Automated Diagnosis of Heart Valve Diseases from Phonocardiogram Signals using Deep Learning

A
Thesis submitted
for the award of the degree of

DOCTOR OF PHILOSOPHY

By

SAMARJEET DAS

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI - 781039, INDIA
OCTOBER 2023

Certificate

This is to certify that the thesis entitled "**Automated Diagnosis of Heart Valve Diseases from Phonocardiogram Signals using Deep Learning**", submitted by **Samarjeet Das** (176102005), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements per the institute's regulations and has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. S. Dandapat,
Professor
Dept. of Electronics and Electrical Engg.
Indian Institute of Technology Guwahati
Guwahati - 781039, India.

Place: Guwahati
Date:

To

The Almighty Lord Jagannath

for His blessings

My Beloved Parents

Mr. Amarjeet Das and Mrs. Kaibalya Das

for their blessings, love and encouragement

&

My Wife

Pragyanshree

for her love, support, sacrifice and patience

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Samarendra Dandapat, for his dedicated mentorship, encouragement, and guidance in life. His insightful feedback has helped me in improving the quality of my thesis. I greatly admire his attitude towards research, creative thinking, and enthusiasm for research work.

I am thankful to my doctoral committee members, Prof. P. K. Bora, Prof. R. Sinha, and Prof. T. Jacob, for their encouragement and valuable suggestions on my work. I am very grateful to them for their insightful comments and constructive criticisms of the work. I am deeply thankful to Dr. Neeraj Kumar Sharma, Assistant Professor at Mehta Family School of Data Science and Artificial Intelligence, IIT Guwahati, for his invaluable suggestions and unwavering guidance during the concluding phase of my Ph.D. journey. I would like to express my gratitude to the Head of the Electrical and Electronics Engineering Department at IIT Guwahati for providing me with the necessary facilities to conduct my research work. I also would like to thank all the anonymous reviewers for their valuable suggestions about my work. I would also like to thank all other faculty members of the Department of Electronics and Electrical Engineering, IIT Guwahati, for their care and support. I am also very thankful to Dr. L. N. Sharma, senior technical officer, and other staff members of the Department for their help and support. I would like to extend my gratitude to the administrative and technical support teams for their invaluable assistance throughout my research journey.

I am thankful to my friend and labmate, Debasish Jyotishi, for his valuable suggestions and critical thinking towards my research work. I am also thankful to my EMST labmates Sibasis, Salman, Himashree, Mousumi, Ato, Sukanya, James, Omesh, Akriti, Sumit, Pooja and Rohan for their unsolicited help and support. I would also like to thank other campus friends, Atanu, Arunima, Manoj, Birjit, Nishant, and Raju. I would also like to thank my seniors, Dr. Tilendra Choudhary, Dr. Alex Paul Kamson, Dr. Vineeta Das, Dr. Eedara Prabhakararao, and Dr. Aniruddha Mazumdar, for their guidance in my PhD journey.

Finally, my heartiest thanks to my parents, my wife, and family members for their constant blessings and silent prayers for successfully completing my thesis work.

Samarjeet Das

Abstract

Heart valve diseases (HVDs) are a major cause of mortality worldwide, particularly in developing and underdeveloped countries. Early diagnosis of HVDs is essential to prevent the progression of these diseases to lethal heart conditions. Phonocardiography (PCG) is a non-invasive, user-friendly, and cost-effective tool that can be used for preliminarily diagnosing HVDs. PCG signals provide a graphical representation of heart sounds (HSs), which can be used to identify characteristic features of HVDs. However, raw PCG signals are often susceptible to noise and artifacts, which can degrade the signal quality and make it challenging to diagnose HVDs manually. Additionally, the wide variability in PCG morphologies due to HVDs can make manual examination subjective and prone to human error. To address these challenges, this dissertation focuses on developing automated deep-learning methods for diagnosing HVDs.

The manifestation of HVDs in PCG cycles shows diverse temporal and morphological variations. Analysis of a complete PCG cycle provides reliable and accurate information related to HVDs. Existing methods have difficulty modeling a PCG cycle due to the non-stationary nature of the PCG signal. These methods have modeled a short segment of the PCG cycle to classify a particular type of HVD. However, considering the wide variabilities of PCG characteristics due to HVDs, these methods may not generalize well to diagnose unseen HVDs. This thesis proposes a novel multi-component oscillatory model for the diagnosis of HVDs. The model captures the morphological variations of the PCG cycles by fitting a half-period sine wave between two successive zero-crossing points. The model parameters, i.e., amplitude and frequency, provide discriminative features to a deep neural network (DNN) classifier to classify the PCG cycles as healthy or pathological. The performance results show that the proposed model is competitive enough with state-of-the-art methods. The model parameters are also used to successfully reconstruct the healthy and pathological PCG cycles. The proposed model is easy to implement as only zero-crossing points have to be detected and fit

a half-wave sine function between them. The improved performance of the proposed model demonstrates its potential for accurate and reliable diagnosis of HVDs.

Heart murmurs (HMs) are the rasping or blowing sound produced due to the turbulence in blood flow through the defective heart valves or septal defects. HMs have a diverse etiology and show temporal variability in the PCG signal. The existing machine learning (ML) methods cannot exploit the temporal variability in the PCG signals associated with HMs. Further, designing a fixed set of hand-crafted features to train an ML model is challenging due to variability in HMs manifestation and is often subjective. To address these issues, we propose a novel acoustic feature-fusion method using deep learning (DL) for HMs classification. In the proposed method, we extract two acoustic features from the PCG signal, i.e., Mel-frequency cepstral coefficients (MFCCs) and linear prediction cepstral coefficients (LPCCs). Then, we fuse these two features using a hierarchical long-short memory (HLSTM) network to exploit the temporal variation in the fused features to detect HMs. Further, a self-attention module is introduced to weigh the encoded vectors of the HLSTM network based on their clinical relevance to improve the HMs classification. The robustness of the proposed method is verified by adding different environmental and body-generated noises for different signal-to-noise (SNR) levels. The proposed method shows a reasonable average run time (2-4 ms) and excellent overall accuracy compared to the existing methods.

The existing wavelet transform (WT) methods to analyze the PCG signal to detect HMs, suffer resolution loss at higher subbands (coarse scales) and lack shift-invariant properties. Further, the extracted features from the WT subbands are fed to an ML classifier to detect HMs. However, these ML classifiers cannot exploit the temporal variations within and across subbands to detect HMs. First, we propose a stationary wavelet transform (SWT) based PCG signal decomposition to address the above issues. The SWT captures the PCG patterns adequately at coarse scales to detect HMs. Then, we propose subband-specific HLSTM networks to exploit the temporal variation across each SWT subband. Further, intra-subband and inter-subband attention modules are proposed to weigh the encoded vectors of HLSTM networks to improve the HMs classification. We test the robustness of the proposed method by adding a variety of

environmental and body-generated noises at different levels of SNR. The proposed method shows improved performance results compared to the existing state-of-the-art methods.

Existing literature on HVDs detection has mainly focused on classifying PCG signals as healthy or pathological. A few studies have analyzed multi-class HMs classification. However, they are limited to small databases. Furthermore, existing DL methods have mainly used transfer-domain approaches to represent PCG segments as 2D images. These images are then fed to a 2D convolutional neural network (CNN) model to classify HMs. However, representing the time-series PCG signal in an image form may not capture all relevant HM information. This thesis proposes a multi-kernel residual CNN (MK-RCNN) model to classify three HM severity stages: absent HMs, soft HMs, and loud HMs, along with low-quality PCGs, to improve the HM assessment. The multi-kernel CNNs with different filter sizes capture scale-specific features from the PCG segment to detect HM patterns. The residual learning helps to extract deep features from deep CNN layers without degrading the performance accuracy. Experimental results show the efficacy of the proposed MK-RCNN model over existing DL models. The proposed method provides a robust end-to-end framework for HM diagnosis without manual intervention.

To summarize, this dissertation presents novel DL methods for automated HVDs diagnosis, including a multi-component oscillatory model, an acoustic feature-fusion using an attention-based-HLSTM network, an SWT-based decomposition with subband-specific attention-based-HLSTM networks, and an end-to-end framework based on the MK-RCNN model. These automated diagnostic systems can help improve the accuracy and reliability of HVDs diagnosis in primary healthcare units.

Keywords: Phonocardiogram (PCG), heart valve diseases (HVDs), heart sound (HS), heart murmurs (HMs), deep learning (DL), hierarchical long short-term memory (HLSTM), self-attention, oscillatory model, residual convolutional neural network (RCNN).

Contents

List of Figures

List of Tables

List of Acronyms

HS	Heart Sound
HVD	Heart Valve Disease
HM	Heart Murmur
MVP	Mitral Valve Prolapse
MR	Mitral Regurgitation
AS	Aortic Stenosis
MS	Mitral Stenosis
N	Normal
A	Abnormal
H	Healthy
P	Pathological
AHM	Absent Heart Murmur
SHM	Soft Heart Murmur
LHM	Loud Heart Murmur
ML	Machine Learning
DL	Deep Learning
WT	Wavelet Transform
TF	Time Frequency
DWT	Discrete Wavelet Transform
SWT	Stationary Wavelet Transform
MFCC	Mel-frequency Cepstral Coefficients
LPCC	Linear Predication Cepstral Coefficients
PRD	Percentage Root Mean Square Difference
PCC	Pearson Correlation Coefficients

HMM	Hidden Markov Model
SVM	Support Vector Machine
RF	Random Forest
DNN	Deep Neural Network
GPU	Graphical Processing Unit
KNN	K-nearest Neighbors
BPF	Band Pass Filter
LPF	Low Pass Filter
MSE	Mean Squared Error
Pr	Precision
OPr	Overall Precision
Re	Recall
ORe	Overall Recall
Se	Sensitivity
OSe	Overall Sensitivity
OSp	Overall Specificity
F1	F-score
Acc.	Accuracy
OA	Overall Accuracy
SD	Standard Deviation
TN	True Negative
TP	True Positive
FP	False Positive
FN	False Negative
SNR	Signal to Noise Ratio
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
HLSTM	Hierarchical LSTM
MKRCNN	Multi-kernel Residual CNN
GAP	Global Average Pooling

FC	Fully Connected Layer
CICC	Computing in Cardiology Challenge
HSM	Heart Sound Murmur
CCDP	CirCor Digiscope Phonocardiogram

1

Introduction

Contents

Cardiovascular diseases (CVDs) are a group of disorders that affect the heart and blood vessels [1]. They are primary causes of mortality worldwide, resulting in about 32% of global deaths [2]. To diagnose CVDs, cardiologists commonly use an electrocardiogram (ECG) signal as a standard tool [3]. However, heart valve diseases (HVDs) pose a challenge in ECG-based detection. HVDs are conditions in which valves that regulate blood flow through the heart may be damaged or diseased. As a result, they can cause the heart to work less efficiently. These anomalies do not alter the heart's electrical activity recorded in the ECG signal. They manifest in the heart sounds (HSs) produced by the movement of cardiac valves [4]. HVDs are prevalent in developing countries, resulting in 1.4 million deaths annually. It is mainly due to the lack of preliminary screening and diagnosis [5]. Therefore, early diagnosis of HVDs is essential to reduce health complications and improve quality of life.

The current diagnostic tools for HVDs include echocardiography (echo), chest X-ray, and cardiac MRI [6]. The need for sophisticated equipment and expertise poses a challenge in underprivileged regions. Cardiac auscultation involves listening to the HSs using a stethoscope. It remains a cost-effective and widely available method for detecting HVDs in primary healthcare units [7]. However, manual auscultation requires subjective expertise and may lead to a miss or false detection, which can be life-threatening or costly. A phonocardiogram (PCG) signal-based computer-aided stethoscope provides a graphical representation of HSs and is a valuable alternative. PCG-based diagnosis is non-invasive, user-friendly, and cost-effective, making it suitable for primary healthcare units.

In recent years, the development of wearable technologies and sensors has led to the emergence of various PCG-based data acquisition systems [8]. However, PCG recordings are often susceptible to artifacts and noises that can degrade signal quality and impact clinical utility. Furthermore, the diverse etiology of HVDs can result in variations in the intensity and morphology of the PCG waveform. It makes a manual analysis for HVD diagnosis challenging, prone to errors, and time-consuming. Therefore, automated diagnosis of HVDs using PCG can be a promising objective for assisting cardiologists in treatment [9].

This dissertation focuses on developing automated and reliable diagnostic methods for HVDs using PCG. This chapter briefly introduces the heart's anatomy and heart sounds (HSs), followed by the PCG signal and its clinical significance. We then review the existing automated methods for diagnosing HVDs using PCG. Further, we highlight the motivation of the research work. Following

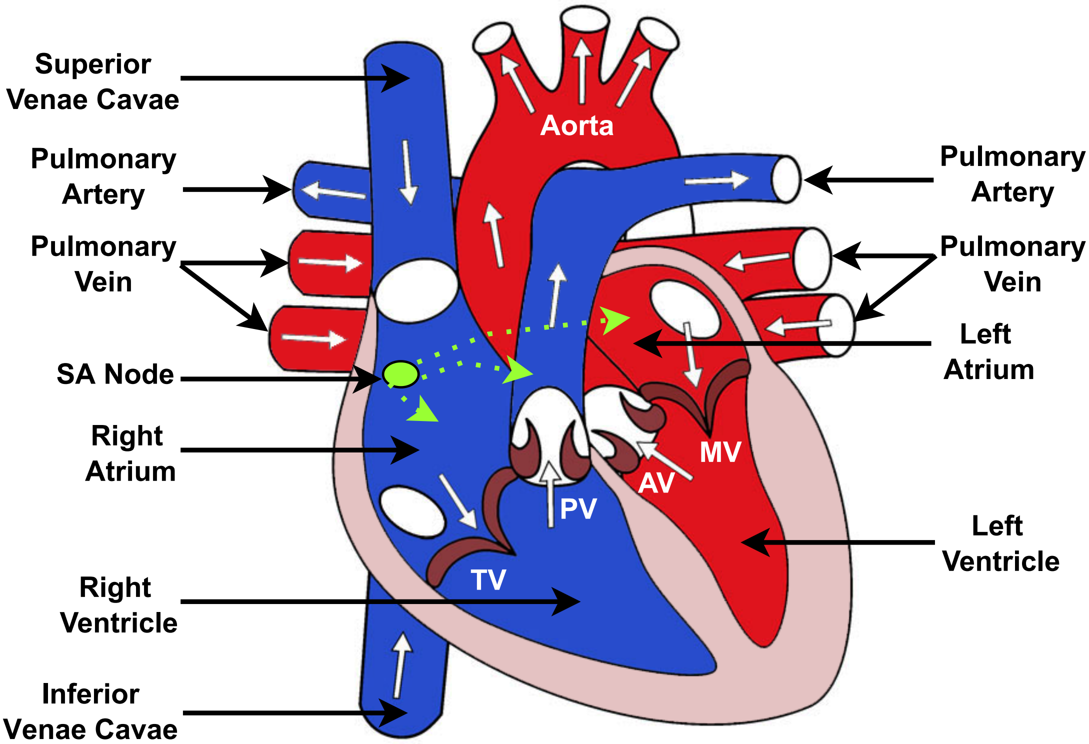

Figure 1.1: Anatomy of the heart. **TV:** Tricuspid Valve; **PV:** Pulmonary Valve; **AV:** Aortic Valve; **MV:** Mitral Valve.

that, we present the contributions and outline of the thesis.

1.1 Anatomy of the Heart and Introduction to Heart Sounds (HSs)

The heart is a vital organ in the human body. It pumps blood, which carries oxygen and nutrients through the blood vessels to the body parts [10]. It also transports metabolic waste, such as carbon dioxide, to the lungs for elimination. The heart is located in the middle compartment of the chest and between the lungs. It is approximately the size of a closed fist. Figure 1.1 shows the anatomy of the heart. The heart contains four chambers. The upper two chambers are the left atrium (LA) and the right atrium (RA), while the lower two chambers are the left ventricle (LV) and the right ventricle (RV). The left heart contains the LA and LV. The right heart contains the RA and RV. The RA receives deoxygenated blood from the body through the superior and inferior venae cavae. At the same time, the pulmonary veins transport oxygenated blood to the LA. The sinoatrial node initiates the heart's electrical activity. It generates periodic electrical signals propagating through the heart, causing its muscle to contract. During atrial contraction, blood stored in the RA is pumped into the RV through the tricuspid valve. At the same time, blood stored in the LA is pumped into the LV through the mitral valve. During ventricular contraction, oxygenated blood in the LV is pumped out of the heart to body parts via the aortic valve, while deoxygenated blood in the RV is pumped to the lungs via the pulmonary valve. The period of ventricular contraction is known as systole, while the period of atrial contraction or ventricular relaxation is known as diastole. One cardiac cycle consists of one sequence of systole followed by diastole.

The human heart produces two audible sounds, known as heart sounds (HSs), during a cardiac cycle. The first HS, known as the lub sound or S1 sound, signifies the beginning of systole. It is generated by the closure of the atrioventricular valves, i.e., the mitral and tricuspid valves. Similarly, the second HS, known as the dub sound or S2 sound, marks the beginning of diastole. It is produced by the closure of the semilunar valves, i.e., the aortic and the pulmonary valves. However, additional HSs during systole, diastole, or both are signs of a pathological heart. These additional HSs include S3, S4, or heart murmurs (HMs). S3 is a ventricular gallop produced by a sudden increase in ventricular blood volume. S4 is an atrial gallop produced by blood flow into a stiffened ventricle. HMs produce rasping or blowing sounds due to HVDs or congenital heart diseases [11, 12]. These pathological HSs can provide crucial diagnostic information for cardiovascular assessment. Table 1.1 presents a brief specification and physiological description of these healthy and pathological HSs. S1 sound lasts 120-150 ms and has a frequency range of 20-200 Hz. S2 sound lasts 80-120 ms and has a frequency range of 20-250 Hz. S3 sound lasts 40-100 ms and has a frequency range of 25-70 Hz. S4 sound lasts 40-80 ms and has a frequency range of 15-70 Hz. HMs typically lie in the frequency range of up to 600 Hz. Clicks and snaps are typically characterized by a 100-800 Hz frequency range. These HSs and their associated frequency ranges provide valuable information for diagnosing HVDs.

1.2 The Phonocardiogram (PCG) Signal and its Clinical Significance

The phonocardiogram (PCG) is a non-invasive, flexible, and cost-effective tool for the automated diagnosis of HVDs [13, 14]. A PCG signal is recorded using a cardiac microphone or pressure transducer placed on the chest surface [15]. As shown in Figure 1.2, four locations on the chest surface are commonly used to acquire PCG signals. These locations are selected based on the

Table 1.1: Healthy and pathological heart sounds (HSs) with their specifications.

Heart Sound	Duration (ms)	Freq. (Hz)	Physiological description of HS
S1	120-150	20-200	Closure of atrioventricular valves
S2	80-120	20-250	Closure of semilunar valves
S3	40-100	25-70	Ventricullar gallop (Early diastolic filling of ventricle)
S4	40-80	15-70	Atrial gallop (Late diastolic filling of ventricle)
Heart Murmurs	—	Up to 600	Turbulent flow of blood through defective heart valves
Clicks	—	100-800	Opening of rigid and calcifies semilunar valves
Snaps	—	100-800	Sudden opening of stiffed mitral valve

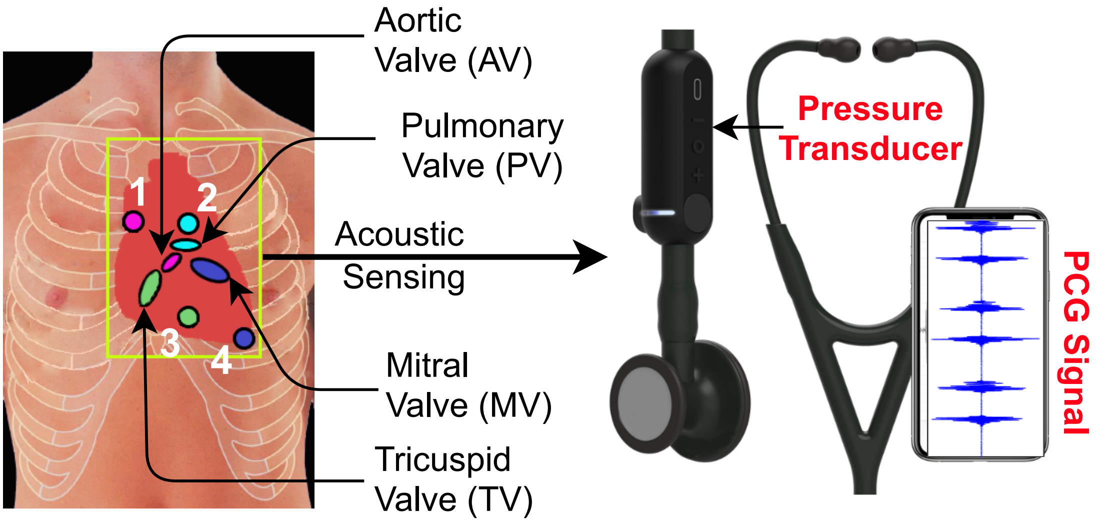

Figure 1.2: PCG signal recording on the chest surface using a pressure transducer. Four locations on the chest surface are used to acquire the PCG signal: 1. near the AV area, 2. near the PV area, 3. near the TV area, and 4. near the MV area.

optimal positions for hearing the HSs [13]. Specifically, the four locations are near the aortic valve (AV), pulmonic valve (PV), tricuspid valve (TV), and mitral valve (MV). The AV area is located at the second right intercostal space, the PV area is in the second intercostal space along the left sternal border, the TV area is in the fourth intercostal space along the left sternal edge, and the MV area is at the cardiac apex, in the fifth intercostal space on the midclavicular line.

A typical healthy PCG recording contains two fundamental HSs, namely S1 (lub or first HS) and S2 (dub or second HS), which occur in every PCG cycle [8,16]. S1 marks the onset of systole, while S2 marks the onset of diastole. Figure 1.4(a) shows a healthy PCG recording. In cases of cardiac pathology, additional HSs may be present, such as clicks, snaps, S3 sound, S4 sound, or HMs.

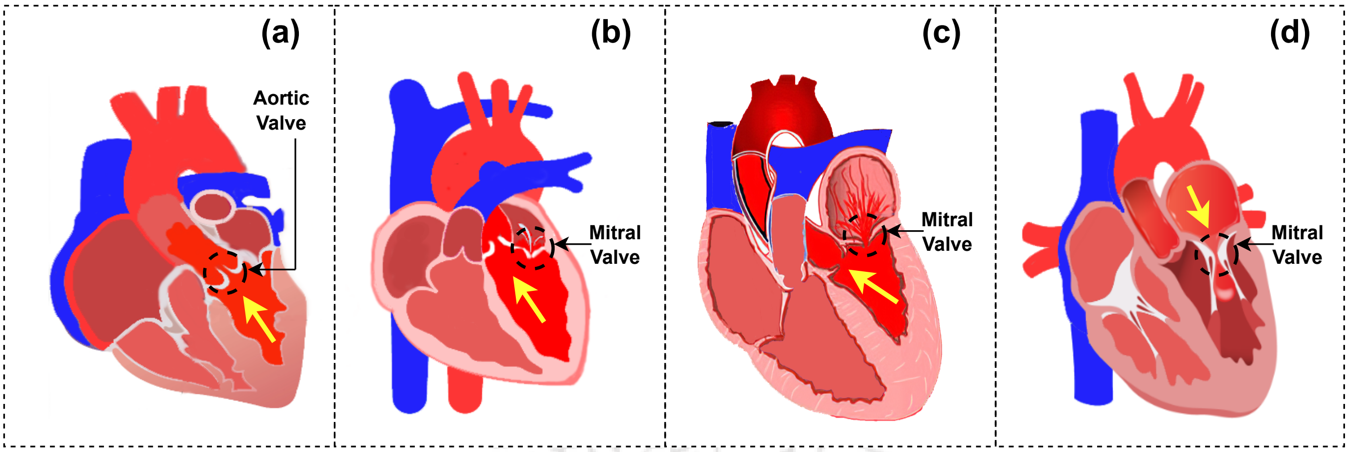

Figure 1.3: Illustration depicting the pathophysiology of heart valve diseases (HVDs). (a) AS murmur, (b) MVP murmur, (c) MR murmur, and (d) MS murmur.

