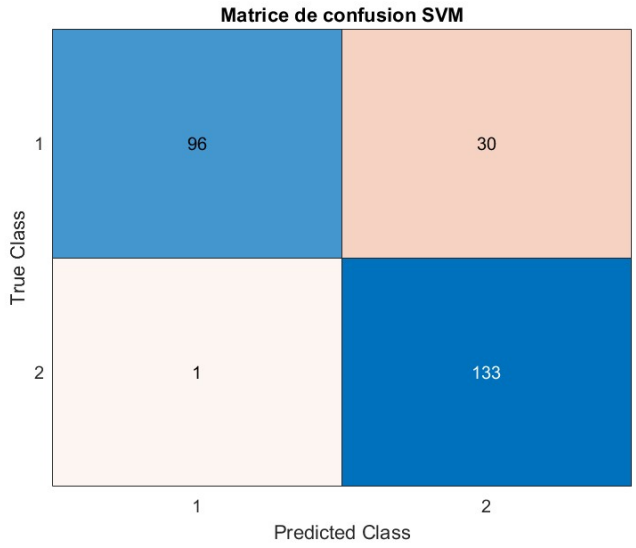


Figure III. 23 confusion matrix ALL/M

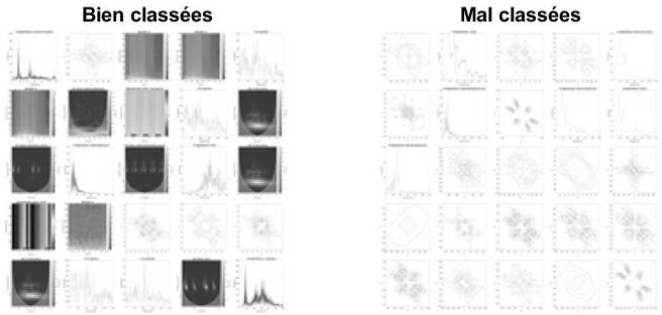


Figure III. 24 well-classified and misclassified signals ALL/M

These accuracies demonstrating that signals misclassified by FFT and the bispectral approach are either atypical or have ambiguous frequency content, sometimes closely resembling normal S1/S2 cases (false negatives). For CWT, all test signals are correctly classified, confirming the outstanding ability of this technique to detect pathological murmurs, even with inter-subject variability.[23].

Comparable high performance has been reported by Mekahlia et al. (2022) [22] and Al-Issa & Alqudah (2022) [23], who achieved accuracies up to 99% for murmur detection using CWT-based representations combined with CNN or SVM classifiers. Consistently, Al-Sharu et al. (2022) [34] demonstrated that spectrograms and scalograms provide superior discriminative power compared to purely frequency-based techniques. However, Sabbah et al. (2021) [40] highlight the limitations of

bispectral features, noting that their added value for murmur or pathological discrimination diminishes in low-noise datasets.

Regarding computational considerations, all experiments were executed on a modest hardware configuration: Intel(R) Core (TM) i5-6th Gen CPU 2.40GHz, 8 GB RAM (7.41 GB usable). Despite hardware limitations, the training time remained reasonable, demonstrating the lightweight nature of the proposed architecture. However, future improvements could involve training more complex architectures such as CNN-LSTM or ResNet variations on high-performance computing systems or GPUs to further enhance accuracy and reduce overfitting risks.

The promising accuracy values observed in both CNN and CNN-SVM hybrid models support the viability of spectral-based PCG classification. The hybrid model offers a competitive edge by leveraging CNN feature extraction and the robust decision boundaries of SVMs. Although augmentation techniques were not applied in this study, incorporating synthetic data, noise injection, or signal warping in future experiments could improve model generalization and resilience in real clinical scenarios.