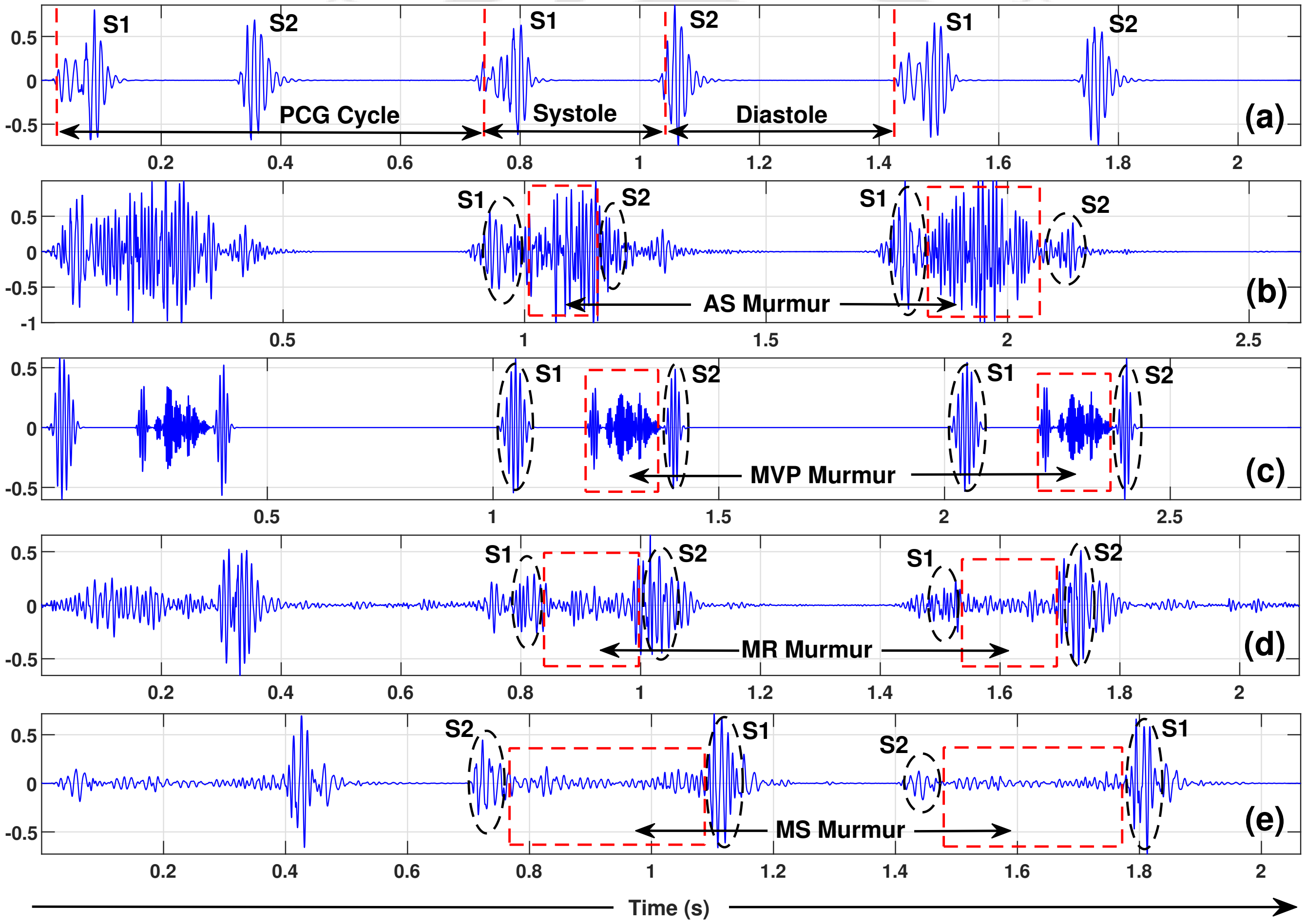

Figure 1.4: Recording of healthy and pathological PCG signals. (a) Healthy PCG signal. PCG signal affected with (b) AS murmur, (c) MVP murmur, (d) MR murmur, and (e) MS murmur.

1.2.1 Clinical Background of Heart Murmurs (HMs)

Heart murmurs (HMs) are subtle abnormal HSs produced mainly due to the turbulent flow of blood through the defective heart valves. There are four main types of HMs: valvular stenosis, valvular insufficiency, congenital anomaly, and increased blood flow through a normal valve [17]. HMs can occur in different phases of the PCG cycle, such as systolic, diastolic, or continuous, and can have various shapes, including crescendo, decrescendo, crescendo-decrescendo, or plateau.

Aortic stenosis (AS) is an HVD that primarily affects the elderly population due to the narrow opening of the aortic valve orifice [18]. As a result, it restricts the blood flow from the left ventricle to the body. Figure 1.3(a) shows the narrow opening of the aortic valve due to AS. The AS is characterized by a systolic murmur with a crescendo-decrescendo shape in the PCG signal. Figure 1.4(b) shows a PCG signal affected by AS murmur. It typically occurs during systole, reaches its maximum intensity in mid-systole, and gradually decreases until the end. AS also may produce a delayed and decreased S2 sound. In addition, the turbulent blood flow due to AS may produce a harsh systolic ejection murmur, which overlays the S1 sound. The AS murmur is usually heard best at the aortic valve area. The analysis of PCG signal affected with AS murmur can provide valuable information about the severity of the HVD.

Mitral valve prolapse (MVP) is a common HVD characterized by the movement of mitral valve leaflets towards the left atrium [19]. Figure 1.3(b) shows the visual representation of MVP with the displacement of the mitral valve leaflets. MVP is usually asymptomatic and considered an innocent HM. The clinical signature of MVP in the PCG signal is a systolic click, which occurs after the S1 sound, and a late systolic murmur. Figure 1.4(c) shows a PCG signal affected by MVP murmur. The click is produced by the abrupt tension on the mitral valve leaflets as they billow into the left atrium during systole. The late systolic murmur is due to the turbulent blood flow from the left ventricle into the left atrium through the prolapsed valve. The MVP murmur typically does not affect the S1 or S2 sound. Sometimes, the prolapse of the mitral valve leaflets can cause a slight delay in the S2 sound.

Mitral regurgitation (MR) is characterized by the backflow of blood from the left ventricle to the left atrium during systole [20]. It is caused due to a defective mitral valve. Figure 1.3(c) demonstrates the backflow of blood in MR. The clinical signature of MR in the PCG signal is a systolic murmur with a continuous or plateau shape. Figure 1.4(d) shows a PCG signal affected by MR murmur. The murmur starts after the S1 sound and continues throughout systole. As a result, the S2 sound is not affected in the MR murmur. The murmur produces a characteristic swishing sound, which can be heard best at the apex of the heart. Early diagnosing MR using PCG is essential, as untreated MR can lead to heart failure and other complications.

Mitral stenosis (MS) occurs due to the narrowing of the mitral valve opening [21]. It reduces the blood flow from the left atrium to the left ventricle. Figure 1.3(d) shows the narrowing of the mitral valve orifice in MS. The clinical signature of MS in the PCG signal is a diastolic murmur with a low-pitched rumbling shape. Figure 1.4(e) shows a PCG signal affected by MS murmur. The murmur occurs during diastole when the left ventricle relaxes and fills with blood from the left atrium. In MS, the mitral valve does not open fully, and as a result, the blood flows through the narrowed opening with turbulence, producing the characteristic rumbling sound. The MS murmur typically does not affect the S1 or S2 sound, but it may cause a delay in the A2 component of the S2 sound due to the delayed closure of the mitral valve. The MS murmur is typically heard best at the cardiac apex. The analysis of PCG signal affected with MS murmur can provide useful information about the HVD.

HMs can also be detected based on their severity stages to diagnose HVDs. The Levine scale, also known as the Levine grading system, is used to classify and measure the severity of HMs [22]. It is named after Dr. Stephen S. Levine, a cardiologist who developed this grading system. The Levine scale provides a standardized way to describe the loudness or intensity of an HM. The scale ranges from grade I to grade VI, with each grade representing a different level of loudness and characteristics of the murmur. Here is a brief overview of the Levine grading system:

- Grade I: Very faint and barely audible HM
- Grade II: Soft but readily audible HM
- Grade III: Moderately loud, easily heard HM
- Grade IV: Loud HM and associated with a palpable thrill or vibration felt over the chest.
- Grade V: Very loud HM that can be heard even if the stethoscope is partially lifted from the chest.

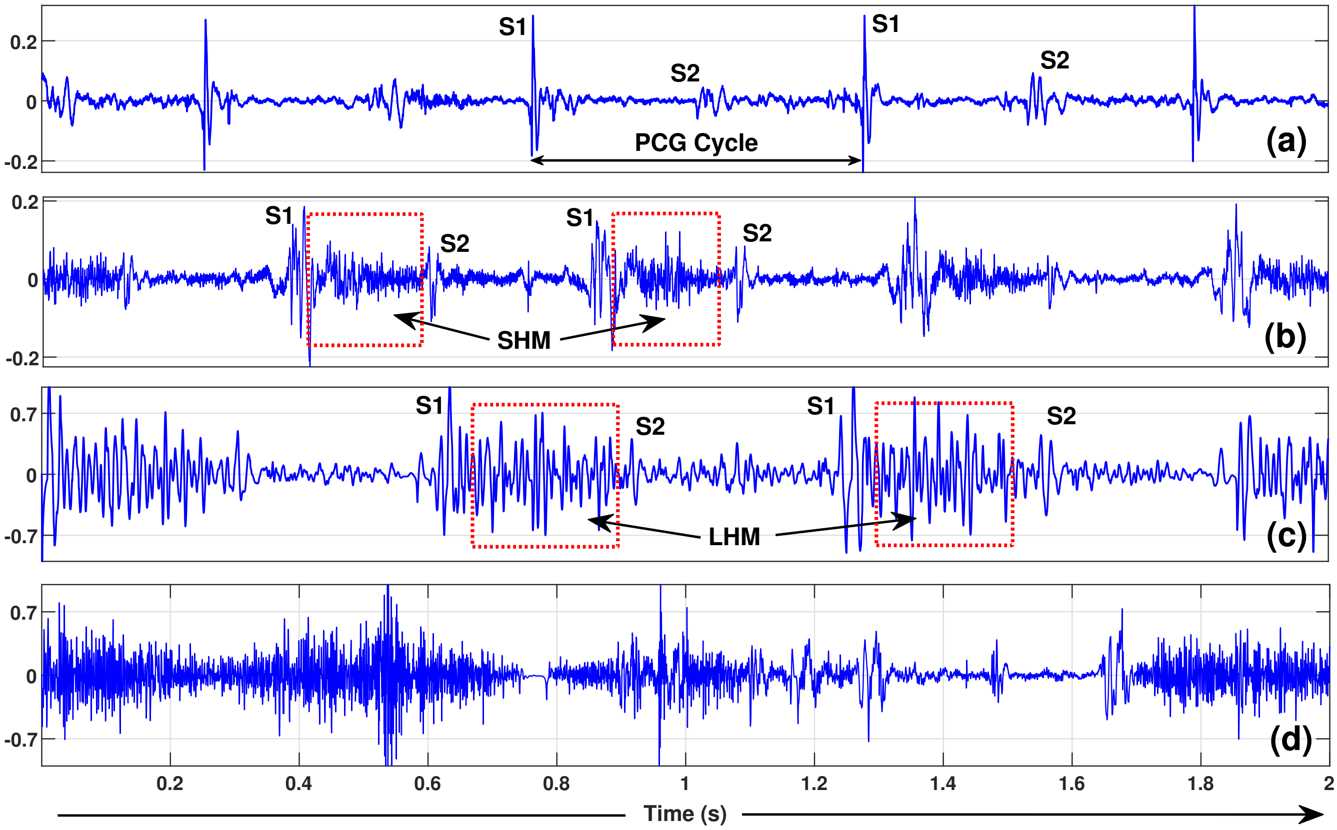

Figure 1.5: Recording of PCG signals. (a) Healthy PCG, (b) Soft heart murmur (SHM) PCG, (c) Loud heart murmur (LHM) PCG, and (d) Low-quality PCG (LQP) recording.

- Grade VI: Loudest HM that can be heard without direct contact with the stethoscope on the chest.

Soft heart murmurs (SHMs) are defined as those with a grading equal to or less than II (grade I and grade II), while loud heart murmurs (LHMs) are defined as those with a grading greater than or equal to III (grade III to grade VI) [22]. Figure 1.5(b) and Figure 1.5(c) show SHM and LHM PCG recordings, respectively. Figure 1.5(a) shows a healthy or absent heart murmur (AHM) PCG for reference. However, the loudness of an HM depends on factors such as background noise, subject anatomy, and the physician's subjective opinion. Poor quality PCG recordings can also affect the accuracy of HM assessment. Figure 1.5(d) shows a low-quality PCG (LQP) recording. Therefore, a robust computer-aided system for automated detection and classification of HMs using PCG is essential to prevent the progression of HVDs. It also benefits patients by assisting cardiologists in providing accurate and timely diagnosis [13].

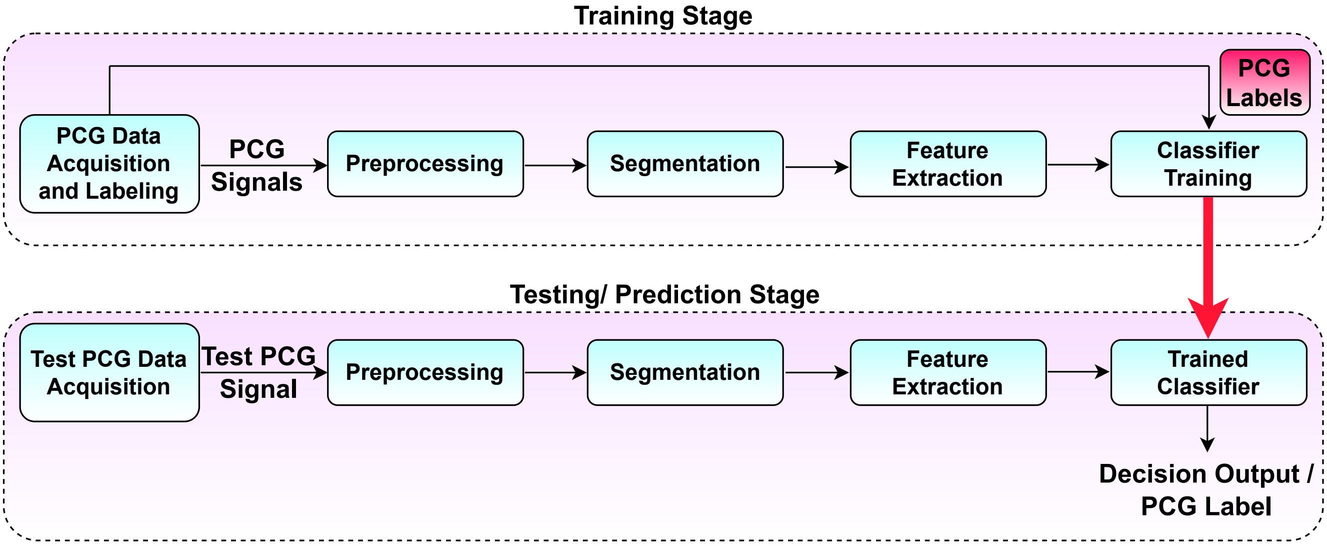

Figure 1.6: Typical framework of a PCG-based automated diagnosis system.

1.3 Automated Diagnosis of HVDs using PCG- A Review

In literature, there has been substantial research for the automated diagnosis of HVDs using PCG signals. Typically, the PCG-based automated diagnosis of HVDs involves four main stages: preprocessing the PCG, segmentation of the PCG, extraction of novel features from the PCG segment, and classification of the PCG segments. Figure 1.6 illustrates the typical framework of a PCG-based automated diagnosis system. In the subsequent sub-sections, we discuss each stage of the diagnosis system.

1.3.1 Preprocessing of the PCG Signal

PCG recordings are often susceptible to various types of noise, such as environmental noise, breathing, and intestinal sounds, which can lead to false detection of HSs. To minimize unwanted signals, researchers have proposed several preprocessing techniques, including bandpass filtering [23], adaptive filtering [24], wavelet transform [25], singular value decomposition (SVD) [26], and empirical mode decomposition (EMD) [27]. Additionally, the friction between the skin and the stethoscope during recording HSs can cause friction spikes in the PCG signal, leading to a low signal-to-noise ratio (SNR). Therefore, a spike removal process is employed to remove these spikes from the filtered PCG signal [28]. As most HSs are recorded using a hand-held electronic stethoscope, these preprocessing techniques are crucial for accurate and reliable PCG-based diagnosis of HVDs.

1.3.2 Segmentation of the PCG Signal

PCG signal segmentation is essential to locate the systolic or diastolic part of the PCG cycle, enabling the classification of pathological HMs in that region. For cyclewise-segmentation, it is required to locate the S1 and S2 sounds from the PCG signal. Locating these two fundamental HSs is challenging due to artifacts and pathological variations in the PCG morphology [29]. Various cyclewise-segmentation methods have been proposed in the literature, including those that utilize the reference of electrocardiogram (ECG) or carotid pulse signals [30, 31], envelope-based methods [32, 33], probabilistic models [28, 29], feature-extraction based methods [34, 35], time-frequency based methods [36, 37], and deep learning models [38, 39]. Due to advanced deep learning (DL) models, researchers have used direct time-interval-based PCG signal segmentation [14, 40]. In this interval-based PCG segmentation, there is no need for locating the S1 and S2 sound. DL models are hypothesized to automatically learn the PCG cycle information from the time-based PCG segments. So, they have avoided the cyclewise-segmentation prior to the feature extraction step. However, this process entirely depends upon the robustness of the DL model. Therefore, whether cyclewise-segmentation prior to feature extraction is necessary or not remains an open question that requires further investigation.

1.3.3 Feature Extraction Methods

Representation of PCG signals in different domains is a useful approach to capture pathological characteristics that aid in improving the detection of HVDs. Time [41–43], frequency [43–46], and time-frequency features have been proposed in the literature for this purpose. Time-frequency methods include, short-time Fourier transform (STFT) [28], continuous wavelet transform (CWT) [47], and discrete wavelet transform (DWT) [48]. Other feature-extraction methods include acoustic features, i.e., Mel-frequency cepstral coefficients (MFCCs) [49] and model-based features [50].

1.3.3.1 Time-domain Features

In the time domain, researchers have mainly extracted features from the pathological bio-markers of the PCG signal [41–43, 51–55]. Chauhan *et al.* [41] computed the duration of S1 sound, systole, S2 sound, and diastole, as well as the average amplitude of S1 and S2 sound to train a hidden Markov model (HMM) for the classification of three types of HMs, i.e., continuous, systolic, and diastolic along with healthy PCG. Whitaker *et al.* [51] have extracted 20 temporal features, including mean and standard deviation of the interval of S1 sound, systole, S2 sound, diastole, and absolute average amplitude ratios. These features were further fed to a support vector machine (SVM) to classify PCG cycles as healthy or pathological. Similarly, the authors in [54] have extracted statistical features, i.e., skewness, kurtosis, and Shannon entropy, to train HMM for detecting pathological PCG cycles. It is worth emphasizing that designing such time-domain features relies on domain expertise. It is also challenging to extract a fixed set of informative features in the time domain to detect HVDs due to the morphological similarity of healthy and pathological PCG signals with noise and artifacts.

1.3.3.2 Frequency-domain Features

In addition to time-domain features, frequency-domain features have been used for the analysis and classification of HVDs [42, 43, 51–53, 56–59]. It is reported that the PCG signals affected with different types of HVDs show high energy concentration in the higher frequency range compared to healthy PCG signals [13]. Various features have been extracted from the spectral domain, including the center of gravity and width of frequency distribution from the spectral envelope of the PCG cycle [52], spectral energy [60], dominant frequency and its magnitude [42], and spectral entropy [42]. Detecting pathological HSs is challenging in the frequency domain due to the overlap of spectral characteristics between HSs and non-physiological events such as respiration and other artifacts. Additionally, the spectral properties of HSs, as illustrated in the spectral properties in Table 1.1, indicate overlapping of S1 and S2 sound. These overlaps further complicate the detection of pathological HSs. Therefore, detecting HVDs in the frequency domain requires careful consideration and analysis to differentiate between physiological and non-physiological events.

1.3.3.3 Time-frequency Domain Features

In earlier works, researchers have extracted time and frequency domain features of PCG signals, like duration and energy of S1 and S2 sound, power spectral density, and bandwidth. The time-domain features are susceptible to noise present in the PCG signal. The frequency-domain features suffer spectral overlap of PCG signal with other physiological events and artifacts.

Therefore, to localize and capture the variations of the PCG signal concurrently on both time and frequency scales, researchers have used various time-frequency domain methods to detect HVDs. These methods include short time Fourier transform (STFT) [41, 44], discrete wavelet transform (DWT) [59], wavelet packet decomposition (WPD) [61], tunable Q-factor wavelet transform (TQWT) [62], wavelet synchro-squeezing transform (WSST) [63], and chriplet transform (CT) [64]. Wang *et al.* have proposed a multi-scale wavelet packet decomposition and extracted entropy features from the PCG signal and applied them to an SVM classifier to identify as normal or abnormal [61]. Patidar *et al.* have used a TQWT method and extracted the sum of average magnitude difference function (SAMDF) features from the PCG cycles [62]. These features are then fed to a least square SVM (LS-SVM) to detect septal defects. Ghosh *et al.* have used a wavelet synchro-squeezing transform to extract magnitude and phase features from the PCG segment and then applied those features to random forest (RF) classifier to detect HVDs [63]. Sawant *et al.* have proposed a fano-factor constrained TQWT to extract 15 time-frequency features from each subband of the PCG signal. Further, these features are fed to a gradient boosting machine model to classify PCG recordings [65]. Ghosh *et al.* have deployed a chriplet transform to perform the time-frequency analysis and extract entropy and energy features and fed them to a composite multi-class classifier for the detection of HVDs [64]. DWT and its variants are mostly used for the automated diagnosis of HVDs using PCG signals. The decomposition of a PCG signal using the DWT method is as follows.

DWT Decomposition Process: To perform DWT decomposition, the input signal $\mathbf{s}[n]$ of length N, is first passed through a high-pass filter $g[n]$ and low-pass filter $h[n]$ to extract high-frequency and low-frequency components of the input signal $\mathbf{s}[n]$, respectively. These two filters are typically designed using the Daubechies, Symlets, Coiflets, or biorthogonal wavelets. After filtering, the signal $\mathbf{s}[n]$ is downsampled by a factor of two, resulting in the detailed coefficients $\mathbf{D}[n]$ and approximation coefficients $\mathbf{A}[n]$. This process is repeated iteratively on the approximation coefficients, resulting in a multilevel decomposition of the signal $\mathbf{s}[n]$. So, the DWT decomposition for level γ is obtained as follows [66].

$$\mathbf{A}_{\gamma}[n] = \sum_{k} h_{\gamma}[n-k] \cdot \mathbf{A}_{\gamma-1}[k] \tag{1.1}$$

$$\mathbf{D}_{\gamma}[n] = \sum_{k} g_{\gamma}[n-k] \cdot \mathbf{A}_{\gamma-1}[k] \tag{1.2}$$

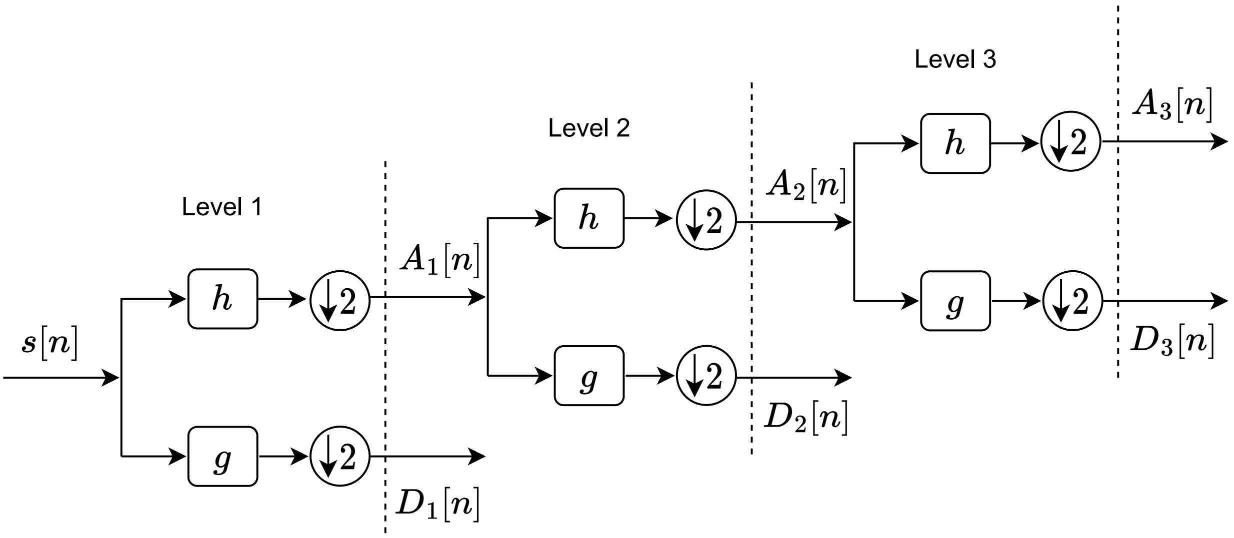

Figure 1.7: The three-level decomposition using discrete wavelet transform (DWT).

Where $\mathbf{A}_\gamma[n]$ and $\mathbf{D}_\gamma[n]$ are the approximation and detail coefficients for γ^{th} level DWT decomposition. Where h_γ and g_γ are the low-pass and high-pass filters for level γ, respectively. DWT does downsample its detail and approximation coefficients for the next level by a factor of 2. Figure 1.7 shows a three-level decomposition of input signal $\mathbf{s}[n]$ using DWT. The total number of subbands for an L- order decomposition is represented as follows:

$$\mathbf{SB}_\gamma[n] = \begin{cases} \mathbf{D}_\gamma[n], & \gamma = 1, 2, ..., L \\ \mathbf{A}_\gamma[n], & \gamma = L + 1 \end{cases} \tag{1.3}$$

where $\mathbf{SB}_\gamma[n]$; $(\gamma = 1, 2, ..., L + 1)$ are the total number of subbands for L- order decomposition. The DWT has several advantages over traditional Fourier-based methods to capture the variations of the PCG signals. The multilevel decomposition of DWT allows the representation of the PCG at different time-frequency characteristics. Further, it helps to extract valuable features from each subband for diagnosing HVDs.

1.3.3.4 Acoustic Features

Mel-frequency cepstral coefficients (MFCCs) have been successfully used for speech analysis by mimicking human auditory perception [67]. It has also shown promising results for acoustic signals like HSs by warping the frequency to a logarithmic scale [41, 44–46, 54, 55, 59, 60]. Son *et*

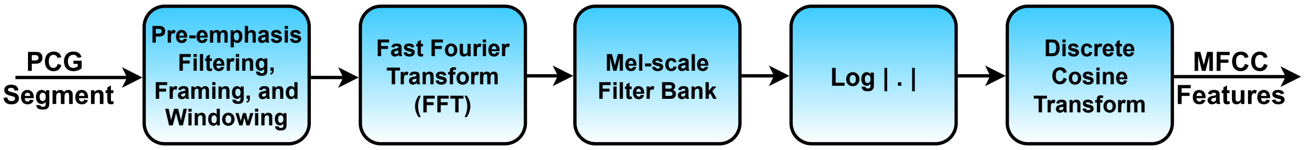

Figure 1.8: Typical steps for MFCC feature extraction from the PCG segment.

al. have extracted 19 MFCC features from the PCG signal to classify five types of HMs [59]. Latif *et al.* have extracted 13 MFCC features from the PCG signal to detect heart valve anomaly [68]. Han *et al.* have fused MFCC features with temporal and statistical features and fed them to a multi-layer perceptron for detection of HVDs using PCG signal [69]. Milani *et al.* have deployed a fusion framework of MFCC and time features and fed it to a supervised classifier for detecting HVDs [70]. Maknickas *et al.* have extracted Mel-frequency spectral coefficients (MFSC), Δ, and Δ^2 features from the PCG signal to classify PCG signals as healthy or pathological [71]. Bozkurt *et al.* have extracted Mel-spectrogram from the PCG signal to classify innocent and pathological HMs [72]. The typical process involves in the extraction of MFCC features is as follows.

MFCC feature extraction: It comprises pre-emphasis filtering, framing, fast Fourier transform (FFT), Mel-scale based filtering, logarithmic transformation, and discrete cosine transformation (DCT) [73]. Figure 1.8 shows the MFCC feature extraction block diagram. First, each PCG segment ($\mathbf{C}[r]$) is passed through a first-order filter with a pre-emphasis coefficient of α. It helps in enhancing high-frequency HS energy contents. Then, each pre-emphasized PCG segment is split into frames. Second, FFT is applied to each frame to obtain the magnitude spectrum. Third, K number of triangular filter banks are used in mel-scale to uniformly divide the entire frequency band of each frame [74]. The Mel-frequency (f_{mel}) is obtained from the linear frequency (f) as follows:

$$f_{\text{mel}} = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{1.4}$$

Fourth, energies are computed for each frame by multiplying each filter bank with its corresponding magnitude spectrum. Finally, DCT is computed from the logarithm of those energies to get the MFCCs for each frame. The formula to obtain the n^{th} coefficient MFCC for t^{th} frame $\mathbf{M}_{n,t}$ is as follows [75]:

$$\mathbf{M}n, t = \sum_{k=1}^{K} (\log \mathbf{F}_{k,t}) \cos \left[\frac{\pi n(k - 0.5)}{K} \right] ; (1 \leq k \leq K) \tag{1.5}$$

where $1 \leq n \leq L$ and L is the desired number of MFCC coefficients. Here, we have considered $L = K$. K is the number of filter banks used for each frame. $\mathbf{F}_{k,t}$ is the k^{th} filter-bank energy of t^{th} frame.

1.3.3.5 Model Based Features

Over the years, attempts have been made to model the PCG signals for HS analysis [50, 76–83]. Joo *et al.* proposed a pole-zero model for porcine prosthetic heart valves in the aortic position for the identification of the valve state [76]. This method provides a better spectral representation of cardiac sounds while lacking time coherence because of the non-stationary behavior of the PCG signal. Similarly, Xu *et al.* proposed a non-linear transient chirp model to extract the aortic and pulmonary components of S2 sound [77]. Zhang *et al.* applied a matching pursuit model to analyze and synthesize normal HSs and murmurs [78]. Other approaches also used a matching pursuit model to diagnose valvular disorders and synthesize the PCG signal [79]. Wavelet-based models are used by Martinez *et al.* [80] and Manikandan *et al.* [81] for compression of the PCG signal. Damped-sinusoidal models are also used in the literature for modeling HSs [50, 82, 83]. Rapid closing of heart valves can be considered as an impulsive excitation in HS and can be reconstructed using a sum of damped sinusoidal. Rasmus *et al.* proposed a windowed sinusoidal model to estimate the S2 sound split [50]. In this field of research, Tang et al. proposed a Gaussian windowed sinusoid model for the telemedicine system [83]. Many parametric and non-parametric models have been used to analyze PCG signals. However, the existing modeling methods have difficulty capturing the morphological patterns of the PCG cycle because of the non-stationary nature of the PCG signal. Most models have relied on modeling short segments of PCG for a specific pathology.

1.3.4 Classification of HVDs

The final stage of the proposed framework involves the classification of HVDs into distinct categories. Two main approaches have been used to achieve this goal: traditional machine learning (ML) and deep learning (DL) methods.

1.3.4.1 Machine Learning (ML) Methods

In the literature, various ML techniques have been applied, such as rule-based classifiers [84], k-nearest neighbors (KNN) [85,86], support vector machine (SVM) [87,88], decision tree (DT) [89], random forest (RF) [90], hidden Markov model (HMM) [41,44], Gaussian mixture model (GMM) [91], naive Bayes theorem [90], and artificial neural network (ANN) [92,93]. In this category, researchers have mainly extracted informative features from the pathological biomarkers of the PCG signals, followed by ML classifiers to detect HVDs. Figure 1.9 (a) shows the typical framework for an ML-based diagnosis system. These ML classifiers have used hand-crafted features from different domains, as discussed in the feature extraction methods 1.3.3, to provide diagnosis decision output. Although these ML methods are easy to train and have improved classification accuracy. They have a few limitations. First, designing distinctive hand-crafted features is highly subjective and depends on domain expertise. Second, these ML methods have generalization and robustness issues, considering the pathological variability in the HVD manifestations.

1.3.4.2 Deep Learning (DL) Methods

In recent years, DL methods have gained significant attention in the field of PCG-based diagnosis systems [94, 95]. These DL methods mainly include deep neural networks (DNNs) [59, 95], recurrent neural networks (RNNs) [68, 96], convolutional neural networks (CNNs) [71, 97], transfer learning models (TLMs) [98, 99], and hybrid models [100, 101]. DL methods for automated HVDs diagnosis can be broadly classified into three categories. The first approach involves a DNN-based method, which follows the classical manual feature extraction and a DNN model to classify HVDs, as depicted in Figure 1.9(a). The second approach is the transfer-domain method, where raw PCG signals or segments are represented in a different domain for compact representation. The transfer-domain signals are then fed to DL models to learn and extract deep features to classify HVDs, as depicted in Figure 1.9(b). The third approach is the end-to-end framework, in which the raw PCG signal is directly fed to a DL model that learns and extracts complex and deep features to classify HVDs, as depicted in Figure 1.9(c). This section provides a detailed discussion of these three approaches.

DNN-based Methods: Various hand-crafted features have been deployed in conjunction with the DNN models for detecting HVDs [59, 69, 70, 94, 95, 102–105]. Table 1.2 shows a brief detail of DNN-based methods in the literature for PCG-based diagnosis of HVDs. For instance, Son

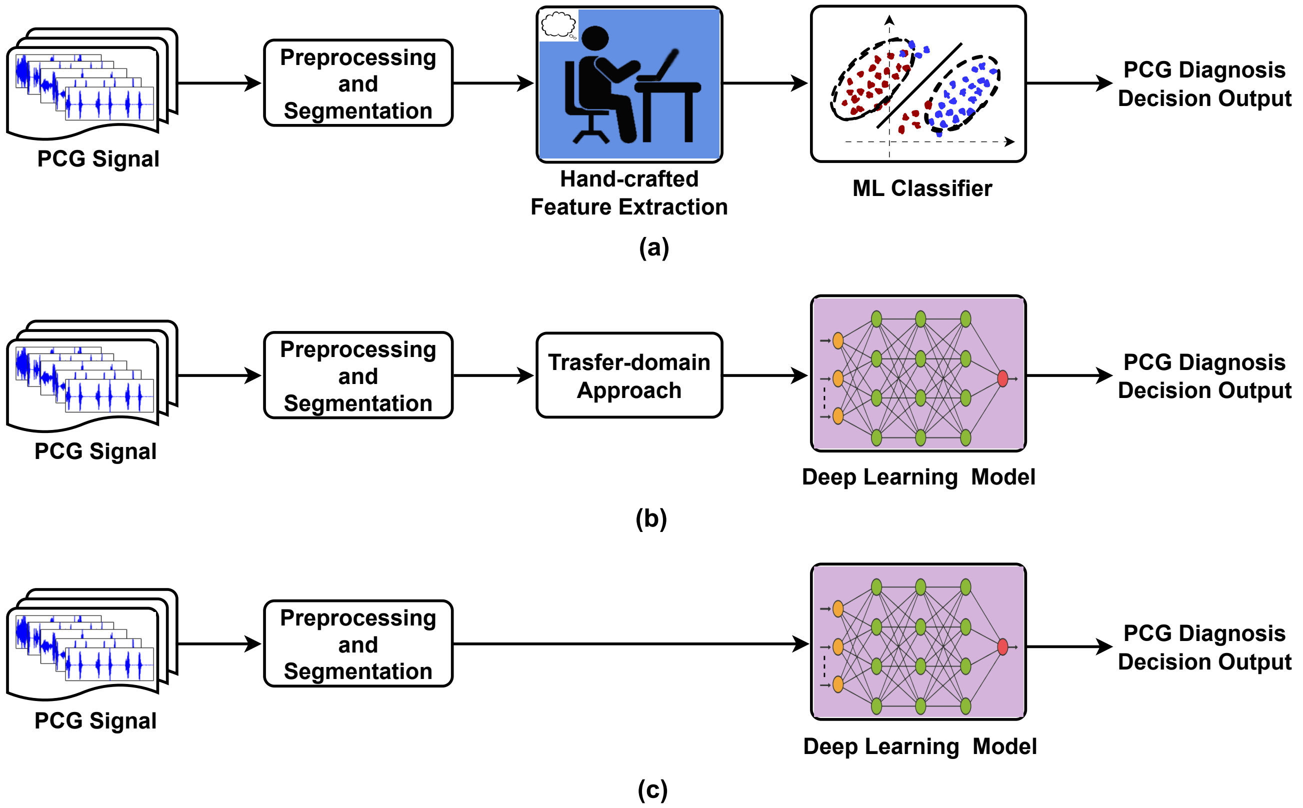

Figure 1.9: Typical frameworks for PCG-based automated diagnosis system. (a) Hand-crafted features with machine learning (ML)-based system, (b) Transfer-domain with deep learning (DL)-based system, and (c) DL-based end-to-end framework system.

et al. employed 19- MFCCs and 24- DWT based features, such as mean, energy, standard deviation, and maximum value of the envelope spectrum in each subband. These features were concatenated and fed to a DNN for multi-class classification of HMs [59]. Millian *et al.* proposed time-domain features, such as amplitude, duration, and intensity of S1 and S2 sounds, and 12-MFCC features from the PCG segment. Further, they were fed to a DNN to classify normal and abnormal HSs [70]. Zabhi *et al.* used 40 time, frequency, and time-frequency features, such as linear prediction coefficients (LPCs), MFCCs, and DWT-based entropy features, for heart sound anomaly detection using PCG signals [95]. Finally, 18 features were selected based on a wrapper feature selection scheme, and an ensemble of 20 feed-forward neural networks was used to detect HSs as normal or abnormal. Although DNNs typically require manual feature extraction, similar to traditional ML-based methods. DNNs can learn complex and abstract representations from the input features. In contrast, traditional ML-based methods may be limited in capturing complex relationships. Moreover, DNNs often require more data and computational resources

Table 1.2: A review of DNN-based methods in the literature for PCG-based diagnosis of HVDs. Acronym: healthy (H), pathological (P), normal (N), aortic stenosis (AS), mitral valve prolapse (MVP), mitral regurgitation (MR), and mitral stenosis (MS).

Method	# Classes	Feature Extraction Details	Database	Result
Chen *et al.*, 2016 [94]	2 (S1, S2)	Acoustic domain (13- MFCCs, 13- Δ, and 13- Δ^2)	Private	OA: 91.12%
Son *et al.*, 2018 [59]	5 (N, AS, MVP, MR, MS)	Time-frequency and Acoustic domain (19- MFCCs and 24- DWT features)	HSM [59]	OA: 89.3%
Han *et al.*, 2019 [69]	2 (H, P)	Acoustic domain (MFCC map)	CICC 2016 [13]	MAcc: 83.42%
Zeng *et al.*, 2021 [103]	2 (H, P)	Time-frequency (Phase space reconstruction features)	CICC 2016 [13]	OA: 97.89%
Milani *et al.*, 2021 [70]	2 (H, P)	Time and Acoustic domain (Amp., duration, intensity, and MFCCs)	CICC 2016 [13]	OA: 93.33%
Das *e al.*, 2022 [104]	2 (H, P)	Time-frequency (Cochleagram features)	CICC 2016 [13]	OA: 98.33%
Zabihi *et al.*, 2016 [95]	2 (H, P)	Time, frequency, and time-frequency (40- features)	CICC 2016 [13]	OA: 91.50%
kay *et al.*, 2016 [105]	2 (H, P)	Time, frequency, time-frequency (70 features)	CICC 2016 [13]	OA: 85.2%

than traditional ML-based methods. Nevertheless, the results obtained from these DNN-based methods indicate that they can potentially improve the accuracy of HVD diagnosis.

Transfer-domain Approaches with DL Methods: One category of DL-based methods for automated diagnosis of HVDs involves the transfer of PCG signals to a different domain for compact representation. These compact representations are then fed to DL models, which typically include CNNs, RNNs, TLMs, or hybrid models. The review of these DL methods for the automated diagnosis of HVDs are as follows.

Researchers have used various transfer-domain approaches using CNN models for the automated diagnosis of HVDs. These models include 2D-CNN [14, 71, 72, 97, 107, 109, 112–115] and 1D-CNN [101, 110, 111]. Table 1.3 presents a brief overview of transfer-domain approaches using CNN models. Deperlioglu *et al.* have used an image of size 100 x 75 pixels obtained from 8s PCG segments, which are then input to a 2D-CNN for classifying normal, murmur, and extrasystole heart sounds [97]. Bozkurt *et al.* have decomposed the PCG segments using gamma tone filter banks and extracted envelopes from each subband. These envelope subbands are logarithmic compressed to form an image, then fed to a 2D-CNN for pathological HS detection [72]. Kui *et al.* have extracted log Mel-frequency spectral coefficients (MFSCs) features from the PCG cycles and input them to a 2D-CNN for classifying as normal or abnormal [114]. Li *et al.* have

Table 1.3: A review of transfer-domain approaches using CNN models for the automated diagnosis of HVDs. Acronym: innocent murmur (IM), pathological murmur (PM), patent ductus arteriosus (PDA), ventricular septal defect (VSD), and atrial septal defect (ASD).

Method	**# Classes**	**Transform-domain Details**	**Database**	**Result**
Maknickas *et al.*, 2017 [71]	2 (H, P)	Acoustic domain (MFSC, Δ, and Δ^2 forms an image of size (3x129x129) pixels)	CICC 2016 [13]	OA: 84.15%
Deperlioglu *et al.*, 2018 [97]	3 (N, Murmur, Extrasystole)	8s PCG segment is converted to an image of size (100x75) pixels	PASCAL- B [106]	OA: 97.9%
Bozkurt *et al.*, 2018 [72]	2 (IM, PM)	Acoustic and Time-frequency domain (Mel-spectrogram, MFCC, and subband envelopes)	Private (UoC-murmur)	ROC: 0.88
Deperlioglu *et al.*, 2019 [107]	2 (H, P)	S1 and S2 sounds image of size (100x75) pixels	CICC 2016 [13]	OA: 97.2%
Li *et al.*, 2019 [108]	2 (H, P)	Time-frequency (Spectrogram Image)	Private	OA: 99.01%
Cheng *et al.*, 2019 [109]	2 (H, P)	Time-frequency (Spectrogram Image)	CICC 2016 [13]	MAcc: 89.50 %
Humayun *et al.*, 2020 [110]	3 (N, mild, moderate)	Time-frequency domain (Four spectral bands decomposed by tconv units filter banks)	Private (Interspeech ComParE)	OA: 81.49%
Li *et al.*, 2020 [111]	2 (H, P)	Time, frequency, and time-frequency domain (497-features)	CICC 2016 [13]	OA: 86.8%
Alqudah *et al.*, 2020 [112]	5 (N, AS, MVP, MR, MS)	Spectral domain (Bi-spectrum Images)	HSM [59]	OA: 98.70%
Chen *et al.*, 2020 [113]	2 (H, P)	Time-frequency (Image using modified frequency WT)	CICC 2016 [13]	MAcc: 95%
Kui *et al.*, 2021 [114]	4 (N, PDA, VSD, ASD)	Acoustic domain (log Mel-frequency spectral coefficients (MFSC) images)	Private	OA: 86.25%
Khan *et al.*, 2021 [115]	3 (N, Murmur, Extrasystole)	Time-frequency (Spectrogram Images)	PASCAL- B [106]	OA: 95.3%
Li *et al.*, 2021 [116]	2 (H, P)	Time-frequency (Spectrogram Images)	CICC 2016 [13]	OA: 85%

performed STFT on the PCG segments and provided the obtained spectrograms as input to a 2D-CNN for automated detection of HSs [116]. Li *et al.* have also extracted spectrograms from PCG segments and employed a denoising autoencoder (DAE) network to obtain feature vectors. These vectors are then fed to a 1D-CNN for classifying HVDs [108]. Overall, these studies have demonstrated the potential of using CNN models for HVD diagnosis. However, the effectiveness of the models may vary depending on the specific dataset and the preprocessing techniques used. It is also important to note that most of the studies reviewed in Table 1.3 focused on the binary classification of HSs as healthy (H) or pathological (P).

Several studies have proposed transfer-domain approaches using RNN models to leverage the temporal information present in the PCG signal for detecting HVDs [45, 68, 96]. Table 1.4 summarizes a brief review of the transfer-domain approaches using RNN models for the automated diagnosis of HVDs. For instance, Latif *et al.* have extracted 13 MFCC features from a segment of 5 PCG cycles and utilized a bidirectional long short-term memory (BLSTM) network to model the

Table 1.4: A review of transfer-domain approaches using RNN models for the automated diagnosis of HVDs. Acronym: Healthy (H) and pathological (P).

Method	# Classes	Transform-domain Details	Database	Result
Latif *et al.*, 2018 [68]	2 (H, P)	Acoustic Domain (MFCCs)	CICC 2016 [13]	MAcc: 98.61%
Zhang *et al.*, 2019 [96]	2 (H, P)	Time-frequency (Temporal quasi-periodic features from different frequency bands of the spectrogram)	CICC 2016 [13]	OA: 94.84%
Khan *et al.*, 2020 [45]	2 (H, P)	Acoustic domain (MFCCs)	CICC 2016 [13]	AUC: 91.39%

Table 1.5: A review of transfer-domain approaches using TLM network for the automated diagnosis of HVDs. Acronym: Healthy (H), pathological (P), and coronary artery disease (CAD).

Method	# Classes	Transform-domain Details	TLM Network	Database	Result
Dominguez *et al.*, 2017 [117]	2 (H, P)	Time-frequency (Gray-scale Sonogram Image)	Modified AlexNet	CICC 2016 [13]	OA: 97%
Wu *et al.*, 2019 [118]	2 (H, P)	Time-frequency and Acoustic domain (Spectrogram, Mel-spectrogram, and MFCCs)	3- VGGNets	CICC 2016 [13]	MAcc: 89.81%
Alaskar *et al.*, 2019 [119]	2 (H, P)	Time-frequency (Scalogram Image)	AlexNet	CICC 2016 [13]	OA: 87%.
Dhar *et al.*, 2021 [98]	2 (H, P)	Time-frequency (Cross-wavelet transform to have an Image)	AlexNet	CICC 2016 [13]	OA: 98%
Pathak *et al.*, 2022 [99]	2 (H, CAD)	Time-frequency (Spectrogram Image)	VGG16	Private	OA: 91.19%

temporal variation in the PCG segment for detecting abnormal HSs [68]. However, the number of studies in this area is relatively small compared to the transfer-domain approaches using CNN models. Further research is needed to explore the potential of RNN models in detecting HVDs.

Several studies have proposed transfer-domain approaches using TLM networks to detect HVDs from PCG signals [99, 117–119]. Table 1.5 summarizes a brief review of the transfer-domain approaches using TLM networks for the automated diagnosis of HVDs. Dhar *et al.* utilized cross-wavelet transform spectrum features extracted from PCG signals and fed them into a pre-trained AlexNet CNN model for HVD detection [98]. Pathak *et al.* converted PCG segments into PCG spectrum and fed them into a pre-trained VGG-16 network for automated detection of coronary artery diseases (CADs) [99]. Transfer-domain approaches using TLM networks show promising results in detecting HVDs, as demonstrated by the studies summarized in Table 1.5. However, further investigations are required to assess the generalizability of these models in detecting HVDs using PCG signals.

In recent years, several transfer-domain approaches using hybrid deep learning models have been proposed to detect heart valve anomalies [40, 100, 120, 122–125]. Table 1.6 summarizes a

Table 1.6: A review of transfer-domain approaches using hybrid models for the automated diagnosis of HVDs. Acronym: Healthy (H) and pathological (P), and coronary artery disease (CAD).