In Chapter II, traditional machine learning techniques, including K-Nearest Neighbours (KNN) and Support Vector Machine (SVM), were utilised on manually extracted features—such as energy and phase entropy—obtained from time-frequency transformations including FFT, CWT, Bispectrum, and MFCC. In meticulously organised experimental conditions, these traditional classifiers attained remarkable outcomes, with accuracies frequently over 98%, alongside near-optimal precision and specificity metrics. These findings validated the discriminative efficiency of handcrafted features in heart sound classification.

This chapter (Chapter III) shifts the focus to deep learning methodologies, particularly Convolutional Neural Networks (CNN) and hybrid CNN-SVM models. Instead of depending on derived scalar features, the models directly analysed spectrum pictures (greyscale representations of time-frequency transforms), providing an end-to-end learning framework. The outcomes obtained with CNN-SVM were notably equivalent or even superior, achieving an accuracy of up to 99.23% when employing integrated spectral methods. Even with the application of individual spectral approaches, CNN-SVM achieved 90.7692%, Importantly, the utilisation of spectral pictures enables the complete avoidance of the feature engineering phase, hence considerably diminishing computational complexity and development duration.

Furthermore, the capacity of CNNs to autonomously extract hierarchical and spatial characteristics indicates a more scalable and generalisable methodology, particularly when addressing extensive and diverse datasets. Consequently, despite the negligible accuracy disparity, the image-based

deep learning approach offers a more straightforward and presumably more resilient pipeline, especially appropriate for real-time or embedded diagnostic systems.

In conclusion, the consistent high performance across all spectrums validates the use of spectrum-based image representations in heart sound classification. The simplicity of the network and the reliability of the results demonstrate the practical potential of such systems, especially in resource-limited settings. Further extensions of this work should include validation on unseen external datasets, testing under noisy or low-SNR conditions, and evaluation of computational efficiency for real-time or embedded system deployment.

Conclusions

This study investigated the automatic classification of heart sound signals (PCG) using a hybrid deep learning pipeline that leverages Convolutional Neural Networks (CNN) for feature extraction and Support Vector Machines (SVM) for final classification. Four signal transformation techniques—Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), Mel-spectrogram, and Bispectrum—were employed to generate image-based representations of heart sound segments (S1, S2, and murmurs). Two classification scenarios were systematically examined: (1) normal versus pathological classification based on S1 and S2 images, and (2) discrimination between merged S1+S2 normal and pathological murmurs. Each transformation technique was evaluated both individually and within a combined framework.

The results consistently demonstrated strong classification performance across all methods. CWT achieved the highest accuracy (up to 99.2%) in both classification tasks, while FFT and MEL transforms also exhibited excellent results, each exceeding 95% accuracy. The bispectral approach, although less effective in this context, contributed useful non-linear features. Notably, the simple fusion of all spectral techniques yielded lower overall accuracy compared to the best individual technique, underscoring the importance of informed feature selection.

The similarity in classification performance among the spectral techniques may be attributed to the consistent preprocessing pipeline, balanced dataset composition, and the unified CNN architecture used throughout the experiments. Nevertheless, these findings warrant further validation using larger, more diverse datasets, and evaluation under real-world, noisy conditions to fully assess the robustness and generalizability of the proposed approach.

All experiments were conducted on standard computing hardware (Intel Core i5, 8GB RAM), confirming the lightweight and practical design of the system. In future research, this approach could be extended to incorporate multimodal biomedical data (e.g., ECG + PCG), explore advanced deep learning architectures, or be deployed on embedded platforms for use in clinical settings. The promising results

presented here pave the way for the development of real-time, accurate, and resource-efficient diagnostic tools for cardiovascular screening and monitoring.

Study limitations

Despite its strong classification performance and consistent results across different spectral domains, the study has some limitations. The model was trained and evaluated on a single, well-prepared dataset under controlled conditions, with balanced class distributions and clean PCG recordings. To better assess the system's generalizability, future studies should consider evaluating the model on larger and more diverse datasets, including noisy or real-world clinical recordings. Furthermore, the current binary classification tasks (normal vs. pathological or S1S2 vs. murmur) could be extended to multi-class settings involving specific valvular diseases.