Method	# Classes	Transform-domain Details	Hybrid Model	Database	Result
Potes *et al.*, 2016 [100]	2 (H, P)	Time-frequency (124- features and four frequency subbands	CNN-DNN	CICC 2016 [13]	OA: 86.02%
Deng *et al.*, 2020 [120]	2 (H, P)	Acoustic domain (13- MFCCs, 13- Δ, 13- Δ^2)	CRNN with DNN	CICC 2016 [13]	OA: 98%
Li *et al.*, 2020 [121]	2 (CAD, Non-CAD)	Time and Acoustic (Hand-crafted features and Mel-spectrogram Image)	CNN-DNN	Private	OA: 90.43%
Gjoreski *et al.*, 2020 [122]	2 (H, P)	Time and Time-frequency (Hand-crafted features and spectrogram Image)	CNN and RF	CICC 2016 [13]	OA: 89%
Dissanayake *et al.*, 2020 [40]	2 (H, P)	Acoustic domain (MFCCs, Δ, and Δ^2)	CNN and LSTM	CICC 2016 [13]	OA: 99.78%
Deperlioglu *et al.*, 2021 [123]	3 (H, Murmur, Extrasystole)	Time domain (Instant energy)	Auto-encoder	PASCAL-B [106]	OA: 99.61%
Huang *et al.*, 2022 [124]	2 (CAD and Non-CAD)	Acoustic (MFCC spectrograms)	CNN-LSTM-RF	Private	OA: 96.05%

brief review of the transfer-domain approaches using hybrid models for the automated diagnosis of HVDs. Deng *et al.* have developed a hybrid model by combining a 2D-CNN, LSTM, and fully connected (FC) layer to classify HSs [120]. They extracted 13-MFCCs, 13-Δ, and 13-Δ^2 features from the PCG signal and fed them to the CNN. The output of the CNN is then given to the LSTM network for HS temporal feature learning, and the encoded vector is fed to the FC layer for classification. Li *et al.* proposed a multi-domain feature and deep learning feature-based model to detect coronary artery disease [101]. They extracted 110 features related to duration, energy, higher-order statistics, and spectral and wavelet-based features from the PCG signal. Further, they extracted MFCC images from the PCG signal and fed them to a CNN network. Finally, they fused both features using a dense layer to detect CAD.

Overall, these transfer-domain approaches using DL models have the potential to revolutionize the diagnosis of HVDs by providing an automated, accurate, and efficient diagnosis.

End-to-End Framework with DL Methods: The final category of DL-based methods involves the end-to-end framework for diagnosing HVDs. In this category, raw PCG signals are fed directly to various DL models, such as CNNs, RNNs, or hybrid models, to classify HVDs. Table 1.7 summarizes a brief review of the end-to-end framework with DL methods for automated HVDs diagnosis. For instance, Baghel *et al.* utilized raw 2.5s PCG segments and fed them to a 1D-CNN to diagnose multiple cardiac diseases [126]. Alkhodari *et al.* fed a 1.2s PCG segment to a combination of CNN and BLSTM to detect HVDs [127]. Shuvo *et al.* proposed the CRNN-based

Table 1.7: A review of end-to-end framework with DL methods for the automated diagnosis of HVDs. Acronym: Healthy (H), pathological (P), and congenital heart disease (CHD).

Method	# Classes	DL Model	Database	Result
Xiao *et al.*, 2019 [131]	2 (H, P)	Two 1D- CNNs	Private	OA: 93.56%
Baghel *et al.*, 2020 [126]	5 (N, AS, MVP, MR, MS)	1D- CNN	HSM [59]	OA: 98.60%
Oh *et al.*, 2020 [129]	5 (N, AS, MVP, MR, MS)	1D- CNN with Residual	HSM [59]	OA: 97%
Xiao *et al.*, 2020 [130]	2 (H, P)	1D- CNN	CICC 2016 [13]	OA: 93%
Avanzato *et al.*, 2020 [132]	5 (N, AS, MVP, MR, MS)	1D- CNN	HSM [59]	OA: 89.6%
Krishnan *et al.*, 2020 [133]	2 (H, P)	1D- CNN	CICC 2016 [13]	OA: 85.74%
Raza *et al.*, 2019 [134]	3 (H, Murmur, Extrasystole)	LSTM	PASCAL- B [106]	OA: 80.80%
Alkhodari *et al.*, 2021 [127]	5 (N, AS, MVP, MR, MS)	1D- CNN with BLSTM	HSM [59]	OA: 97.87%
Liu *et al.*, 2022 [135]	2 (CHD, Non-CHD)	1D- CNN with RNN	Private	OA: 96.7%
Shuvo *et al.*, 2021 [128]	5 (N, AS, MVP, MR, MS)	1D- CNNs with BLSTM	HSM [59]	OA: 99.60%
Li *et al.*, 2021 [14]	2 (H, P)	1D- CNN with GRU	CICC 2016 [13]	OA: 95.50%

CardioXNet architecture for the automated detection of HVDs [128]. Other studies have also utilized end-to-end frameworks with raw PCG signals, including CNNs [129–133], RNNs [134], and hybrid models [14, 135] for detecting HVDs. So, the end-to-end DL-based framework for the automated diagnosis of HVDs has shown promising results. These end-to-end framework approaches eliminate the need for hand-crafted features, which can be time-consuming and prone to errors. Furthermore, end-to-end frameworks enable the DL models to learn informative features directly from the PCG signals, improving diagnostic accuracy.

1.4 Motivation of the Research Work

Although several methods have been proposed for the automated diagnosis of HVDs using PCG, a few research challenges have not been addressed as follows.

- As discussed earlier, the pathological manifestation of HVDs in a PCG cycle has diverse temporal and morphological variations. Due to artifacts and pathological variations in PCG signals, existing time [41, 51] and frequency [42, 60] domain feature extraction approaches encounter difficulties in detecting the HVDs. To automatically detect HVDs require analysis of a complete PCG cycle. Existing modeling methods have difficulty to model a PCG cycle due to the non-stationary nature of the PCG signal [76, 77]. These methods focused on modeling a short segment of the PCG cycle, i.e., only S1 or S2 sound [50]. However, analyzing a particular segment may not be generalized well to unseen HVDs. Therefore, we hypothesize that modeling a PCG cycle can provide better generalization and robustness for diagnosing

HVDs.

- Heart murmurs (HMs) have a diverse etiology, resulting in temporal variability in PCG signals. Feature-fusion-based methods lack the effective integration of complementary information for HMs detection [70, 121]. Additionally, finding a fixed set of hand-crafted features that can adequately train an ML model is challenging due to the variability in disease manifestations. The existing ML and DL-based methods have limitations in capturing the temporal dependencies in the PCG signal associated with the HMs [59, 63]. Therefore, an effective feature-fusion algorithm and a DL model that leverages temporal information are needed to enhance HM detection.

- Existing wavelet transform (WT) based methods used for HMs detection suffer from resolution loss at higher subbands (coarse scales) and lack shift-invariant properties [59, 61, 62, 72]. As a result, crucial information related to HMs is often missing in the coarse scales. Moreover, existing wavelet approaches extract features from wavelet subbands and employ an ML classifier to detect HMs. However, an HVD pathology shows temporal variations within and across different subbands. So, effectively handling these subbands' temporal information is essential for the robust detection of HMs. Therefore, there is a need for a WT that adequately captures PCG patterns in coarse scales. Further, there is a need for a DL model to utilize the temporal information across each subband and extract relevant clinical information to classify HMs.

- Existing automated methods for diagnosing HVDs mainly focus on the binary classification of PCG signals as healthy or pathological. A few studies have attempted to perform multi-class analysis of HMs [59]; however, small databases have limited them. Furthermore, most DL methods for HMs analysis represent the PCG signal to transfer domain and fed to a DL model to classify HMs [14, 71, 114, 115]. Therefore, there is a need for a robust end-to-end framework that can automate the analysis of PCG signals without manual intervention. Such a framework will be convenient for use in primary healthcare to assist experts in diagnosing HVDs.

1.5 Contributions of the Work

Motivated by the above issues, the thesis aims to develop automated diagnosis methods using PCG for reliable and efficient HVDs detection. The contributions of the thesis are as follows.

- A multi-component oscillatory model has been investigated for PCG signals for the automated diagnosis of HVDs. The proposed model captures and accurately represents the PCG cycle's morphological variations by fitting a half-period sine wave between two zero-crossing points. The extracted model parameters, i.e., amplitude and frequency, provide discriminative features to a DNN classifier for detecting PCG cycles as healthy or pathological. Our proposed method is computationally efficient and easy to implement, requiring only the calculation of zero-crossing points. We evaluate the efficacy of the proposed method using a publicly available challenge database. The proposed method achieves competitive results as compared to the existing state-of-the-art methods.

- A novel acoustic feature-fusion-based hierarchical LSTM (HLSTM) network is proposed to classify HMs. First, MFCCs and linear prediction cepstral coefficients (LPCCs) features are extracted from the PCG signals. Then, these two features are fused in two novel ways using a hierarchical LSTM (HLSTM) network with a self-attention module. The first model is the series feature-fusion HLSTM (SFF-HLSTM) model. The second model is the parallel feature-fusion HLSTM (PFF-HLSTM) model. HLSTM encodes the temporal information of the fused features in different abstractions. Then, the self-attention module aggregates these encoded vectors based on their clinical relevance to detect HMs. The efficacy of the proposed method is evaluated using two publicly available databases. The proposed method exhibits improved performance compared to the existing state-of-the-art methods.

- A stationary wavelet transform (SWT) based PCG signal decomposition is proposed for adequately capturing PCG patterns in coarse scales to detect HMs. We employ subband-specific HLSTM networks to utilize each SWT subband's temporal and relative temporal information. Then, the encoded vectors of each subband-specific HLSTM network are fed to their respective intra-subband attention layer for weight aggregation based on their clinical relevance. Further, the inter-subband attention layer weight aggregates these subband-specific representations to improve the detection of HMs. The proposed method is tested and validated using two open-access databases. The proposed method shows competitive performance with the existing baseline methods. The proposed method's impressive performance and generalization can help to detect HVD anomalies during preliminary check-ups in healthcare units.

- A novel multi-kernel residual convolutional neural network (MK-RCNN) model has been proposed for classifying the severity stages of HMs. The model utilizes multi-kernel CNN-based feature extractors to extract scale-specific features from the raw PCG segments.

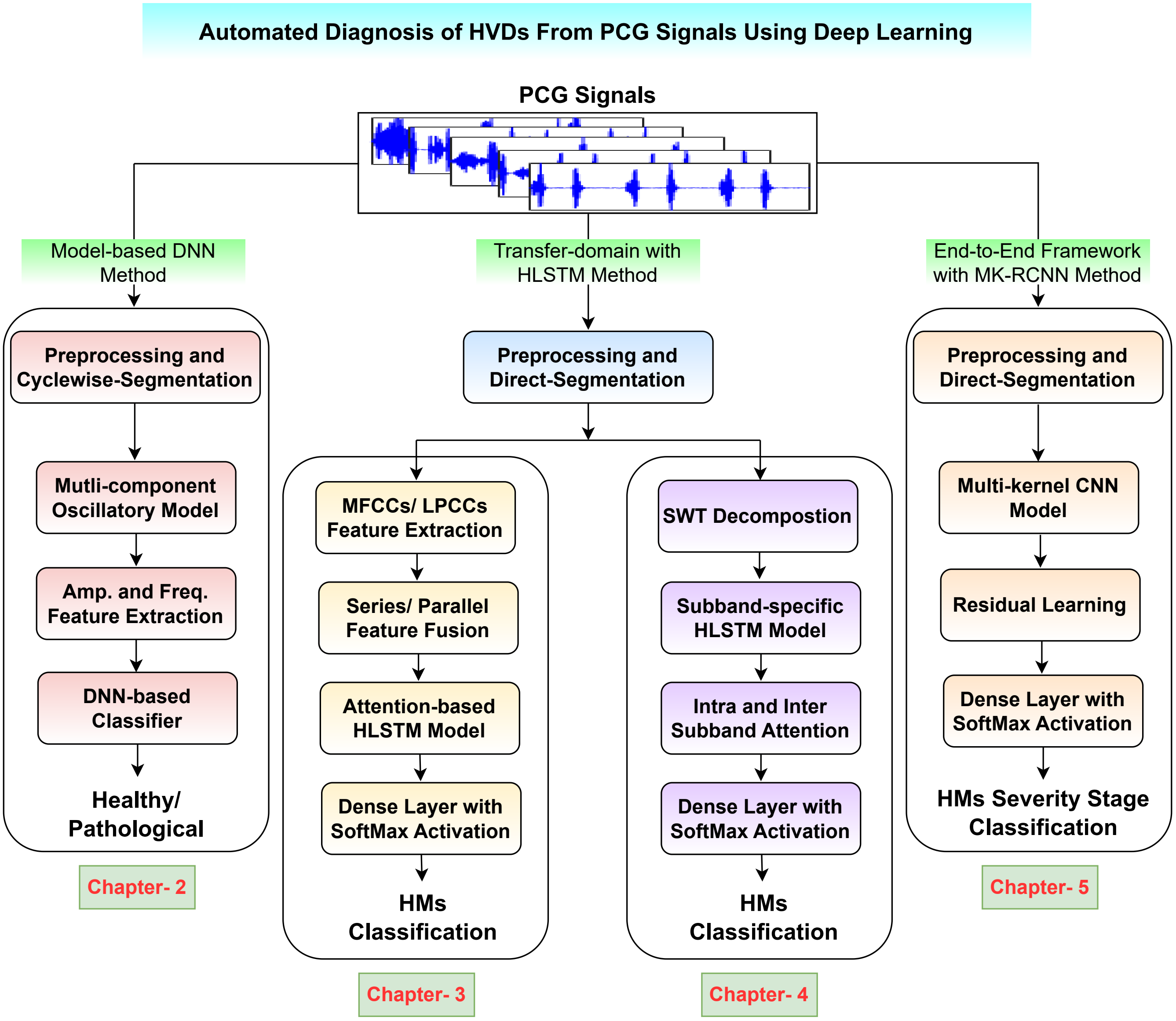

Figure 1.10: Graphical representation of the working chapters of this dissertation.

Furthermore, using residual learning helps to extract deep features from the CNN layers without degrading the performance accuracy. The proposed end-to-end framework eliminates manual interventions, which can be time-consuming and prone to errors. The results show that the proposed MK-RCNN model outperforms the state-of-the-art methods for HM severity stages classification. The proposed method can be a reliable tool for the automated diagnosis of HMs and provide better patient treatment options.

1.6 Organization of the Thesis

The thesis is organized into three working chapters, as illustrated in Figure 1.10. **Chapter** 2 proposes a multi-component oscillatory model to classify PCG cycles as healthy or pathological. The model extracts model parameters, i.e., amplitude and frequency, by fitting a half-period sine wave function between two consecutive zero-crossing points of the PCG cycle. The extracted parameters are then fed to a DNN classifier for HVDs detection. **Chapter** 3 presents a novel acoustic feature-fusion-based hierarchical LSTM (HLSTM) network for HMs classification. MFCCs and LPCCs features are extracted from PCG signals. Then, they are fused using a hierarchical LSTM (HLSTM) network with a self-attention module. The HLSTM captures temporal information across different abstractions, while the self-attention module aggregates encoded vectors based on clinical relevance to detect HMs. **Chapter** 4 introduces a novel approach for HMs detection using a stationary wavelet transform (SWT) based PCG signal decomposition. The SWT aims to capture PCG patterns effectively in coarse scales. Subband-specific hierarchical HLSTM networks leverage the temporal information within each SWT subband. Based on their clinical relevance, encoded vectors from subband-specific HLSTM networks are then fed into intra-subband attention layers for weight aggregation. Further, an inter-subband attention layer aggregates these subband-specific representations to enhance HM detection. **Chapter** 5 presents a multi-kernel residual convolutional network (MK-RCNN) model-based end-to-end framework for the automated detection of HM severity stages. The multi-kernel CNN extracts scale-specific features from the raw PCG segments, while residual learning in the CNN layers facilitates extracting deep features without performance degradation. Finally, **Chapter** 6 summarizes the research work and highlights some future directions.

2

Multi-component Oscillatory Model Based Detection of Pathological HSs

Contents

In the previous chapter 1, we discussed the genesis of different HVDs and their morphological variations observed in a PCG signal. Different types of HVDs produce different patterns in the PCG cycle, which are subtle and difficult to differentiate. In addition, there may be variations in the PCG morphology due to the same type of HVD, depending on factors such as the disease's severity, the patient's age, and other underlying heart conditions.

The existing time-domain features to detect HVDs show morphological similarity between healthy and pathological PCG signals [41, 51]. Detecting pathological HSs is challenging in the frequency domain due to the overlap of spectral characteristics between HSs and non-physiological events such as respiration and other artifacts [42, 60]. Numerous modeling methods have been reported in the literature which extract novel features for the analysis and classification of HSs as healthy or pathological [136]. These methods include pole-zero models [56, 76], chirp signal models [77], matching pursuit models [78, 79], wavelet models [80], damped sinusoidal models [50, 82, 137, 138]. However, these models have difficulty capturing the morphological patterns of the PCG cycle because of the non-stationary nature of the PCG signal. Most existing models have focused on modeling a short segment of the PCG cycle, i.e., only S1 or S2 sound. Whereas analyzing S1 or S2 sound, or only the systolic or diastolic part, may not be sufficient to diagnose HVDs. A complete analysis of the PCG cycle, including the timing, duration, and intensity of each component, is necessary to diagnose HVDs accurately. The PCG cycle provides valuable information about the functioning of the heart valves and can help identify abnormalities that may indicate the presence of an HVD. Therefore, we hypothesize that modeling a PCG cycle can provide a better generalization to analyze different types of HVDs and will provide a more robust performance to classify the PCG cycles as healthy or pathological.

This chapter proposes a novel multi-component oscillatory model to detect HVDs. The model captures the morphological variations of the PCG cycle by fitting a half-period sine function between consecutive zero-crossing points. The model parameters, i.e., amplitude and frequency, provide discriminative features to the DNN classifier to detect the PCG cycles as healthy or pathological. The proposed method is tested and validated using the Physionet challenge 2016 database [13].

The rest of the chapter is organized as follows: Section 2.1 discusses the proposed multi-component oscillatory model-based method to classify HSs. Section 2.2 presents the results and discussion of the proposed method. Section 2.3 ends with the summary.

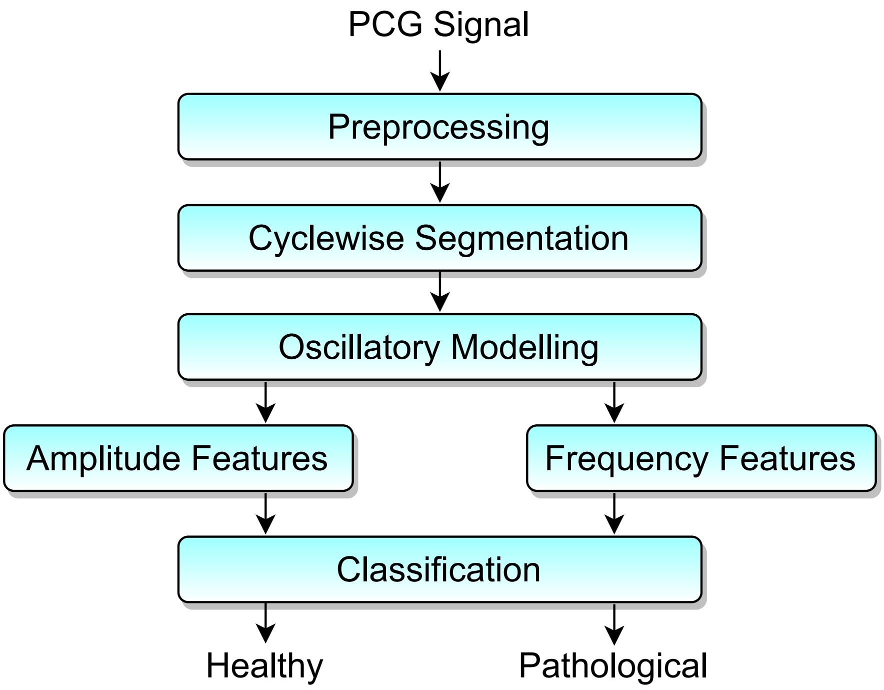

Figure 2.1: The schematic outline of the proposed modeling-based method to classify HSs.

2.1 Multi-component Oscillatory Model Based Method to Classify HSs

This section presents the proposed modeling method to detect pathological HSs using PCG signals. It consists of four parts: preprocessing the raw PCG signal, segmentation into PCG cycles, proposed model-based feature extraction, and detecting PCG cycles as healthy or pathological. Figure 2.1 shows the block diagram of the proposed method.

2.1.1 Preprocessing of the PCG signal

The presence of unwanted signals such as lung sound, diaphragm movement, and surrounded speech affects the PCG signal recording [138]. It leads to false detection or miss of HVDs in the PCG signal. These noises also create a challenge to segment the PCG signal. Therefore, it is necessary to eliminate these noises before segmentation. So, the raw PCG signal is filtered using a sixth-order Butterworth bandpass filter having a lower and higher cut-off frequency of 20 Hz and 900 Hz, respectively. Figure 2.2(a) and Figure 2.2(b) show the raw and filtered PCG signals, respectively. In most clinical setups, HSs are generally recorded using a hand-held electronic stethoscope. So, friction spikes can be generated in the PCG signal due to the movement between the skin and the stethoscope. It leads to a low signal-to-noise ratio. Therefore, these spikes are

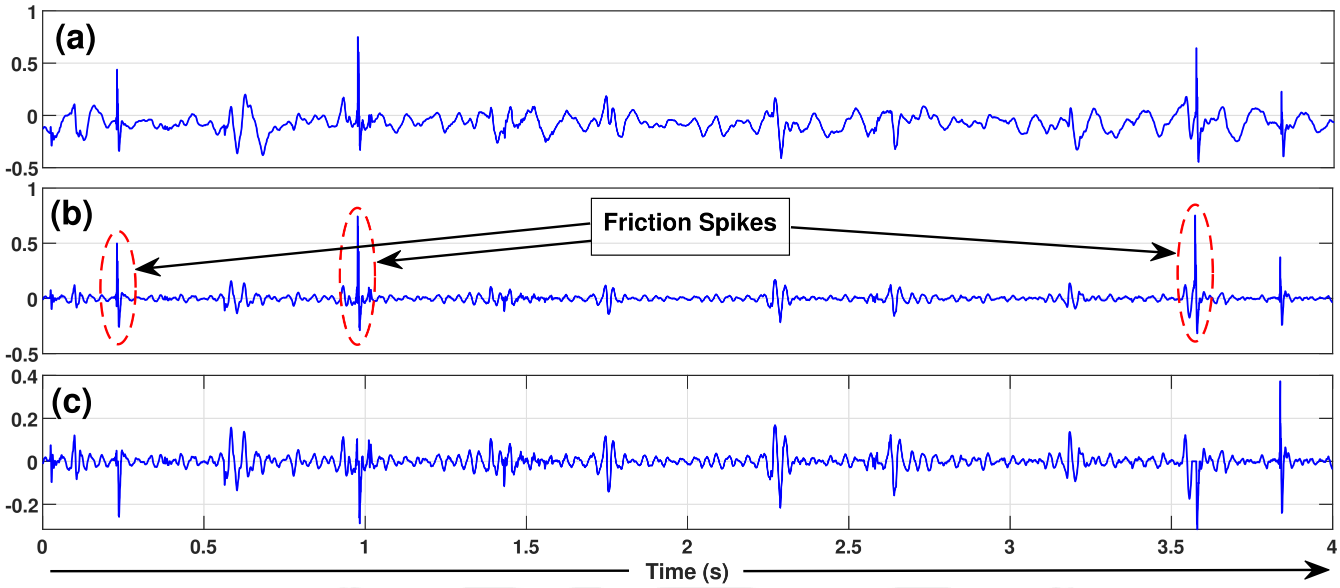

Figure 2.2: Preprocessing of the PCG signal. (a) Raw PCG signal, (b) Filtered PCG signal, and (c) Spike-removed filtered PCG signal.

further removed from the filtered PCG signal using a spike removal process proposed by Schmidt *et al.* [28]. Figure 2.2(c) shows the spike-removed filtered PCG signal. The steps involved in removing the friction spikes using Schmidt's algorithm are as follows:

(i) Each filtered PCG signal is divided using a 500 ms rectangular window.

(ii) The maximum absolute amplitude (MAA) is computed from each window.

(iii) Then, the median value of MAA's from all the windows is computed. If any window MAA exceeds three times the median MAA value, follow these sub-steps. Otherwise, go to step (iv) directly.

  (a) First, the window with the highest MAA value is considered.

  (b) In that particular window, the top of the noise spike is chosen as the location of the MAA.

  (c) The first zero-crossing before and after that location is selected.

  (d) All the sample points between these two zero-crossings are replaced by zeros.

  (e) Go to step (ii).

(iv) The process is completed.

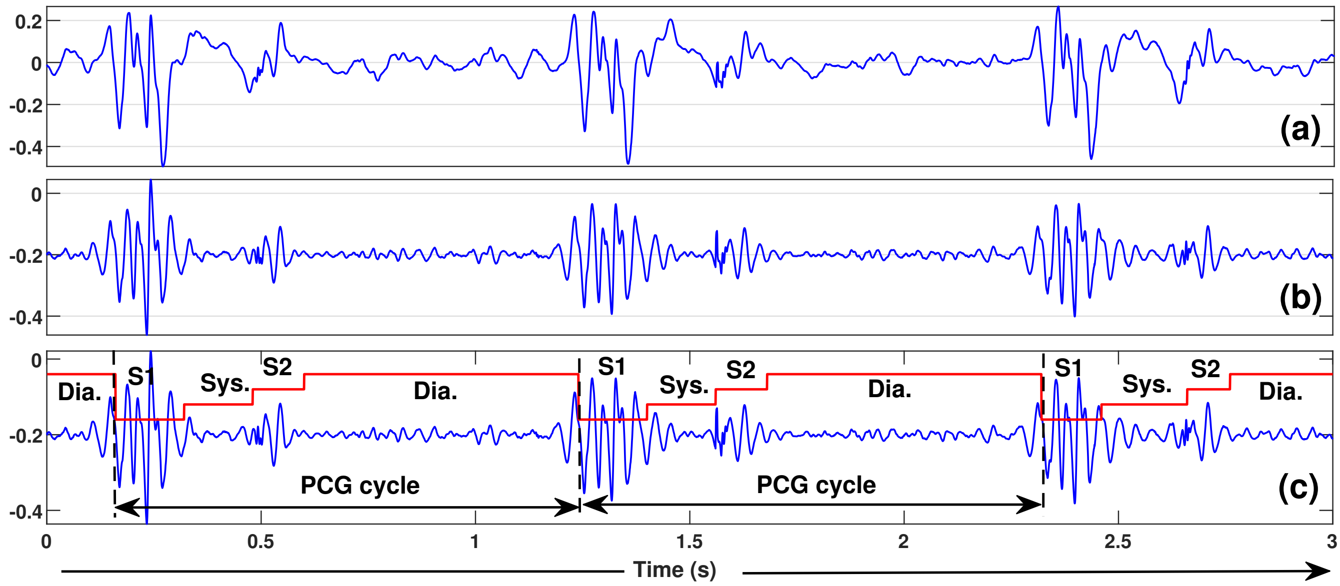

Figure 2.3: Preprocessing and segmentation of a raw PCG signal. (a) Raw PCG signal, (b) Preprocessed PCG signal, and (c) Segmented PCG cycles with their four HS states, namely, S1, systole, S2, and diastole.

2.1.2 Segmentation into PCG Cycles

The PCG signal segmentation is an essential step to locate a particular region, i.e., the systolic or diastolic part of the PCG cycle, and subsequently, classify the pathological murmurs in that region. Thus, the preprocessed PCG recordings are segmented cyclewise. In this work, we deployed the logistic regression-hidden semi-Markov models (LR-HSMM) method to identify PCG cycles [29]. The model is a first-order nonergodic hidden Markov model (HMM) with four hidden states, i.e., S1, systole, S2, and diastole. The model incorporates the duration distribution information of each state to enhance the PCG segmentation. Therefore, the model is called the hidden semi-Markov model (HSMM) and is defined as

$$\lambda = (\mathsf{A}, \mathsf{B}, \pi, p) \tag{2.1}$$

where λ is the model, A is the transition probability matrix between hidden states, B is the distribution of observation sequence, and p is the state duration probability.

The HSMM model uses logistic regression (LR) to estimate observation probability compared to the Gamma, or Gaussian distribution, reported in earlier work [28, 139]. Incorporating LR allows more significant discrimination between states, i.e., S1, systole, S2, and diastole. Logistic regression is a binary classifier that provides a binary response based upon feature space or predictor variable using a logistic function, computed by the following equation [140].

$$\sigma(x) = \frac{1}{1+\exp(-x)} \tag{2.2}$$

The probability of a particular state at time instant t, given the input observation (O_t), can be calculated by using the above equation as:

$$P[q_t = \zeta_j | O_t] = \sigma(\omega^{'} O_t) \tag{2.3}$$

where ω is the weight vector applied to each input observation. It is trained iteratively using the re-weighted least square method. Finally, the probability of observation at time instant t given the state $(q_t = \zeta_j)$ is calculated using Bayes' rule as:

$$b_j(O_t) = P[O_t | q_t = \zeta_j] = \frac{P[q_t = \zeta_j | O_t] \times P(O_t)}{P(\zeta_j)} \tag{2.4}$$

where $P(O_t)$ is obtained from the features of the training dataset by using a multivariate normal (MVN) distribution. $P(\zeta_j)$ is the initial state probability of the j^{th} state. The LR-HSMM method uses four signal-derived features, i.e., wavelet envelope, homomorphic envelope, Hilbert envelope, and power spectral density envelope, as the observed sequences. For the evaluation of the LR-HSMM model, R-peaks and the end of T-waves of the ECG signal are used as the reference to locate the S1, and S2 sounds, respectively. Figure 2.3(a) and Figure 2.3(b) show a raw PCG recording, the preprocessed PCG recording. Figure 2.3(c) shows the segmentation of the PCG recording. Each cycle-wise PCG segment has its respective four HS states, i.e., S1 sound, systole, S2 sound, and diastole.

2.1.3 Proposed multi-component oscillatory model

We have proposed a multi-component oscillatory model to extract the model parameters from the PCG cycles and their respective four HS states. Further, these model parameters are deployed to classify the PCG cycles and the four HS states as healthy or pathological.

2.1.3.1 Extraction of model parameters

The flow diagram to extract model parameters using the proposed multi-component oscillatory model is shown in Figure 2.4. First, the PCG cycle $(\mathbf{S}^1)$ is fed as input to extract first-order model parameters. Hereafter, the residual signal is fed as input to extract the following order model

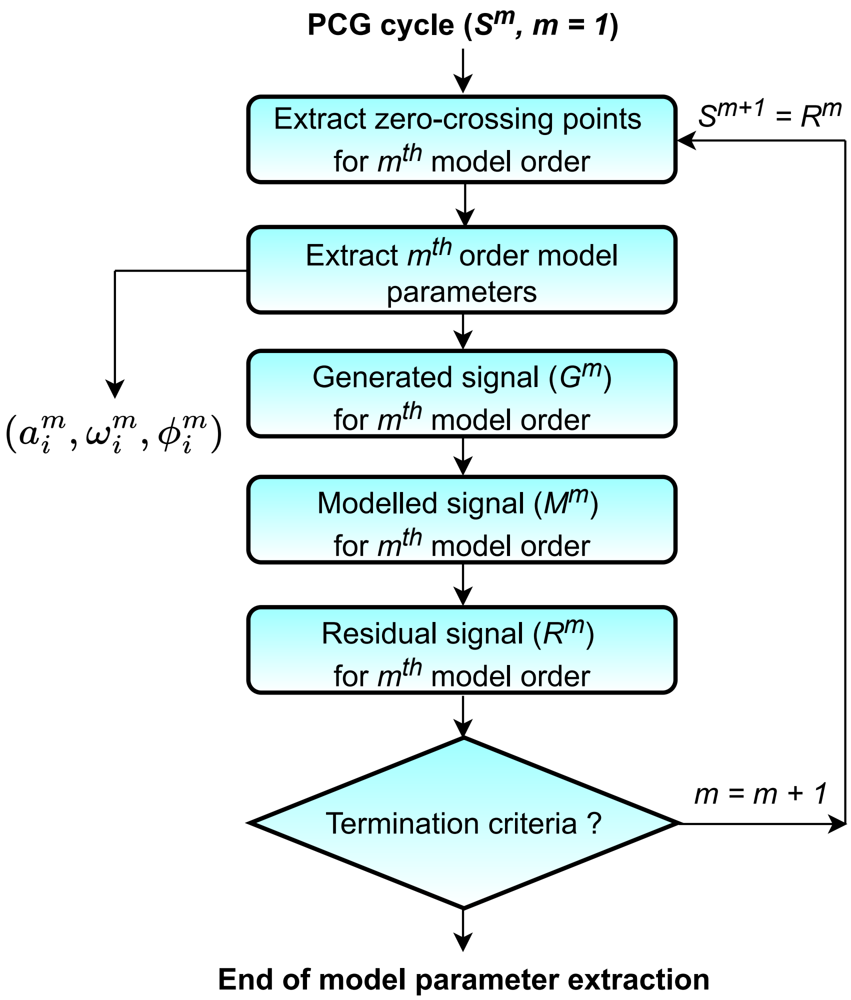

Figure 2.4: Extraction of model parameters using the proposed multi-component oscillatory model.

parameters till the termination criteria is satisfied. Here, every iteration corresponds to a model order.

First, the zero-crossing points are computed from the input signal at each iteration. Let $\mathbf{S}^m$ be the input signal for m^{th} iteration and t_1^m, t_2^m, t_3^m, ..., t_Z^m are its successive zero-crossing instants. Where Z is the total number of zero-crossing points of $\mathbf{S}^m$. The model parameters for the m^{th} order are obtained by fitting a half period of sine wave function between each successive zero-crossing point of the input signal (S^m). A half-period sine wave function is represented by its amplitude, frequency, and phase, which forms a parameter-set. Hence, the total parameter-sets for the m^{th} order is $Z - 1$, which is one less than the number of zero-crossing points (Z). A parameter-set

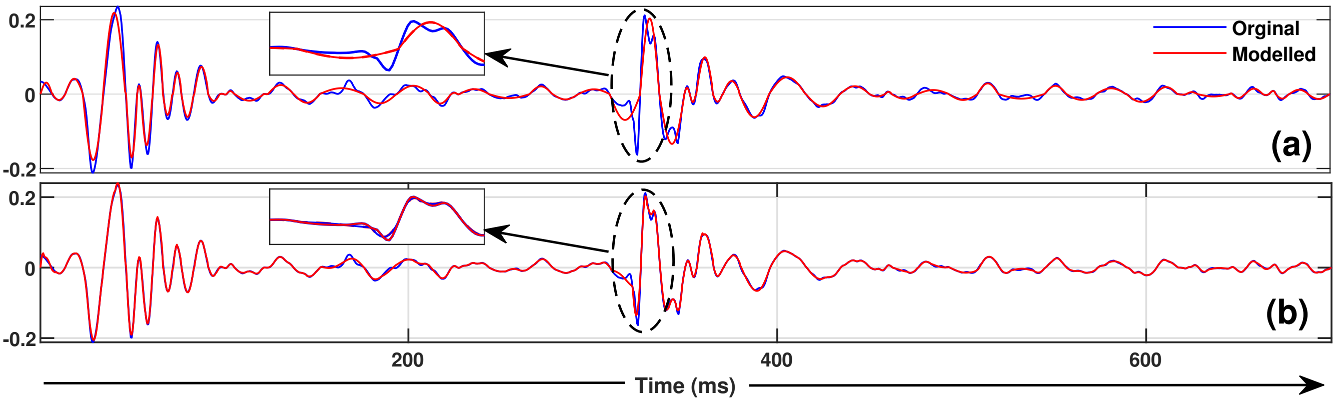

Figure 2.5: Reconstruction of a healthy PCG cycle. Using (a) model order- 1 and (b) model order- 2.

between two consecutive zero-crossing points, i.e., t_i^m and t_{i+1}^m is represented as follows [141]:

$$a_i^m = \frac{\pi}{2}\text{mean}(\mathbf{S}^m)\Big|_{t_i^m}^{t_{i+1}^m} \tag{2.5}$$

$$\omega_i^m = \frac{\pi}{t_{i+1}^m - t_i^m} \tag{2.6}$$

$$\phi_i^m = -\frac{\pi t_i^m}{t_{i+1}^m - t_i^m} \tag{2.7}$$

where (a_i^m, ω_i^m, ϕ_i^m) is the parameter-set for i^{th} zero-crossing point, m^{th} model order and $i \in [1, Z-1]$. Thus, the generated signal ($\mathbf{G}^m$) for m^{th} model order and i^{th} zero-crossing point, is calculated as follows

$$\mathbf{G}^m(t) = a^m(t)\sin(\omega^m(t)\cdot t + \phi^m(t)) \tag{2.8}$$

where $a^m(t) = a_i^m$, $\omega^m(t) = \omega_i^m$, $\phi_i^m(t) = \phi_i^m$, $t \in [t_i^m, t_{i+1}^m]$, and $i \in [1, Z-1]$. The modeled signal ($\mathbf{M}^m$) for m^{th} model order is computed as follows

$$\mathbf{M}^m = \sum_{j=1}^{m} \mathbf{G}^j \tag{2.9}$$

Finally, the residual signal ($\mathbf{R}^m$) for m^{th} model order and the input for $(m+1)^{th}$ model order are obtained as follow

$$\mathbf{R}^m = \mathbf{S}^1 - \mathbf{M}^m, \quad \mathbf{S}^{m+1} = \mathbf{R}^m \tag{2.10}$$

Figure 2.5 and Figure 2.6 depict a healthy and pathological PCG cycle reconstruction using the proposed multi-component oscillatory model up to model order two. It can be observed from both Figure 2.5 and Figure 2.6 that the first model order almost captures the respective

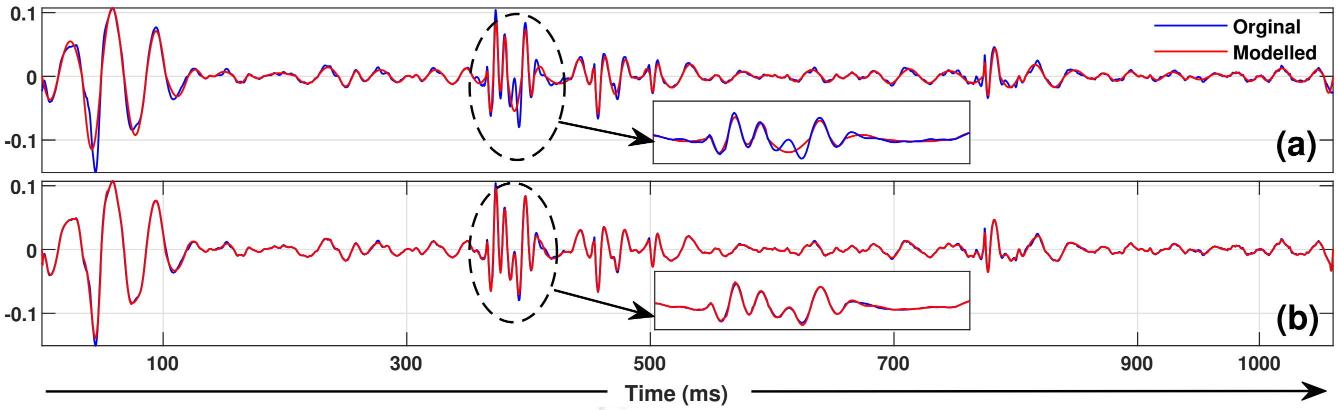

Figure 2.6: Reconstruction of a pathological PCG cycle. Using (a) model order- 1 and (b) model order- 2.

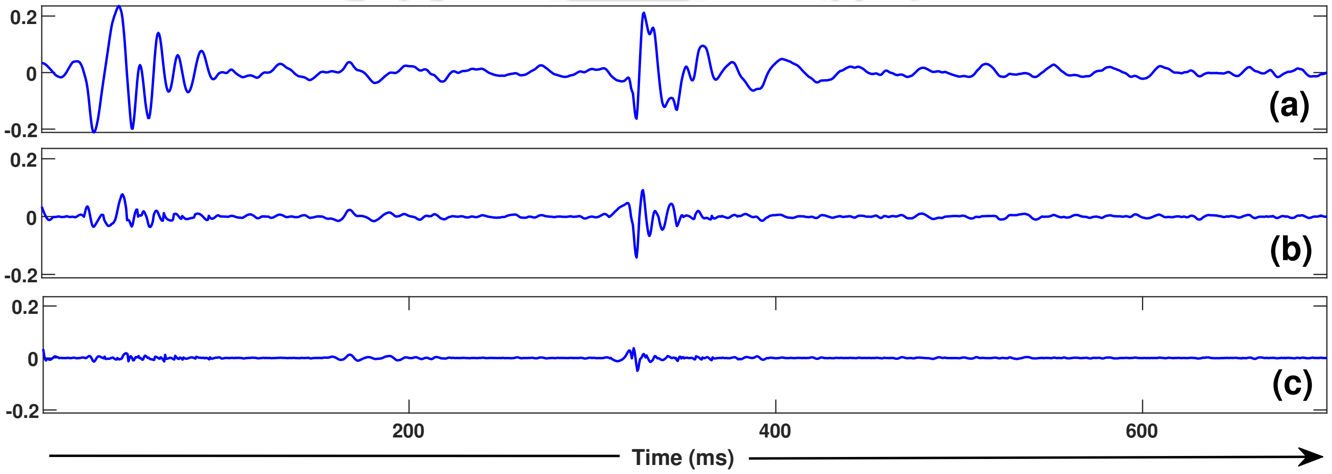

Figure 2.7: The input signal after each iteration for the above-discussed healthy PCG cycle. (a) The original PCG cycle, (b,c) The first and second residual signals, respectively.

PCG cycle patterns. On the other hand, adding second model order helps capture the sudden transient morphological variations of the PCG cycles. Thus, the use of multi-order enhances the representation of the PCG cycles. It also helps capture specific temporal information to improve HVD detections.

Figure 2.7 and Figure 2.8 illustrate the input signal feed to the proposed model for the above-discussed healthy and pathological PCG cycles, respectively. It can be observed from Figure 2.7 and Figure 2.8 that the energy of the input signals starts decreasing with successive iterations. It is almost negligible after the second model order in healthy and pathological PCG cycles, which can be seen in Figure 2.7(c) and Figure 2.8(c).

We have computed the average reconstruction quality for all the healthy and pathological PCG cycles, and their four HS states up to model order eight. We have used two performance

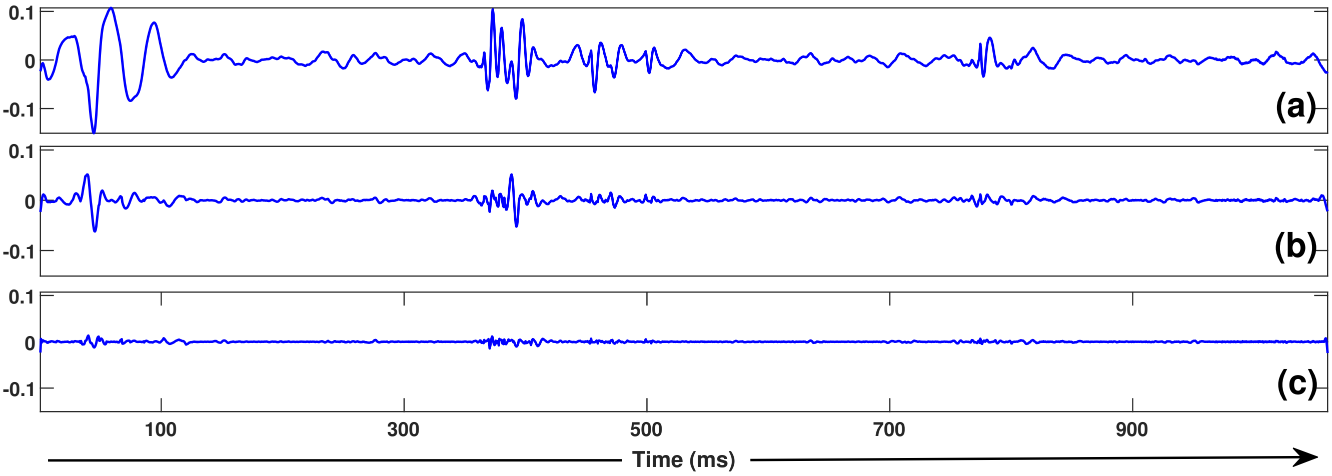

Figure 2.8: The input signal after each iteration for the above-discussed pathological PCG cycle. (a) The original PCG cycle, (b,c) The first and second residual signals, respectively.

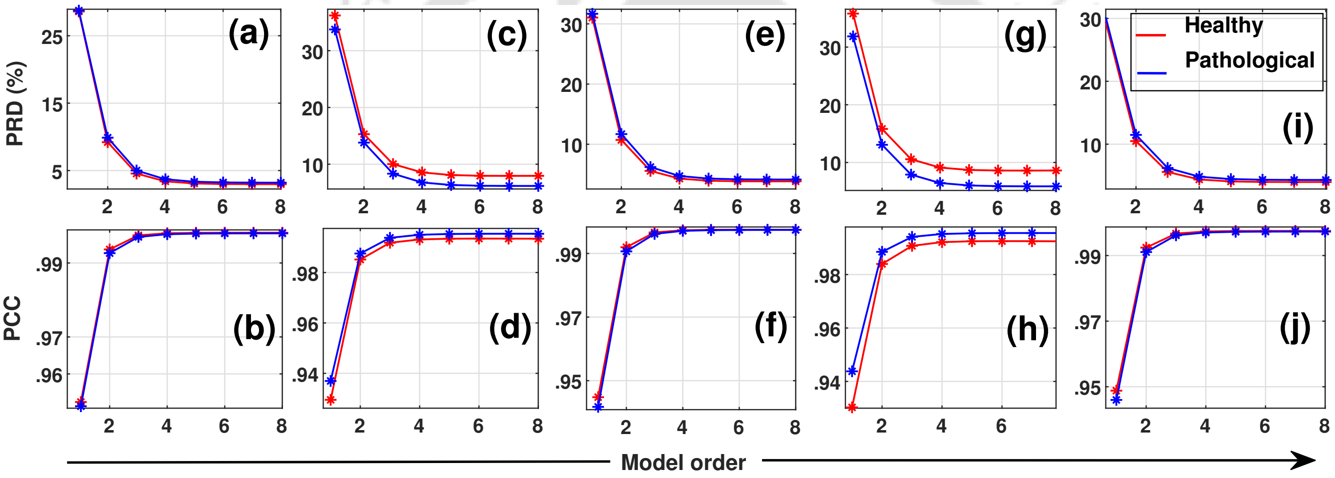

Figure 2.9: The average reconstruction results between original and modeled signals up to model order eight using PRD (%) and PCC measures. (a,b) For S1 sounds, (c,d) For systoles, (e,f) For S2 sounds, (g,h) For diastoles, and (i,j) For PCG cycles.

measures to quantify the reconstruction results, i.e., percentage root-mean-square difference (PRD) [81], and Pearson correlation coefficient (PCC) [142] between the original and the modeled PCG cycles for different model orders. The mathematical formulation of these two metrics is defined as follows:

$$\text{PRD} = \sqrt{\frac{\sum_{n=1}^{N}(\mathbf{S}_n^1 - \mathbf{M}_n^m)^2}{\sum_{n=1}^{N}(\mathbf{S}_n^1)^2}} \times 100 \tag{2.11}$$

$$\text{PCC} = \frac{E[\mathbf{S}^1\mathbf{M}^m]}{\sigma_{\mathbf{S}^1}\sigma_{\mathbf{M}^m}} \tag{2.12}$$

where $\mathbf{S}^1$ is the original PCG cycle, and $\mathbf{M}^m$ is the modeled PCG cycle for m^{th} order. N represents the number of samples present in the PCG cycle. $E[\mathbf{S}^1\mathbf{M}^m]$ is the correlation, and $\sigma_{\mathbf{S}^1}$ and $\sigma_{\mathbf{M}^m}$

are the standard deviations of $\mathbf{S}^1$ and $\mathbf{M}^m$, respectively. Figure 2.9 shows the average PRD (%) and PCC value between original and modeled PCG cycles and their four HS states up to model order eight for both the healthy and pathological classes. It can be seen from Figure 2.9 that the average PRD (%) value decreases as the model order increases for both healthy and pathological classes. Its value does not change significantly above the second model order. It can also be observed from Figure 2.9 that the correlation between the original and modeled PCG cycles and their four HS states increases with the increase of the model order from one to eight for both the healthy and pathological PCG classes. It is almost constant after the second model order. PCC value reaches more than 99% at model order two. The most significant change occurs between model orders one to two in both cases. So, it can be concluded from Figure 2.9 that the proposed multi-component oscillatory model can capture the morphological variations of the PCG cycle and its four HS states for both healthy and pathological classes.

2.1.3.2 Proposed model-based features

The parameter-sets of the proposed oscillator model are used as features to classify PCG cycles and their respective HS states. The size of the feature vector depends upon the model order. It can be observed from Figure 2.7 and Figure 2.8 that the intensity of the input signals starts decreasing with successive iterations. Its value is almost negligible after the second model order in healthy and pathological classes. Thus, the parameter-sets up to model order two are used for classification. With the addition of higher model order, it may introduce noise and decrease the classification performance. This work fits a half-period sine wave function between two zero-crossing points. So, we have not considered the phase. Therefore, the parameter-set for each half-period sine wave consists of amplitude and frequency. An HVD can cause changes in the intensity and shape of the fundamental HSs (S1 and S2 sound). It can also cause additional vibration due to turbulence in blood flow through the defective heart valve, which can be reflected in the systolic or diastolic part of the PCG cycle. So, in this work, we have used prominent amplitudes and their corresponding frequencies as features from model orders one and two for detecting pathological PCG cycles. The number of parameter-sets from each model order is fixed through experimentation.