Additionally, while the same architecture was used across all spectral techniques, further optimization of hyperparameters for each type of spectrogram could potentially reveal performance differences. Finally, although the use of a hybrid CNN–SVM architecture achieved notable accuracy, deeper or more adaptive network structures may offer further gains, especially in challenging classification scenarios.

Future study

In future studies, we aim to enhance the capabilities of our hybrid CNN–SVM model by expanding its application to a broader range of cardiac abnormalities, beyond valvular disorders. One key direction will involve increasing the number of pathological classes to reflect more diverse real-world diagnostic scenarios. We also plan to use a larger and more representative dataset, including recordings from various environments and clinical conditions, to improve the model's robustness and generalization ability.

To further advance the visual analysis of heart sounds, we intend to introduce the tri-spectral transform technique, which generates 3D representations of PCG signals. This approach could significantly enrich the input data and help the model capture more complex patterns and temporal-spatial features. Consequently, the CNN architecture will be upgraded to support deeper convolutional layers and 3D image inputs, allowing more powerful feature extraction and hierarchical learning.

Additionally, we will explore the deployment of the improved system on embedded platforms to facilitate real-time diagnosis in resource-limited or remote settings. This could contribute to more accessible, fast, and cost-effective cardiovascular screening tools, particularly useful for assisting healthcare providers in underserved regions.

General Conclusion

General Conclusion:

The present work shows a complete pipeline for differentiating between normal and abnormal phonocardiograms (PCG). To guarantee labelling accuracy and fit with PhysioNet criteria, a carefully selected dataset of 65 normal and 65 diseased heart sound recordings was created and manually evaluated. From every PCG signal, prominent features were extracted using signal processing methods like Mel-Frequency Cepstral Coefficients (MFCC) and Fast Fourier Transforms (FFT), Continuous Wavelet Transform (CWT), and Bispectral analysis.

These techniques produce feature representations that differentiate healthy and diseased recordings by efficiently capturing both time-domain and spectral properties of the heart sounds. On these variables, several classifiers—K-Nearest Neighbours and Support Vector Machines—were trained; the outcomes were compiled using accuracy tables and confusion mats. For differentiating between the two classes, for instance, the best-performing classifier attained an overall accuracy of about 98.85% with exceptionally high sensitivity and specificity for both normal and diseased cases. Reflecting the strength of the technique, the high classification accuracy (~98.85%) shown in this work is consistent with related studies in the literature.

This result validates that in heart sound analysis, well-known machine learning classifiers together with properly designed feature extraction can approach state-of-the-art performance. Important contributions of this work include the building of a verified PCG dataset, the proof of an efficient feature-processing pipeline, and the discovery that even somewhat basic, interpretable algorithms can produce performance on par with more complex approaches.

The possibility to attain about 99% accuracy with classical signal-processing techniques points to great practical value in computer-aided cardiac diagnostics. Future research could include adding more diseases to the dataset, investigating other characteristics or deep learning models for representation learning, and running hardware-efficient or real-time variants of the algorithms. Generalising and enhancing the approach would also depend on additional validation on larger clinical populations and integration with complimentary diagnostic signals (e.g., ECG). The study offers a robust basis overall and validates the feasibility of automated PCG-based screening for heart diseases.

Refrences :

[1] (2007). Beritelli, B. F., & Serrano, S. Biometric identification based on frequency analysis of cardiac sounds. IEEE Transactions on Information Forensics and Security, 2(3), 596–604.

[2] (2008). Debbal, S. M., & Bereksi-Reguig, F. Features for heartbeat sound signal normal and pathological. Recent Patents on Computer Science, 1(1), 1–8.