Table 2.1: Confusion matrix for binary classification of HVDs. Acronym: true positive (TP), false negative (FN), false positive (FP), and true negative (TN).

Actual Class	Predicated Class	
	Pathological	Healthy
Pathological	TP	FN
Healthy	FP	TN

2.1.4 Classification of PCG Cycles and Its Four HS States

The selected parameter-sets from model orders one and two are fed to a supervised classifier to detect the PCG cycles as healthy or pathological. Four binary supervised classifiers, i.e., support vector machine (SVM), k-nearest neighbor (KNN), random forest (RF), and deep neural network (DNN), are deployed to compare the performance accuracy. The SVM classifier is tested with two kernel functions, i.e., (radial basis function) RBF and polynomial (POLY), and the performance of the KNN classifier is evaluated with the two numbers of nearest neighbors, i.e., k= 5 and 10. Similarly, the RF classifier efficiency is measured with two estimator values, i.e., 5 and 10. The DNN model is also tuned with the number of layers, number of neurons per layer, and different values dropout ratios, learning rate, and batch size to improve the performance accuracy to detect the pathological PCG cycles. The general model architecture of a DNN is given in Appendix A.

2.1.4.1 Evaluation Scheme and Performance Measures

A standard five-fold cross-validation method is performed to evaluate the classification of pathological HSs. It ensures generalizable results and prevents overfitting to a specific subset of the data. Table 2.1 shows a confusion matrix for the binary classification of HVDs. Here, TP is the true positive value, TN is the true negative value, FP is the false positive value, and FN is the false negative value of the confusion matrix.

The proposed method is evaluated using five standard measures: precision (Pr), sensitivity (Se), specificity (Sp), F1-score (F1), and overall accuracy (OA). The mathematical formulation of these five measures is given as follows [143].

$$\text{Pr}(\%) = \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100 \tag{2.13}$$

$$\text{Se}(\%) = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100 \tag{2.14}$$

Table 2.2: A brief details of the CICC 2016 database. Acronym: H- Healthy and P- Pathological.

Dataset	# Subjects	# PCG Recordings	Recording Environment	Used Sensor
training set-a	121	H: 117, P: 292	Home-visit, Hospital	Welch-Allyn Meditron
training set-b	151	H: 386, P: 104	Hospital	3M-Littmann
training set-c	32	H: 8, P: 24	Hospital	Audioscope
training set-d	38	H: 27, P: 28	Hospital, Lab setup	Prototype
training set-e	509	H: 1958, P: 183	Hospital, Lab setup	MLT201, 3M-Littmann
training set-f	60	H: 80, P: 34	Hospital	JABES
Total	911	H: 2576, P: 665	—-	—-

$$\text{Sp}(\%) = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100 \tag{2.15}$$

$$\text{F1}(\%) = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \times 100 \tag{2.16}$$

$$\text{OA}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100 \tag{2.17}$$

2.2 Experimental Results and Discussion

This section discusses the details of the PCG database, the results of the proposed method for PCG cycle classification, the performance comparison with the existing methods, and finally, the results of HS states classification.

2.2.1 CICC 2016 Database

The proposed method is evaluated using the publicly available PhysioNet/ Computing in Cardiology Challenge (CICC) 2016 database [13, 144]. It comprises six datasets, i.e., training sets- a, b, c, d, e, and f. Each dataset is recorded independently from different clinical and non-clinical (such as home visits and lab) setups across the globe. Table 2.2 briefly describes the CICC 2016 database. It includes a total of 3241 PCG recordings from 911 subjects. Of these, 2576 records belong to the healthy (H) class, and 665 to the pathological (P) class comprising various HVDs. An echocardiography examination or expert cardiologist verifies the ground truth of the recordings. Most of the PCG recordings are corrupted by various environmental and body noises [13]. The duration of each PCG recording varies between 5 s to 120 s. The PCG data was acquired using different sensors with different sampling rates for each training set. So, each recording was

Table 2.3: Fixing of model order- 1 parameter-sets.

Parameter-sets	OA (%)
40	87.92
50	89.06
60	**90.20**
70	90.15

Table 2.4: Fixing of model order- 2 parameter-sets.

Parameter-sets	OA (%)
5	90.50
10	90.57
15	**90.65**
20	90.63

downsampled to 2000 Hz before being made available publicly. This database is used to classify PCG cycles and its four HS states as healthy or pathological.

2.2.2 Fixing the Model Parameters for PCG Cycle Classification

After preprocessing and segmenting the PCG signals, 56245 healthy and 11123 pathological PCG cycles are obtained. Each PCG cycle is modeled using the proposed multi-component oscillatory model, and parameter-sets (amplitude and frequency) are extracted up to model order two. First, we have experimented with fixing parameter-sets' value for model order- 1. Table 2.3 shows the OA results for different values of parameter-sets to classify PCG cycles. It can be observed from Table 2.3 that 60 parameter-sets, i.e., 60- prominent amplitudes (AM1) and their corresponding frequencies (FM1), show better OA (%). Finally, they are used for our proposed work. Along with the model order- 1 parameter-sets, we have experimented with fixing the value of parameter-sets for model order- 2. Table 2.4 shows the OA results for different values of parameter-sets for model order- 2. It can be observed from Table 2.4 that along with 60 parameter-sets of model order- 1, 15 parameter-sets, i.e., 15- prominent amplitudes (AM2) and their corresponding frequencies (FM2), show better OA (%). So, 75 parameter-sets or 150 parameters (amplitudes and frequencies) are selected through experimentation and used as features to classify PCG cycles as healthy or pathological. We have used the SVM-RBF classifier to perform the above experiments.

The proposed model's performance is evaluated using three binary supervised classifiers, SVM, KNN, and RF. We have tested the OA (%) for different configurations of these classifiers. Table 2.5 shows the performance results for a different combination of features to classify PCG cycles. Combining all features, i.e., AM1, FM1, AM2, and FM2, shows a better OA of 90.65 % using the SVM-RBF classifier.

The statistical analysis of the proposed model features can reveal the qualitative and quantitative characteristics differences between healthy and pathological PCG cycles. The proposed

Table 2.5: Overall accuracy (%) performance results for a different combination of model features.

Proposed model features	**Classifiers**					
	SVM		KNN		RF	
	RBF	POLY	K = 10	K = 5	E = 10	E = 5
AM1-AM2	84.42	83.92	85.85	85.58	86.05	84.84
FM1-FM2	87.89	83.91	86.83	86.60	87.23	86.53
AM1-FM1-AM2-FM2	**90.65**	88.11	88.72	88.65	89.84	89.29

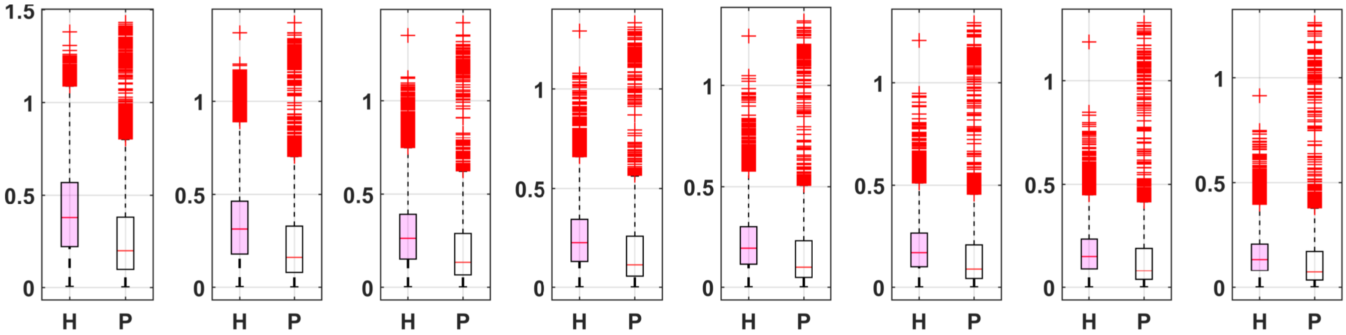

Figure 2.10: Box plot of first eight amplitude features. Acronym: H- Healthy and P- Pathological.

features are analyzed using a box plot and statistical measure t-test. Figure 2.10 shows the box plot of the first eight prominent amplitude features, i.e., Feat1 to Feat8. These features have a higher median value for the healthy class than the pathological class. The distribution characteristics show that the features have a discriminate potential to classify the healthy and pathological PCG cycles. The quantitative analysis of the model features is performed using a statistical hypothesis test, i.e., a t-test [145]. T-test gives two values, i.e., t-value and p-value. A higher t-value and lower p-value show good discrimination capability of a feature. Table 2.6 shows the t-test result for the first eight amplitude features. It can be seen from Table 2.6 that the features show higher t-values and p-values less than 0.0001.

2.2.3 Performance Results on PCG cycles Classification

The performance of the proposed method to classify the pathological PCG cycles is shown in Table 2.7. All the performance measures are five-fold cross-validated. The performance measures are evaluated using SVM-RBF and DNN classifier. We have experimented with different network architectures of the DNN model. The DNN model having three hidden layers, a hidden dimension of 200 for the first two hidden layers and 50 for the last hidden layer, shows the optimum OA (%) result. The hyper-parameter tuning of the DNN model is performed using an Adam optimizer

Table 2.6: T-test result performed on the proposed model-based features.

Feature	Feat1	Feat2	Feat3	Feat4	Feat5	Feat6	Feat7	Feat8
t- value	58.27	55.82	53.28	50.08	45.17	40.66	35.44	30.25
p-value	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$

Table 2.7: The performance results to classify healthy and pathological PCG cycles.

Performance metric	**Classifiers**	
	SVM-RBF	DNN
Precision (%)	78.57	75.77
Sensitivity (%)	59.63	66.03
Specificity (%)	96.78	95.82
F1-score (%)	67.80	70.57
Overall accuracy (%)	**90.65**	**90.91**

with a learning rate of 0.0001. The batch size is set to 128, and the model is trained for 300 epochs. The results of both the SVM-RBF and DNN classifiers show an OA of 90.65% and 90.91%, respectively. The SVM-RBF classifier shows a higher Sp of 96.78% compared to the DNN classifier's Sp of 95.82%, indicating that the SVM-RBF classifier is better at correctly identifying healthy PCG cycles. On the other hand, the DNN classifier shows a higher Se of 66.03% compared to the SVM-RBF classifier's Se of 59.63%, indicating that the DNN classifier is better at correctly identifying pathological PCG cycles. Regarding Pr, the DNN classifier shows a slightly lower Pr (75.77%) than the SVM-RBF classifier (78.57%). The F1 for both the SVM-RBF and DNN classifiers is 67.80% and 70.57%, respectively.

The proposed modeling method is compared with the existing state-of-the-art methods in the literature to classify PCG cycles. The comparison results are shown in Table 2.8. The reported methods have used time, frequency, time-frequency, and MFCC features to classify PCG signals. However, the proposed work outperforms these methods with a better OA of 90.91%. It indicates that the proposed multi-component oscillatory model captures the temporal dynamics and fine-grained frequency resolution that detect specific frequency components associated with HVDs. The model handles the PCG signal's non-stationary behavior, transient phenomena, and overlapping frequency components. Further, it can decompose the PCG signal into its constitute oscillatory components, facilitating detailed analysis of the cardiac sounds.

Table 2.8: Comparison of the proposed work with the existing state-of-the-art methods to classify PCG cycles as healthy or pathological.

Author	Feature extraction method	Classifier	OA (%)
Potes *et al.* [100]	Time-frequency	AdaBoost and CNN	86.02
Zabhi *et al.* [95]	Time-frequency	Ensemble of NNs	85.90
Kay *et al.* [105]	Wavelet, MFCC	Drop connected NN	85.20
Bobillo *et al.* [146]	Time-frequency, wavelet, MFCC	LR, SVM and KNN	84.54
Homsi *et al.* [147]	Time, frequency, wavelet	Ensemble of classifiers	84.48
Proposed method	**Multi-component oscillatory model**	SVM-RBF	**90.65**
Proposed method	**Multi-component oscillatory model**	DNN	**90.91**

2.2.4 Performance Results on HS States Classification

We have also deployed the proposed multi-component oscillatory model to classify the HS states, i.e., S1 sound, systole, S2 sound, and diastole, as healthy or pathological. After preprocessing and segmenting the PCG signals, 56245 healthy and 11123 pathological HS states are obtained with respect to each PCG cycle. Each HS state is modeled using the proposed multi-component oscillatory model, and parameter-sets are extracted up to model order two. We have experimented to fix the parameter-sets for each model order. Eight prominent amplitudes and their corresponding frequencies from model order one, while four prominent amplitudes and their corresponding frequencies from model order two are used as features. The performance of the proposed model is evaluated with two supervised classifiers, SVM and k-nearest neighbors (KNN). Different configurations are utilized for these classifiers, such as SVM-linear, SVM-RBF, KNN (k=5), and KNN (k=10). We have also compared the proposed model results with the state-of-the-art sinusoidal model [148]. The features of the sinusoidal model are estimated as follows: (i) First, Fourier transform (FT) is obtained for each HS segment, (ii) Second, from the magnitude spectrum of the FT, first L prominent magnitudes and their corresponding frequencies, i.e., $[(A_1, f_1), (A_2, f_2), ..., (A_L, f_L)]$ are obtained. Here, L represents the order of the sinusoidal model. We have experimented with different values of L. The $L = 6$ shows the optimum results to classify the HS segments. The reconstruction of a HS segment $\mathbf{s}[n]$ using the L^{th}- order sinusoidal model is obtained as follows:

$$\mathbf{s}_r[n] = \sum_{l=1}^{L} a_l \sin(2\pi f_l T_s n + \theta_l) \tag{2.18}$$

where a_l, f_l, θ_l represents the amplitude, frequency, and phase of the l^{th}- order sinusoidal model. $\mathbf{s}_r[n]$ is the reconstructed HS segment. Where $n = (0, 1, 2, ..., N-1)$ and N is the total number of

Table 2.9: Comparison of OA (%) results for classification of HS states using the proposed and sinusoidal model.

Category	Proposed model				Sinusoidal model			
	SVM		KNN		SVM		KNN	
	RBF	Linear	k = 5	k = 10	RBF	Linear	k = 5	k = 10
S1 sound	**83.32**	79.60	80.18	80.99	82.76	78.83	80.13	80.29
Sys. segment	**82.61**	79.05	79.68	80.22	81.88	78.85	79.00	79.41
S2 sound	**83.18**	81.08	81.23	81.85	81.12	78.85	78.42	78.85
Dia. segment	**81.88**	78.76	78.92	79.73	80.43	78.81	78.65	79.31

samples of the HS segment. The performance results to classify HS states are provided in Table 2.9. It is observed that our proposed model outperforms the classical sinusoidal model across all the classifier configurations. The proposed model with the SVM-RBF classifier achieves better performance among all. It produces OA (%) of 83.32, 82.61, 83.18, and 81.88 for the S1 sound, systole, S2 sound, and diastole, respectively.

2.3 Summary

In this chapter, we proposed a novel multi-component oscillatory model to classify PCG cycles and its four HS states as healthy or pathological. We analyzed the proposed model in detail for the representation and recognition of PCG cycles and their respective HS states. The representation results show that the PCC value reaches almost 99% at model order two for healthy and pathological PCG cycles. For the classification of PCG cycles, the proposed method has shown an OA of 90.91% using the proposed model-based features with DNN. The results show that the proposed method is competitive enough with the existing state-of-the-art methods. The proposed method is easy to compute and implement as only zero-crossing points have to calculate and fit a half-period sine wave function between two consecutive zero-crossing points.

3

Acoustic Feature-fusion and Hierarchical LSTM Network for HMs Classification

Contents

In the last chapter 2, a multi-component oscillatory model is investigated to classify PCG cycles as healthy or pathological. The model extracts amplitude and frequency features by fitting a half-period sine wave between two successive zero-crossing points of the PCG cycle. However, designing such hand-crafted features needs domain expertise and is often subjective. Sometimes, finding a fixed set of discriminative features for training a DL model is challenging due to the variability in the HVDs manifestations. The existing feature-fusion-based approaches [70, 121] fail to provide complementary information to detect HVDs. Further, classical ML models [59, 63] and DNN, CNN-based deep learning (DL) models [70, 100] cannot exploit the temporal information present in the PCG signal. So, there is a need for an effective feature-fusion algorithm along with a DL model that exploits the temporal information of the fused features to enhance the detection of HVDs.

PCG is quasi-periodic with repetitive PCG beats. A PCG beat is divided into four basic PCG waveforms (i.e., S1 sound, systole, S2 sound, and diastole). When HVDs originate in heart valves, the heart produces some rasping or blowing sound, known as HMs, instead of a regular rhythm sound (lub-dub). It is reflected as morphological and temporal variations in the four basic PCG waveforms discussed in chapter 1. To capture these acoustic variations and extract useful information for detecting HMs, we deploy two acoustic features: MFCCs and linear prediction cepstral coefficients (LPCCs). MFCCs and LPCCs features have been successfully used to analyze speech signals [149]. MFCCs have also shown promising results for the PCG signal analysis by mimicking the human auditory perception [44, 59]. LPCCs are based on linear prediction (LP) coefficients and were initially developed to model the vocal-tract resonance characteristics of speech production. In the case of PCG signals, the production of HSs shows sharp resonance patterns, i.e., S1 and S2 sound, accompanied by HMs in the case of HVDs. These resonance patterns carry diagnostic information for the classification of HMs. It is hypothesized that LPCCs help capture these resonance patterns of HSs in the spectral domain to diagnose HMs. Combining these two features is hypothesized to enhance the spectral representation, improve the noise robustness, and provide a compact representation for the PCG signal. Therefore, this chapter proposes two feature-fusion models by systematically combining MFCCs with LPCCs. To exploit the temporal and relative temporal variation among the fused features, we propose a hierarchical LSTM (HLSTM) network. Further, we have introduced a self-attention module to aggregate the encoded vectors of the HLSTM network based on their clinical relevance to improve the classification of HMs.

The rest of the chapter is organized as follows. Section 3.1 presents the proposed framework for HMs classification. Section 3.2 discusses the experimental results for the acoustic feature-fusion and HLSTM network. The chapter summary is presented in section 3.3.

3.1 Acoustic Feature-fusion and HLSTM Network to Classify HMs

This section presents the proposed acoustic feature-fusion and attention-based HLSTM network for classifying HMs. First, the raw PCG signal is preprocessed, which includes filtering, spike removal, and blind segmentation. Second, extraction of MFCC and LPCC features from the PCG segments. Third, MFCC and LPCC features are fused using the HLSTM network with a self-attention module. Finally, the classification of HMs. Figure 3.1 shows the block diagram of the proposed framework. The detailed description of each step of the framework is as follows.

3.1.1 Preprocessing of PCG Signal

The raw PCG signal is first preprocessed, which is discussed in the last chapter 2.1.1. Figure 3.2(a) and Figure 3.2(b) show the raw and preprocessed PCG signal, respectively. In most existing works, after preprocessing, the PCG signal is segmented cycle-wise to analyze the HVDs on respective cycles [13, 68]. The segmentation of the PCG signal into PCG beats requires locating S1 and S2 sounds. Locating these two HSs is challenging and leads to false detection due to artifacts and pathological variations in PCG morphologies [29]. So, we have segmented each preprocessed PCG signal $\mathbf{S_p}[r]; (r = 1, 2, 3, ..., R)$ blindly (without locating S1 and S2 sound) using a 2s rectangular window with a 75 % overlap. Figure 3.2(c) and Figure 3.2(d) show the first two consecutive PCG segments of the preprocessed PCG signal. Each 2 s PCG segment $\mathbf{C}[r]; (r = 1, 2, 3, ..., Q)$ contains at least one complete PCG cycle information [121]. This simple segmentation avoids the complicated HS segmentation algorithms to segment PCG signals cycle-wise [29]. It also helps with data augmentation for model training and testing to classify HMs.

3.1.2 Feature Extraction

MFCC and LPCC features are extracted from each PCG segment ($\mathbf{C}[r]; (r = 1, 2, 3, ..., Q)$). The steps involved in extracting these two features are as follows.

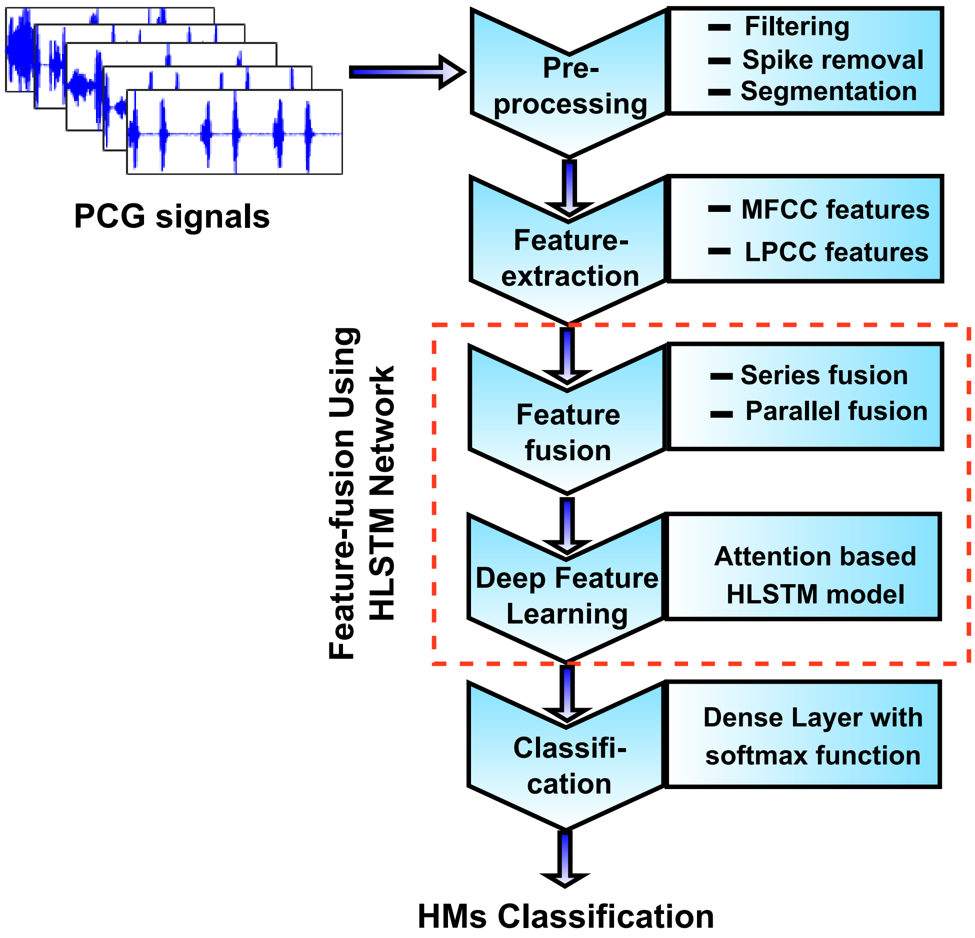

Figure 3.1: Workflow of the proposed framework for automated HMs classification.