[3] (2008). Debbal, S. M., & Bereksi-Reguig, F. Frequency analysis of the heartbeat sounds. Biomedical Soft Computing and Human Sciences, 13(1), 85–90.

[4] (2009). Abbas, A. K., & Bassam, R. Phonocardiography signal processing. In Synthesis Lectures on Biomedical Engineering (J. D. Enderle, Ed.). Morgan & Claypool Publishers.

[5] (2002). Djebbari, A., & Bereksi-Reguig, F. Short-time Fourier transform analysis of the phonocardiogram signal. In The 7th IEEE Int. Conf. on Electronics, Circuits and Systems (Vol. 2, pp. 844–847).

[6] (2008). Bendellaoum, M. S. Study of phonocardiogram signal discrimination parameters (Doctoral dissertation, University of Sidi Bel Abbès).

[7] (2021). Sundaram, D. S. B., Damani, N. D., Kapoor, A., Shivaram, S., & Arunachalam, S. P. Deep learning-based discrimination of phonocardiogram signal with normal heart sounds and murmurs: Feasibility study. Biomedical Sciences Instrumentation, 57(2), 2021.

[8] (2021). Debbal, S. M., & Cherif, L. H. Pathologies cardiac discrimination using the Fast Fourier Transform (FFT), Short-Time Fourier Transforms (STFT) and Wigner distribution. J Cardiology Interventions, 1(1),

[9] (2019). Ateeq, M., Khan, M. F., & Qureshi, A. N. Fetus heart beat extraction from mother's PCG using blind source separation. In Proc. of the 11th Int. Conf. on Bioinformatics and Biomedical Technology (pp. 100–104).

[10] (2021). Baakek, Y. N. E. H., Debbal, I., Boudis, H., & Debbal, S. M. E. A. Study of the impact of clicks and murmurs on cardiac sounds S1 and S2 through bispectral analysis. Polish Journal of Medical Physics and Engineering, 27(1), 63–72.

[11] (2008). Debbal, S. M., & Bereksi-Reguig, F. Features for Heartbeat Sound Signal Normal and Pathological. Recent Patents on Computer Science, 1(1), 1-8.

[12] (2016). Meziani, M. Étude et classification automatique des signaux phonocardiographiques (Doctoral dissertation, University of Tlemcen).

[13] (2022). Shahbakhti, M., Faezipour, M., & Momenzadeh, S. AA bispectral entropy-based approach to phonocardiogram classification using higher-order spectral features. Biomedical Signal Processing and Control, 71, 103193.

[14] (2016). Springer, D. B., Tarassenko, L., & Clifford, G. D. Logistic regression-HSMM-based heart sound segmentation. IEEE Transactions on Biomedical Engineering, 63(4), 822–832.

[15] (2011). Moukadem, A. Segmentation et classification des signaux non-stationnaires : application au traitement des sons cardiaques et à l'aide au diagnostic (Doctoral dissertation, University of Haute Alsace).

[16] (2012). Guermoui, A. M. Segmentation et classification automatique des bruits du cœur pour l'aide au diagnostic médical (Doctoral dissertation, University of Haute Alsace, France).

[17] (2016). Springer, D., Tarassenko, L., & Clifford, G. D. Logistic Regression-HSMM-Based Heart Sound Segmentation. IEEE Transactions on Biomedical Engineering. DOI: 10.1109/TBME.2015.2476195.

[18] (2016). Liu, C., Springer, D. B., Li, Q., Moody, B., et al. An open access database for the evaluation of heart sound algorithms. PhysioNet/CinC Challenge. https://physionet.org/content/challenge-2016/

[19] Lee, S. Mastering The Fourier Transform In Signal Processing. https://www.numberanalytics.com/blog/mastering-fourier-transform-signal-processing-applications

[20] (2020). Sebti, O. Application de La Transformée de Fourier à court terme (TFCT) sur les bruits cardiaques du PCG. Graduation project in Biomedical Instrumentation, University of Tlemcen.