3.1.2.1 MFCC feature extraction:

MFCC features have been used extensively for the PCG signal analysis [59, 70, 120, 150]. It comprises pre-emphasis filtering, framing, fast Fourier transform (FFT), Mel-scale-based filtering, logarithmic transformation, and discrete cosine transformation (DCT) [73]. First, each PCG segment ($\mathbf{C}[r]$) is passed through a first-order filter with a pre-emphasis coefficient (α) of 0.95. It helps in enhancing high-frequency HS energy contents. Then, each pre-emphasized PCG segment is split into 79 frames using a 50 ms Hamming window with a 50% overlap. This windowing helps provide continuity at the edges and ensures a smoother transition from one frame to another. Second, FFT is applied to each frame to obtain the magnitude spectrum. Third, K number of triangular filter banks are used in mel-scale to uniformly divide the entire frequency band of each frame [74]. The optimum value of K is set through experiments, which is discussed

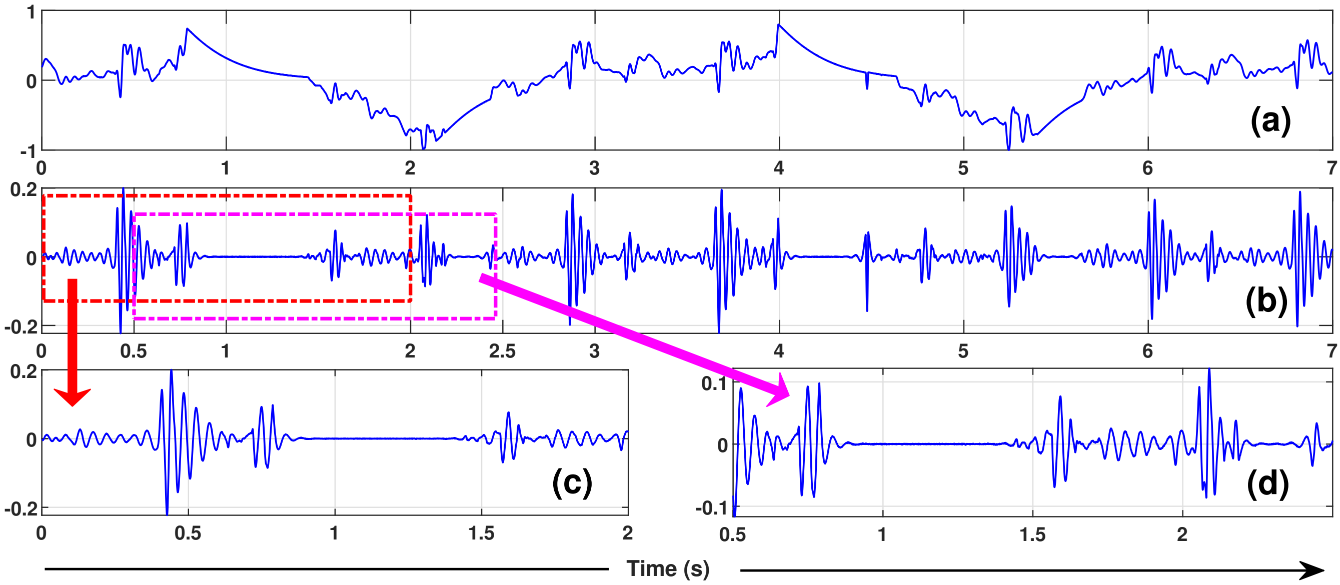

Figure 3.2: Preprocessing and blind segmentation of the PCG signal. (a) Raw PCG signal, (b) Preprocessed PCG signal, (c) and (d) First two consecutive PCG segments.

in Section 3.2. Fourth, energies are computed for each frame by multiplying each filter bank with its corresponding magnitude spectrum. Finally, DCT is computed from the logarithm of those energies to get the MFCCs for each frame. The formula to obtain the n^{th} coefficient MFCC for t^{th} frame $\mathbf{M}_{n,t}$ is as follows [75]:

$$\mathbf{M}n,t = \sum_{k=1}^{K} (\log \mathbf{F}_{k,t}) \cos\left[\frac{\pi n(k-0.5)}{K}\right]; (1 \leq k \leq K) \tag{3.1}$$

where $1 \leq n \leq L$ and L is the desired number of MFCC coefficients. We have considered $L = K$ for this work. K is the number of filter banks used for each frame. $\mathbf{F}_{k,t}$ is the k^{th} filter-bank energy of t^{th} frame.

3.1.2.2 LPCC feature extraction:

The fundamental idea of the LPCC is derived from linear prediction coefficients (LPC). LPCC is the compact representation of LPC in the cepstrum domain. The production of the PCG signal shows sharp resonance patterns (S1 and S2) accompanied by some murmurs (in case of abnormality). These resonance patterns carry diagnosis information of the heart valves. It is of interest to detect these resonance patterns' position and strength. LPC captures the spectrum of these sharp resonances and provides information regarding HMs. Figure 3.3(a) and Figure 3.3(b) show a normal PCG segment with its original spectrum and LPC envelope. Figure 3.3(c) and

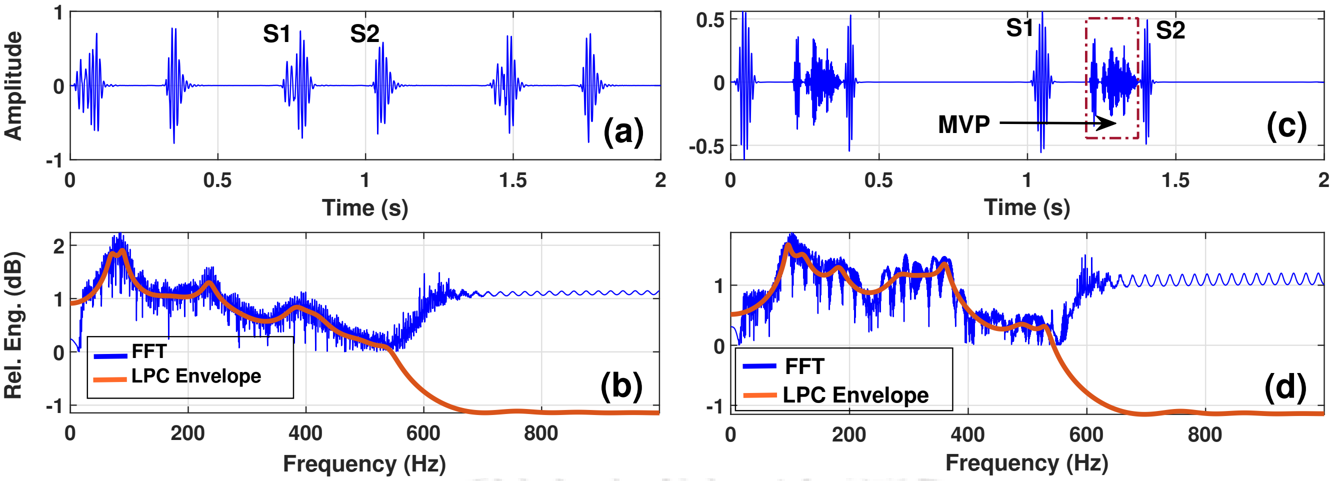

Figure 3.3: The PCG segment and its LPC spectral analysis. (a,b) A normal PCG segment along with its spectrum and LPC envelope, (c,d) A MVP PCG segment along with its spectrum and LPC envelope.

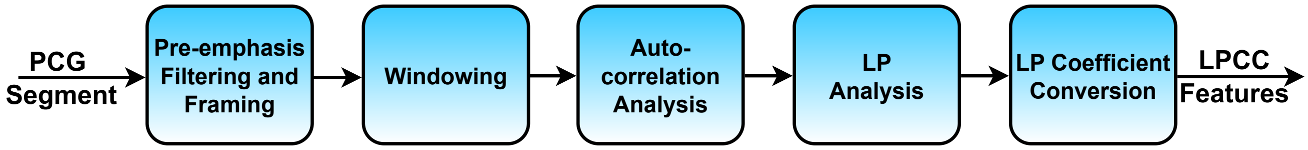

Figure 3.4: Steps involved for LPCC feature extraction from the PCG segment.

Figure 3.3(d) show an MVP PCG segment with its original spectrum and LPC envelope. It can be observed from Figure 3.3 that LP spectra capture the normal and MVP murmur information smoothly and efficiently. Therefore, extracting LPCC features can provide specific information for detecting HMs. The LP assumes that a linear combination of past P samples can predict the current signal sample. It can be formulated as follow:

$$\hat{s}(n) = \sum_{i=1}^{P} \alpha_i s(n-i) \tag{3.2}$$

where $\hat{s}(n)$ is the predicated value of $s(n)$, α_i is the LP coefficient and it is a P^{th} order auto-regressive (AR) process. Figure 3.4 shows the block diagram of the LPCC feature extraction process. First, each PCG segment ($\mathbf{C}[r]$) is passed through a first-order filter with a pre-emphasis coefficient (α) of 0.95. Second, each pre-emphasized PCG segment is split into 79 frames using a 50 ms Hamming window with a 50% overlap. Third, we apply Yule-Walker equations to each frame, which use auto-correlation functions to estimate the LPC coefficients (α_i) [151]. So, the

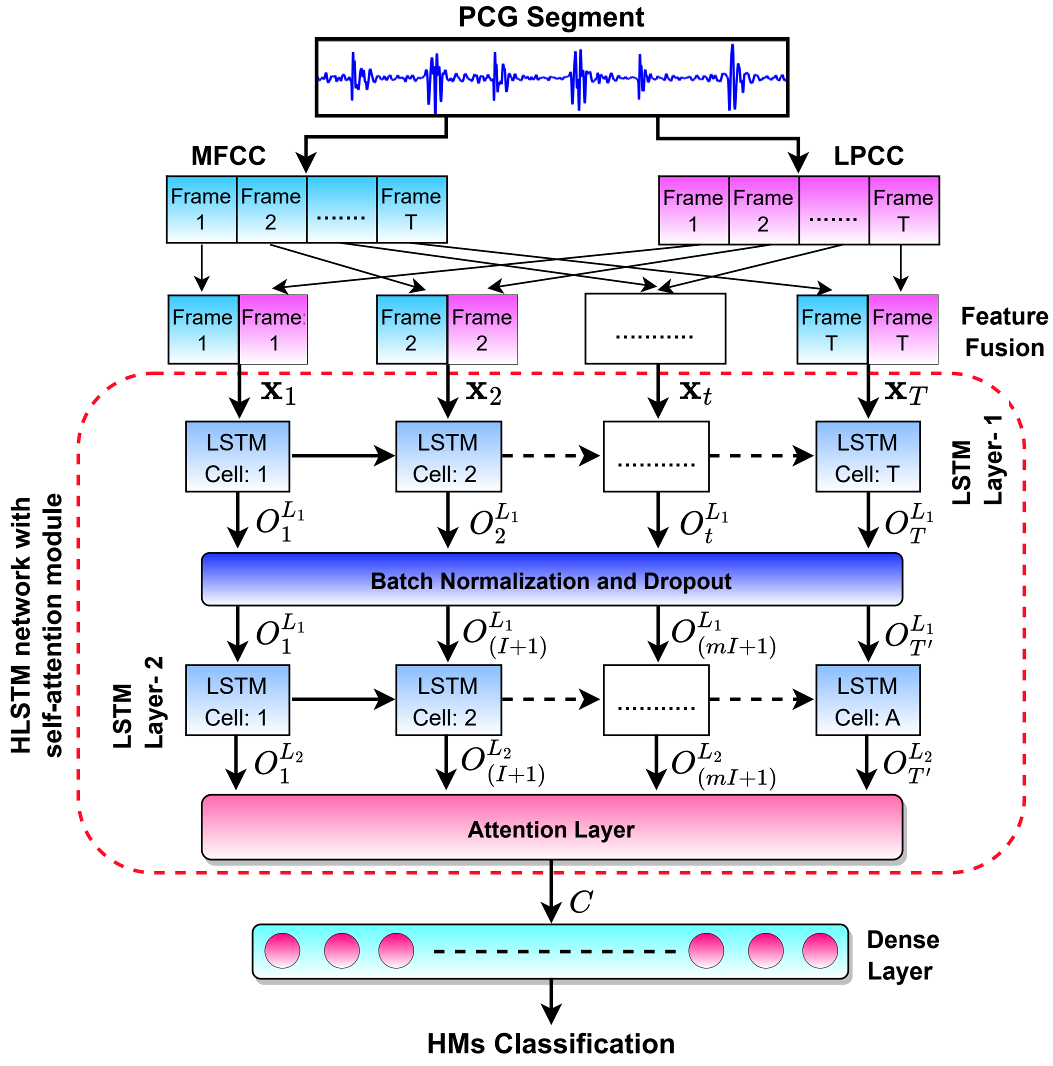

Figure 3.5: Series feature fusion (SFF) model using hierarchical LSTM (HLSTM) with self-attention to classify HMs.

m^{th} LPCC coefficient for t^{th} frame $\mathbf{L}_{m,t}$ is obtained by using a recursive formula as follows [152]:

$$\mathbf{L}_{m,t} = \begin{cases} \log_e P, \quad m = 0 \\ \alpha_m + \displaystyle\sum_{k=1}^{m-1} \left[\frac{k}{m} \cdot \alpha_{m-k} \cdot \mathbf{L}_{k,t} \right], \quad 1 \leq m \leq P \\ \displaystyle\sum_{k=1}^{m-1} \left[\frac{k}{m} \cdot \alpha_{m-k} \cdot \mathbf{L}_{k,t} \right], \quad P < m < P' \end{cases} \tag{3.3}$$

where LP coefficients are defined as $[\alpha_1, \alpha_2, ..., \alpha_P]$ and P is the order of the process. The optimum value of P is set through experiments and is discussed in Section 3.2.

3.1.3 Feature-fusion Using Attention-based HLSTM

A self-attention-based HLSTM network is used to fuse the MFCC and LPCC features in two novel ways to detect HMs. One model is the series feature fusion HLSTM (SFF-HLSTM) model. Another is the parallel feature fusion HLSTM (PFF-HLSTM) model.

3.1.3.1 SFF-HLSTM model:

The obtained MFCC ($\mathbf{M}_n$) and LPCC ($\mathbf{L}_m$) features of each frame for a PCG segment ($\mathbf{C}[r]$) are appended sequentially to obtain the input sequences for the SFF-HLSTM model. Figure 3.5 shows the SFF-HLSTM model. The input sequence $\mathbf{x}_t$ at timestamp t is formed as follows:

$$\mathbf{x}_t = [(\mathbf{M}_{n,t}), (\mathbf{L}_{m,t})] \tag{3.4}$$

where $\mathbf{M}_{n,t} \in \mathbb{R}^K$ is the t^{th} frame MFCC feature vector with a dimension of K. $\mathbf{L}_{m,t} \in \mathbb{R}^P$ is the t^{th} frame LPCC feature vector with a dimension of P. This forms the input sequences $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T$ to the first layer of the self-attention-based HLSTM network as shown in Figure 3.5. Where $\mathbf{x}_t \in \mathbb{R}^{(K+P)}$ and T are the total number of sequences or timestamps.

3.1.3.2 PFF-HLSTM model:

The obtained MFCC ($\mathbf{M}_n$) and LPCC ($\mathbf{L}_m$) features from each frame of a PCG segment ($\mathbf{C}[r]$) are fed separately as input sequences to the two HLSTM networks for PFF-HLSTM model. Figure 3.6 shows the PFF-HLSTM model. The first network is the MFCC-HLSTM network with a self-attention module. The second network is the LPCC-HLSTM network with a self-attention module. So, the input sequences to the MFCC-HLSTM network are formed as follows:

$$(\mathbf{x}_{m1}, \mathbf{x}_{m2}, ..., \mathbf{x}_{mT}) = [(\mathbf{M}_{n,1}), (\mathbf{M}_{n,2}), ..., (\mathbf{M}_{n,T})] \tag{3.5}$$

where $\mathbf{M}_{n,T} \in \mathbb{R}^K$ is the T^{th} frame MFCC feature vector and $\mathbf{x}_{mT} \in \mathbb{R}^K$ is the input for timestamp T, as shown in Figure 3.6. Similarly, the input sequences to the LPCC-HLSTM network are formed as follows:

$$(\mathbf{x}_{l1}, \mathbf{x}_{l2}, ..., \mathbf{x}_{lT}) = [(\mathbf{L}_{n,1}), (\mathbf{L}_{n,2}), ..., (\mathbf{L}_{n,T})] \tag{3.6}$$

where $\mathbf{L}_{n,T} \in \mathbb{R}^P$ is the T^{th} frame LPCC feature vector and $\mathbf{x}_{lT} \in \mathbb{R}^P$ is the input for timestamp T, as shown in Figure 3.6.

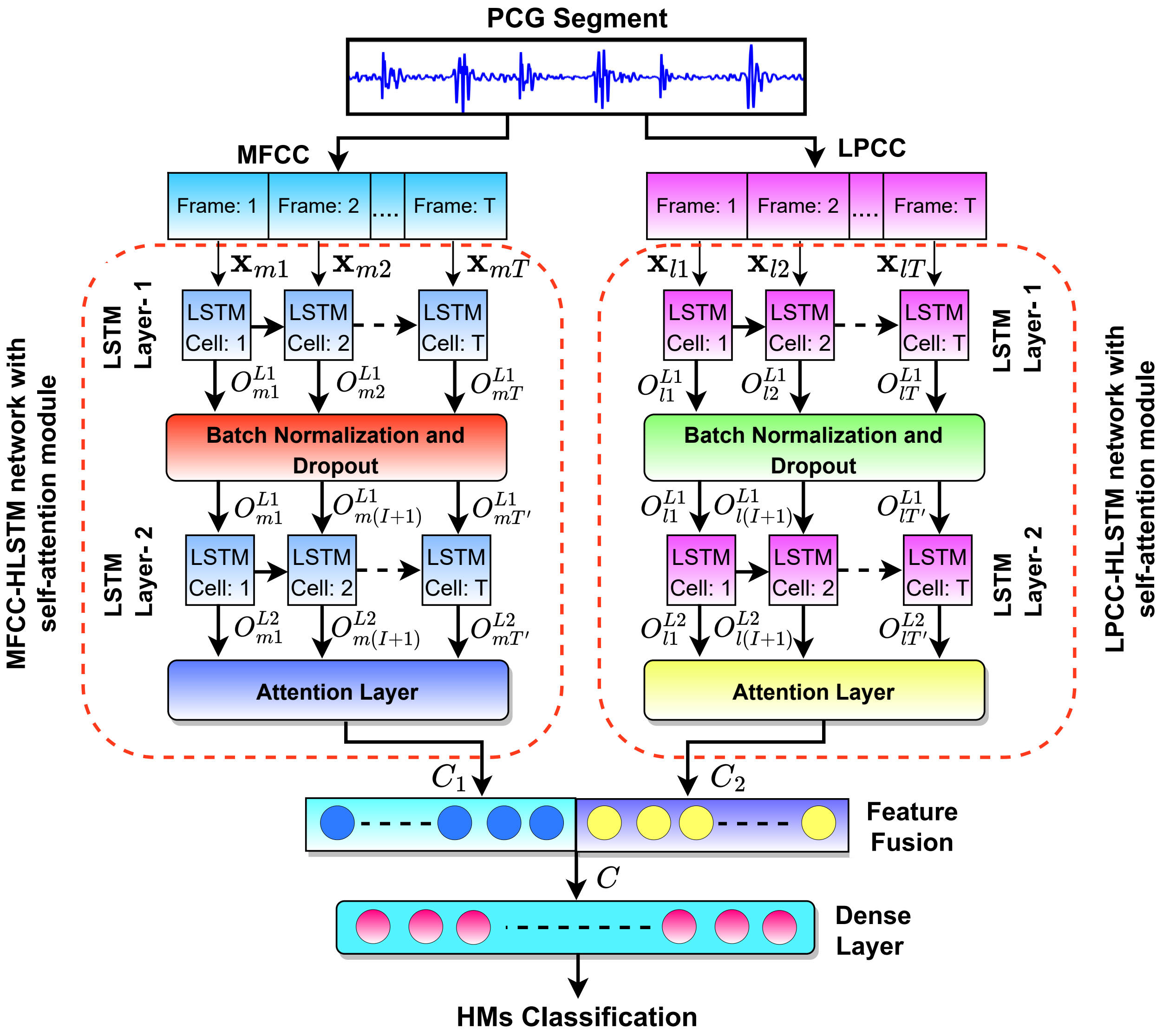

Figure 3.6: Parallel feature fusion (PFF) model using hierarchical LSTM (HLSTM) with self-attention to classify HMs.

3.1.3.3 Proposed HLSTM network with self-attention module:

This work proposes a self-attention-based HLSTM network for the fusion of acoustic features for both the SFF-HLSTM and PFF-HLSTM models. The basic idea of the HLSTM network is based on the LSTM model. LSTM is widely used because of its ability to learn long-term dependencies of sequential data [153]. Figure 3.7 shows the structure of an LSTM cell. The input to the LSTM cell at present state t is $\mathbf{x}_t$, hidden state (h_{t-1}) and cell state (c_{t-1}) from the previous LSTM state. An LSTM cell contains three gates, i.e., input gate (i_t), output gate (o_t), and forget gate (f_t), to add or remove information. The forget gate erases the irrelevant information present in the previous

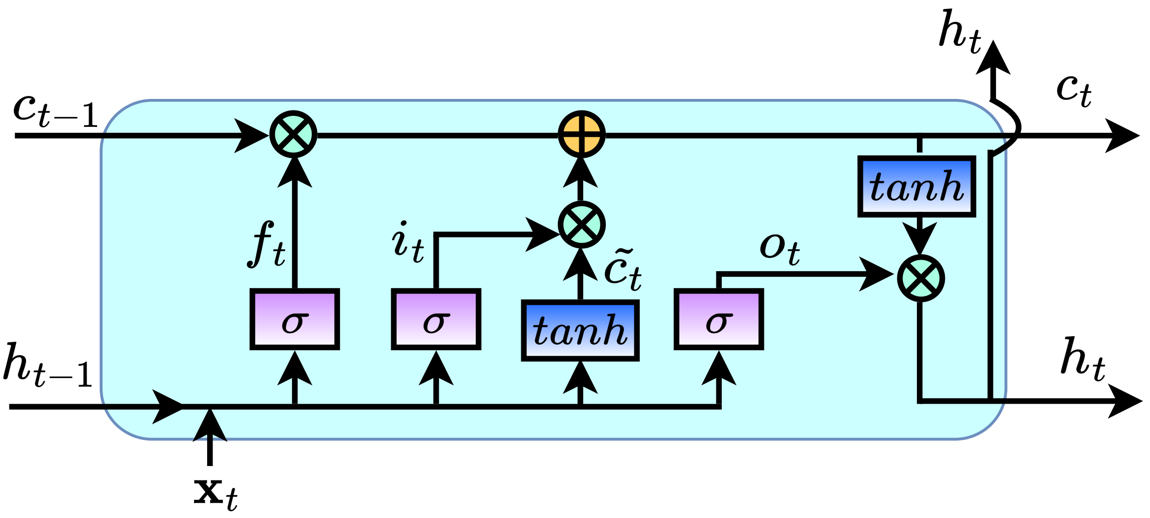

Figure 3.7: Architecture of an LSTM cell.

cell state c_{t-1} and expresses as:

$$f_t = \sigma(\mathbf{W}_f \cdot [\mathbf{x}_t, h_{t-1}] + \mathbf{b}_f) \tag{3.7}$$

where $\mathbf{W}_f$ and $\mathbf{b}_f$ represents the weight matrix and bias vector, respectively. The sigmoid (σ) is the activation function to limit the f_t to $[0, 1]$. i_t and $\tilde{c}_t$ determine the part of relevant information to retain from the present state. So, LSTM selectively updates its present cell state (c_t) by adding the present state relevant information with the previous state relevant information. The following equations compute c_t:

$$\tilde{c}_t = \tanh(\mathbf{W}_g \cdot [\mathbf{x}_t, h_{t-1}] + \mathbf{b}_g) \tag{3.8}$$

$$i_t = \sigma(\mathbf{W}_i \cdot [\mathbf{x}_t, h_{t-1}] + \mathbf{b}_i) \tag{3.9}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{3.10}$$

where [$\mathbf{W}_g$, $\mathbf{b}_g$] and [$\mathbf{W}_i$, $\mathbf{b}_i$] are weight matrix and bias vector of $\tilde{c}_t$ and i_t, respectively. $\tanh()$ is the hyperbolic tangent function. $\odot$ is the point-wise multiplication operator. The output gate (o_t) controls the flow of information and is fed to the next state as h_t. It is formulated as follows:

$$o_t = \sigma(\mathbf{W}_o \cdot [\mathbf{x}_t, h_{t-1}] + \mathbf{b}_o) \tag{3.11}$$

$$h_t = o_t \odot c_t \tag{3.12}$$

where $\mathbf{W}_o$ and $\mathbf{b}_o$ are the weight matrix and bias vector of the LSTM output (o_t), respectively.

Although the LSTM network performs well for sequential data, it does not capture the hierarchical information present in the temporal domain as CNN captures in the spatial domain. To overcome this limitation of LSTM, we propose an HLSTM model to detect HMs. HLSTM captures the temporal information present in a PCG segment related to HMs. It also exploits the relative temporal information among the basic PCG waveforms (i.e., S1 sound, systole, S2 sound, and diastole) related to HMs. HLSTM is designed by stacking two layers of LSTM with update intervals [154].