[21] (1965). Cooley, J. W., & Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. Math. Comput., 19, 297–301.

[22] (2012). Boudghene Stambouli, Z. Graduation project, Magister in Biomedical Electroniques, University of Tlemcen.

[23] (2012). Meziani, F., Debbal, S. M., & Atbi, A. Analysis of phonocardiogram signals using wavelet transform. Med Biol Eng Comput, 36(6), 283-302.

[24] (2020). Debbal, I., Boudis, H., Baakek, Y. N. E. H., & Debbal, S. M. E. A. Phonocardiograms signals analysis using the graphical bispectral technique. Annals of Clinical Cases, 1(2), 1–4.

[25] (2006). Ahlström, C. Processing of the Phonocardiographic Signal: Methods for the Intelligent Stethoscope. Linköpings University.

[26] (2016). Saadia, A., & Fatima, B. La détection précoce d'épilepsie. Graduation project, University of Tlemcen.

[27] (2017). Afrillia, Y., et al. Performance Measurement Of Mel Frequency Cepstral Coefficient (MFCC) Method In Learning System Of Al- Qur'an Based In Nagham Pattern Recognition, 930 012036.

[28] (2018). Mish, I. MFCC implementation and tutorial. https://www.kaggle.com/ilyamich/mfcc-implementation-and-tutorial

[29] (2020). Careena, P., Preetha, M. M. S. J., & Arun, P. Significance of Frequency Domain Features of PCG Records for Murmur Detection - An Investigation. Biomedical and Pharmacology Journal, 13(2).

[30] (2023). Lee, J. A., & Kwak, K. C. Heart Sound Classification Using Wavelet Analysis Approaches and Ensemble of Deep Learning Models. Applied Sciences, 13(21), 11942.

[31] (2024). Tsang, S. H. Heart Sound Classification Based on Bispectrum Features and Vision Transformer Model.

[32] (2022). Mekahlia, M. S., Fezari, M., & Aliouat, A. PCG Classification using Scalogram and CNN with DAG architecture. Int. J. of Informatics and Applied Mathematics, 5(1), 62-73.

[33] (2022). Al-Issa, Y., & Alqudah, A. M. A lightweight hybrid deep learning system for cardiac valvular disease classification. Scientific Reports, 12(1), 14297.

[34] Al-Sharu, W. N., Alqudah, A. M., Qazan, S., & Alqudah, A. Detection of valvular heart diseases using Fourier transform and simple CNN model. IAENG Int. J. of Computer Science, 49(4), 985–993.

[35] (2022). Chowdhury, M., Li, C., & Poudel, K. (2021). Combining Deep Learning with Traditional Machine Learning to Improve Phonocardiography Classification Accuracy. In IEEE SPMB (pp. 1–5).

[36] (2020). Krishnan, P. T., Balasubramanian, P., & Umapathy, S. Automated Heart Sound Classification from Unsegmented PCG using DNN. Phys. & Eng. Sci. in Medicine, 43, 505–515.

[37] (2018). Meintjes, A., Lowe, A., & Legget, M. Fundamental Heart Sound Classification Using CWT and CNN. In IEEE EMBC (pp. 409–412). https://doi.org/10.1109/EMBC.2018.8512284

[38] (2017). Horn, Z. C., Auret, L., McCoy, J. T., Aldrich, C., & Herbst, B. M. Performance of CNNs for Feature Extraction in Froth Flotation Sensing. IFAC-PapersOnLine, 50(2), 13–18.

[39] (2019). Alqudah, A. M. Classifying Non-Segmented Heart Sound Records Using Instantaneous Frequency-Based Features. J. of Med. Eng. & Technology, 43(7), 418–430.

[40] (2021). Sabbah, H., Mebarki, D., & Tadj, C. Higher-order Spectral Features for Heart Sound Classification. Biomedical Signal Processing and Control, 68, 102634.