In the SFF-HLSTM model (Figure 3.5), the input to the first layer HLSTM ($\mathbf{L}_1$) is the PCG segment fused features which are represented as:

$$\mathbf{x}^{\mathbf{L}_1} \in \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t, ..., \mathbf{x}_T\} \tag{3.13}$$

where $\mathbf{x}_t \in \mathbb{R}^{d_{in}}$ is the fused MFCC and LPCC features at timestamp t and a dimension of $d_{in} = (K + P)$. So, a total of T timestamps of fused-feature correspond to a PCG segment, fed to the $\mathbf{L}_1$. The outputs of $\mathbf{L}_1$ are represented as follows:

$$\mathbf{O}^{\mathbf{L}_1} \in \{O_1^{L_1}, O_2^{L_1}, ..., O_t^{L_1}, ..., O_T^{L_1}\} \tag{3.14}$$

where $\mathbf{O}^{\mathbf{L}_1} \in \mathbb{R}^{d_{h1}}$ are the outputs of $\mathbf{L}_1$ and d_{h1} is the dimension of each output. So, $\mathbf{L}_1$ exploits both the HS perceptual and production temporal information of the PCG segment.

The second layer of HLSTM ($\mathbf{L}_2$) utilizes the relative temporal information in the basic PCG waveforms. It is performed by feeding the output of $\mathbf{L}_1$ at an interval of I to the input of $\mathbf{L}_2$. The value of interval I can have an impact on the performance of the HLSTM network to detect HMs. A smaller value of I may allow the network to capture more detailed information, leading to overfitting or a more complex network architecture. A larger interval I may simplify the network architecture and potentially lose vital information related to HMs. So, the optimum value of I is computed experimentally and is discussed in section 3.2. The input and output of $\mathbf{L}_2$ is represented as follows:

$$\mathbf{X}^{\mathbf{L}_2} \in \{O_1^{L_1}, O_{(I+1)}^{L_1}, O_{(2I+1)}^{L_1}, ..., O_{(mI+1)}^{L_1}, ..., O_{T'}^{L_1}\} \tag{3.15}$$

$$\mathbf{O}^{\mathbf{L}_2} \in \{O_1^{L_2}, O_{(I+1)}^{L_2}, O_{(2I+1)}^{L_2}, ..., O_{(mI+1)}^{L_2}, ..., O_{T'}^{L_2}\} \tag{3.16}$$

where $m = \{0, 1, 2, ..., [\frac{T}{I}] - 1\}$. $T' = [\frac{T}{I}] - 1\}$ is the last time stamp that is fed to the input of $\mathbf{L}_2$. $\mathbf{O}^{\mathbf{L}_2} \in \mathbb{R}^{d_{h2}}$ are the outputs of $\mathbf{L}_2$ and d_{h2} is the dimension of each output.

HLSTM encodes the temporal and relative temporal information of the PCG segment into output vectors $\mathbf{O}^{\mathbf{L}_2}$. To further enhance the representation power of the output vectors in the direction of HMs detection, the HLSTM encoding blocks are fed to a self-attention (SA) layer. The SA layer selectively assigns more weights to the informative feature vectors and fewer weights to less useful feature vectors [155]. So, the encoded outputs of $\mathbf{L}_2$ are fed to a self-attention layer to assign a score which is calculated as follows:

$$S_t^{L_2} = \mathbf{W}_a \cdot O_t^{L_2} + \mathbf{b}_a \tag{3.17}$$

where $S_t^{L_2}$ and $O_t^{L_2} \in \mathbb{R}^{d_{h2}}$ are the t^{th}- timestamp score and output of the $\mathbf{L}_2$. $[\mathbf{W}_a \in \mathbb{R}^{d_a}, \mathbf{b}_a \in \mathbb{R}]$ are the weight vector and bias of the attention layer. So, the attention weight of $O_t^{L_2}$ at timestamp t is obtained as:

$$\alpha_t = \text{softmax}\left[S_t^{L_2}\right] = \frac{\exp(S_t^{L_2})}{\sum_{t=1}^{T'} \exp(S_t^{L_2})} \tag{3.18}$$

where T' is the total number of timestamps of L_2. Then, a context vector ($\mathbf{C}$) is obtained by the following equation:

$$\mathbf{C} = \sum_{t=1}^{T'} c_t = \sum_{t=1}^{T'} \alpha_t \cdot O_t^{L_2} \tag{3.19}$$

Finally, this context vector is fed to the classification block to get the final decision output, as shown in Figure 3.5.

The PFF-HLSTM model (Figure 3.6) has two HLSTM networks, i.e., the MFCC-HLSTM and LPCC-HLSTM networks. So, the PFF-HLSTM model's objective is to exploit both MFCC and LPCC features independently for a PCG segment to extract complementary information for detecting HVDs. For the MFCC-HLSTM network, the first layer ($\mathbf{L}_1^m$) exploits the temporal HS perceptual information of the PCG segment. The second layer of the MFCC-HLSTM network ($\mathbf{L}_2^m$) is deployed to exploit the relative temporal HS perceptual information among the MFCC features at different timestamps. Consequently, it exploits the relative temporal information among the basic PCG waveforms. So, the input to $\mathbf{L}_2^m$ is the output of $\mathbf{L}_1^m$ at an interval of I (Figure 3.6) and is represented as $O_{m1}^{L_2}, O_{m(I+1)}^{L_2}, ..., O_{mT'}^{L_2}$. Similarly, the first layer of the LPCC-HLSTM network ($\mathbf{L}_1^l$) exploits the temporal HS production information of the PCG segment. The second layer of the LPCC-HLSTM network ($\mathbf{L}_2^l$) is deployed to exploit the relative temporal HS production information among the LPCC features at different timestamps. Consequently, it exploits the relative temporal information among the basic PCG waveforms. So, the input to $\mathbf{L}_2^l$ is the output of $\mathbf{L}_1^l$ at an interval of I (Figure 3.6) and is represented as $O^{L_2}_{l1}, O^{L_2}_{l(I+1)}, ..., O^{L_2}_{lT'}$. To further enhance the representation power of the output vectors of MFCC-HLSTM ($\mathbf{O}^{\mathbf{L}^m_2}$) and the output vectors of LPCC-HLSTM network ($\mathbf{O}^{\mathbf{L}^l_2}$) to detect HMs, both the output vectors are fed to their respective SA layers. Based on their attention weights, two context vectors ($\mathbf{C}_1$ and $\mathbf{C}_2$) are obtained for the MFCC-HLSTM and LPCC-HLSTM network, respectively (Figure 3.6). Finally, they are fused to get the context vector ($\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$).

3.1.4 Classification

The context vector **C** of the SFF-HLSTM and PFF-HLSTM models are fed to a dense layer followed by a softmax activation function to get the decision output. The softmax gives the probability distribution over different classes. It is calculated as follows:

$$\mathbf{y}^{(i)} = softmax\left(\mathbf{W}_d \cdot \mathbf{C} + \mathbf{b}_d\right) \tag{3.20}$$

where [$\mathbf{W}_d \in \mathbb{R}^{d_d}, \mathbf{b}_d \in \mathbb{R}$] are the learnable weight vector and bias of the dense layer. $\mathbf{y}^{(i)}$ is the softmax probability for i^{th} class. We have used a multi-class cross-entropy loss function ($\mathcal{L}$) to train the model parameters and minimize the loss. The total cross-entropy loss for all the training data is calculated as follows:

$$\mathcal{L} = -\sum_{j=1}^{N}\sum_{i=1}^{m} t_j^{(i)} \cdot \log\left[\mathbf{y}_j^{(i)}\right] \tag{3.21}$$

where m is the number of classes, N is the total training data, $t_j^{(i)}$ is the true label for i^{th} class, and j^{th} data point.

3.2 Experimental Results and Discussion

3.2.1 Databases

In this work, two publicly available PCG databases, PhysioNet/ Computing in Cardiology Challenge (CICC) 2016 [13, 144], and heart sound murmur (HSM) [59] are used for experimental purposes. The details of the CICC 2016 database is given in the previous chapter 2.2.1. A brief detail of the HSM database is as follows.

3.2.1.1 HSM Database

In the HSM database, the PCG recordings were collected from random sources, such as books (Auscultation Skills CD, Heart Sound Made Easy) and websites (48 different websites provided the data, including Washington, Texas, 3M, Michigan, and other websites) [59]. It consists of a total 1000 number of PCG recordings. They are equally divided into five classes, i.e., normal (N), aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP). The data were resampled to 8000 Hz. They were converted to a mono channel using cool edit software. This database is used for multi-class classification of PCG signals, i.e., N, AS, MS, MR, or MVP.

3.2.2 Evaluation Scheme and Performance Measures

CICC 2016 database is used for binary classification of PCG segments as healthy or pathological. After preprocessing and blind segmentation, there are 102723- healthy and 31757- pathological PCG segments. To handle the imbalanced nature of the database, we train our model using a balanced dataset (*Bal-DS*). *Bal-DS* comprises 31757- pathological PCG segments and 31757- healthy PCG segments (selected using a simple random sampling algorithm). The *rest* 70966 healthy PCG segments are used for testing with the trained model. The *Bal-DS* is divided into five-fold and represented as follows:

$$\textit{Bal-DS} = [(train_i, test_i),\ i \in [1, 5]] \tag{3.22}$$

According to equation 3.22, for each fold i, $train_i$ is used for training the model, and $test_i$ and *rest* data are used to evaluate the model. So, ultimately, we have used all the CICC 2016 databases for each fold.

The HSM database is used for the multi-class classification of HS segments. After preprocessing and blind segmentation, it contains 400 HS segments for each N, AS, MVP, MR, and MS class. For both the databases, the model's performance is computed using precision (Pr), recall (Re), F1- score (F1), and accuracy (Acc) [156]. All the performance measures are obtained by using a five-fold cross-validation.

Table 3.1: The OA (%) results for different MFCC filter banks (K) and for different LPCC orders (P) to detect HMs.

K	OA (%)	P	OA (%)
20	94.69 ± 0.42	15	93.02 ± 0.44
30	$\mathbf{95.60 \pm 0.14}$	20	93.96 ± 0.47
40	95.45 ± 0.09	25	$\mathbf{94.57 \pm 0.24}$
50	95.16 ± 0.12	30	94.50 ± 0.41

3.2.3 Network Architecture

Each PCG segment **C**$[r]$ with 2s duration is split into 79 frames using a 50 ms Hamming window with a 50% overlap. Then, MFCC and LPCC features are extracted from each frame or timestamp. In this work, we have proposed a novel HLSTM network-based feature fusion algorithm that projects the MFCC and LPCC features to a joint feature space to improve the HMs detection. So, the value of the optimum number of filter banks of MFCC (K) and the order of the LPCC (P) depends on the LSTM architecture. So, we have experimented with fixing the values of K and P for the proposed work. Table 3.1 shows the overall accuracy (OA) results for different values of K and P to detect HVDs. All the results are five-fold cross-validated. It can be observed from Table 3.1 that 30 MFCC filter banks and 25 order LPCC show better OA (%) to detect HVDs. Finally, they are used for the proposed SFF-HLSTM and PFF-HLSTM models.

SFF Model: There are 79 input frames or sequences that are fed as input to the first layer ($\mathbf{L}_1$) of the HLSTM network. The dimension of each input sequence x_i is 55 ($d_i = 55$). These input sequences are learned by the $\mathbf{L}_1$ with a hidden vector of size 150 ($d_{h1} = 150$). Then, the output vectors of $\mathbf{L}_1$ are fed to the $\mathbf{L}_2$ to learn a higher-level abstract representation of dimension 100 ($d_{h2} = 100$). The output vectors of $\mathbf{L}_2$ emphasize intra and inter-frame diagnostic information, thereby reducing frame redundancies. Later, these output vectors are fed to a self-attention module with a latent space size of 100 ($d_a = 100$) to generate a context vector (**C**). This vector attains the most discriminative high-level representation. Finally, the 100-dimensional context vector is fed to the dense layer and softmax function to classify HMs.

PFF Model: There are 79 input frames or sequences fed as input to the two HLSTM networks. For the MFCC-HLSTM network, the dimension of each input sequence x_{mi} is 30 ($d_{mi} = 30$). For the LPCC-HLSTM network, the dimension of each input sequence x_{li} is 25 ($d_{li} = 25$). These input sequences of both the networks are learned by their respective $\mathbf{L}_1$ with a hidden vector

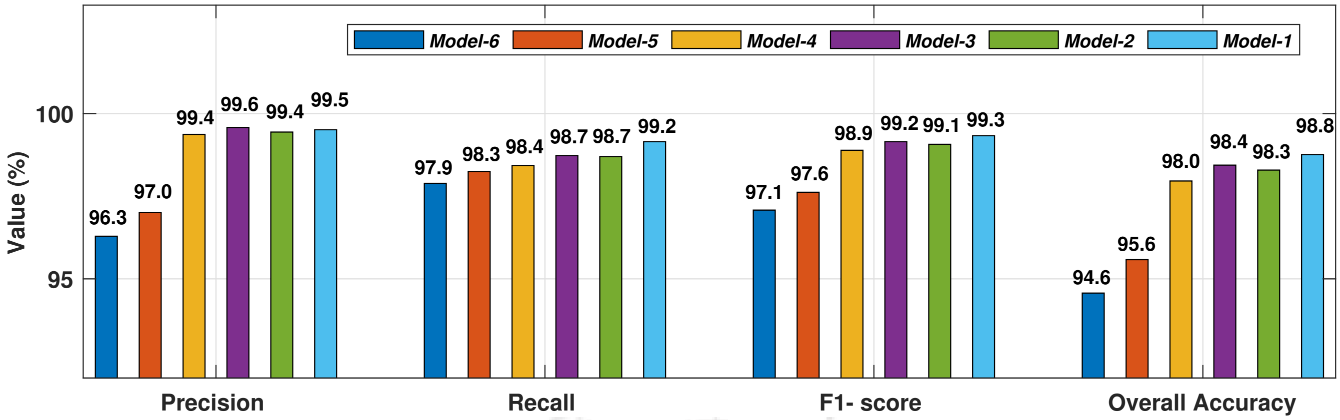

Figure 3.8: The performance comparison of the proposed models with four other models. Model-1: Proposed SFF-HLSTM model, Model-2: Proposed PFF-HSLTM model, Model-3: SFF-LSTM model , Model-4: PFF-LSTM, Model-5: MFCC-LSTM and Model-6: LPCC-LSTM.

of size 150 ($d_{h1} = 150$). Then, the output vectors of $\mathbf{L}_1$ are fed to their respective $\mathbf{L}_2$ to learn a higher-level abstract representation of dimension 100 ($d_{h2} = 100$). Later, these output vectors are fed to their respective self-attention module with latent space size of 100 ($d_a = 100$) to generate two context vectors ($\mathbf{C}_1$ and $\mathbf{C}_2$). We appended $\mathbf{C}_1$ and $\mathbf{C}_2$ to get one context vector $\mathbf{C}$. Finally, the 200-dimensional context vector is fed to the dense layer and softmax function to classify HMs.

We deployed the batch normalization layer and dropout layer (dropout ratio = 0.4) after the $\mathbf{L}_1$ layer to prevent the models from over-fitting. The proposed SFF and PFF models' optimal hyperparameter selection is determined based on experiments. This study uses an Adam optimizer with a batch size of 200 for model training. The learning rate of the model is 0.005, and the total number of epochs to train the model is 350. Python programming language with PyTorch library implements the deep learning models. A Titan P100 GPU server is used to perform the experiments.

3.2.4 Effectiveness of the Proposed Models

We have analyzed the performances of the proposed SFF-HLSTM model (Model-1) and PFF-HSLTM model (Model-2) with the SFF-LSTM model (Model-3), PFF-LSTM (Model-4), MFCC-LSTM (Model-5) and LPCC-LSTM (Model-6). The self-attention module is included for all the models. We have used the CICC 2016 database to compare these six models. Figure 3.8 shows the comparison results for all models' Pr, Re, F1, and OA. All the results are five-fold cross-validated. It can be observed from Figure 3.8 that fused-feature models (Model-3 and 4)

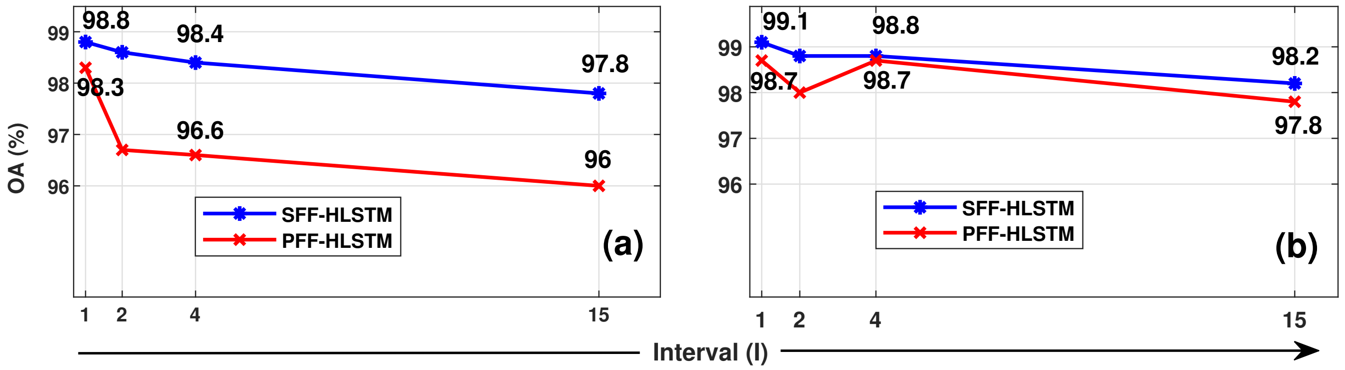

Figure 3.9: Variation of overall accuracy (%) of the proposed SFF-HLSTM and PFF-HLSTM models for different interval lengths. (a) Results for CICC 2016 database and (b) Results for the HSM database.

show better performance results than individual feature models (Model-5 and 6). Model-3 and 4 show a 3% increase in Pr, a 1% increase in Re and F1, and almost a 3% increase in OA than Model-5 and 6. It indicates that fused features capture more information related to HVDs than individual features. Moreover, the proposed fused feature models using the HLSTM network (Model-1 and 2) show slightly better performance than the LSTM network models (Model-3 and 4). Model-1 and 2 show a 0.5% increase in OA than Model-3 and 4. It indicates that proposed fused-feature models using HLSTM can capture the temporal information of HVDs in different higher-level abstractions. Among Model-1 and 2, Model-1 (SFF-HLSTM) shows slightly better performance results than Model-2 (PFF-HLSTM). It implies that initial feature fusion in Model-1 captures the multi-scale temporal information better than in Model-2.

We experimented with fixing the interval (**I**) for the proposed SFF-HLSTM and PFF-HLSTM models. Figure 3.9 shows the OA (%) results for different intervals (**I**) for both the CICC 2016 and HSM databases. It can be observed from Figure 3.9 that both the SFF-HLSTM and PFF-HLSTM models show the best OA (%) at **I** of 1 for both the databases. The OA (%) decreases with the increase in the **I** value. It indicates that all the outputs of $\mathbf{L}_1$ contain meaningful relative temporal information among each timestamp for the detection of HVDs and are fed as input to the $\mathbf{L}_2$ (Figure 3.5).

3.2.5 Results on the CICC 2016 Database

Table 3.2 shows the fold-wise and average HVD detection performance results for the proposed SFF-HLSTM and PFF-HLSTM models using the CICC 2016 database. All the results are five-fold cross-validated. The SFF-HLSTM model shows an OA (%) of 98.76, average Pr (%) of 99.51,

Table 3.2: The fold-wise and average performance measures (%) for the proposed SFF-HLSTM and PFF-HLSTM models using the CICC 2016 database.

Model	Fold	Pr (%)	Re (%)	F1 (%)	Acc (%)
SFF-HLSTM	1	99.69	98.95	99.32	98.75
	2	99.48	99.19	99.33	98.77
	3	99.62	99.06	99.34	98.78
	4	99.59	99.21	99.40	98.89
	5	99.15	99.36	99.25	98.62
	Avg.	**99.51 ± 0.21**	**99.15 ± 0.16**	**99.33 ± 0.05**	**98.76 ± 0.10**
Model	**Fold**	**Pr (%)**	**Re (%)**	**F1 (%)**	**Acc. (%)**
PFF-HLSTM	1	99.73	98.63	99.18	98.49
	2	99.25	98.95	99.10	98.34
	3	99.43	98.55	98.99	98.14
	4	99.34	98.88	99.11	98.36
	5	99.47	98.51	98.98	98.13
	Avg.	**99.44 ± 0.18**	**98.64 ± 0.14**	**99.07 ± 0.08**	**98.29 ± 0.15**

Table 3.3: Comparison of the proposed method results with the existing state-of-the-art methods in the literature for the CICC 2016 database.

Related Work	Feature-extraction Method	Classification Method	OA (%)
Potes *et al.*, 2016 [100]	Time-Frequency	AdaBoost, CNN	86.02
Noman *et al.*, 2019 [54]	MFCC, Time-Frequency	Markov Model	91.20
Zhang *et al.*, 2019 [96]	Temporal, Quasi-Periodic Features	LSTM	94.66
Suyi *et al.*, 2021 [121]	No	CNN, GRU	95.50
Milani *et al.*, 2021 [70]	MFCC, Time, Frequency	LDA, ANN	93.33
Mei *et al.*, 2021 [157]	Wavelet Scattering	SVM	93.64
This Study	MFCC, LPCC	SFF-HLSTM	**98.76**
This Study	MFCC, LPCC	PFF-HLSTM	**98.29**

average Re (%) of 99.15, and average F1 (%) of 99.33. The PFF-HLSTM model shows an OA (%) of 98.29, average Pr (%) of 99.44, average Re (%) of 98.64, and average F1 (%) of 99.07. Among the proposed models, the SFF-HLSTM model shows slightly better performance results than the PFF-HLSTM model. It can also be observed from Table 3.2 that the proposed models have a slight standard deviation in all the performance measures across the five-folds. It shows the excellent generalization of the proposed models for detecting HVDs.

The comparison results of the proposed method with the existing state-of-the-art methods are shown in Table 3.3. The methods presented in Table 3.3 have used the CICC 2016 database. It can be observed that both the proposed models outperform the existing methods in terms of OA (%). The proposed models outperform the time, frequency, time-frequency, and wavelet-based

Table 3.4: The classification results both class-wise and overall using the proposed SFF-HLSTM and PFF-HLSTM models for the HSM database.

Model	Class	Pr (%)	Re (%)	F1 (%)	Acc. (%)
SFF-HLSTM	N	99.01	99.66	99.33	99.03
	MVP	99.07	98.79	98.92	98.71
	MR	99.56	99.07	99.30	98.71
	AS	99.43	99.76	99.53	99.35
	MS	98.24	98.58	98.40	99.68
	Avg.	**99.06** ± **0.51**	**99.17** ± **0.52**	**99.10** ± **0.45**	**99.10** ± **0.42**
Model	**Class**	**Pr (%)**	**Re (%)**	**F1 (%)**	**Acc. (%)**
PFF-HLSTM	N	99.10	98.48	98.46	99.03
	MVP	99.10	98.48	98.46	97.42
	MR	98.66	98.15	98.36	99.03
	AS	99.54	99.29	99.41	99.03
	MS	97.28	98.57	97.90	99.03
	Avg.	**98.74** ± **0.87**	**98.59** ± **0.42**	**98.52** ± **0.55**	**98.71** ± **0.72**

Table 3.5: Comparison of the proposed method results with the existing state-of-the-art methods in the literature for the HSM database.

Related Work	Feature-extraction Method	Classification Method	OA (%)
Son *et al.*, 2018 [59]	DWT	SVM	91.60
Son *et al.*, 2018 [59]	DWT, MFCC	SVM	97.90
Ghosh *et al.*, 2019 [63]	WSST	RF	95.13
Ghosh *et al.*, 2020 [64]	Chirplet Transform	Composite	98.33
This Study	MFCC, LPCC	SFF-HLSTM	**99.10**
This Study	MFCC, LPCC	PFF-HLSTM	**98.71**

feature approaches with a large margin of nearly 10% in OA. The proposed models also show better OA (%) results than MFCC and fusion of MFCC with other features-based approaches.

3.2.6 Results on the HSM Database

Table 3.4 shows the performance of multi-class classification of HMs using the HSM database. It can be observed that both proposed SFF-HLSTM and PFF-HLSTM models show impressive both class-wise and overall Pr, Re, F1, and OA measures. The proposed method attains OA of 99.10 % and 98.71 % using the SFF-HLSTM and PFF-HLSTM models. For both the proposed models, it can be observed that the MVP class shows a slight reduction of Acc. (%) than other classes. It may be because of the morphological similarity of the MVP class with the MR class. It can also be observed from Table 3.4 that the proposed models have a slight standard deviation in all the performance measures across the five-folds. It shows the excellent generalization of the

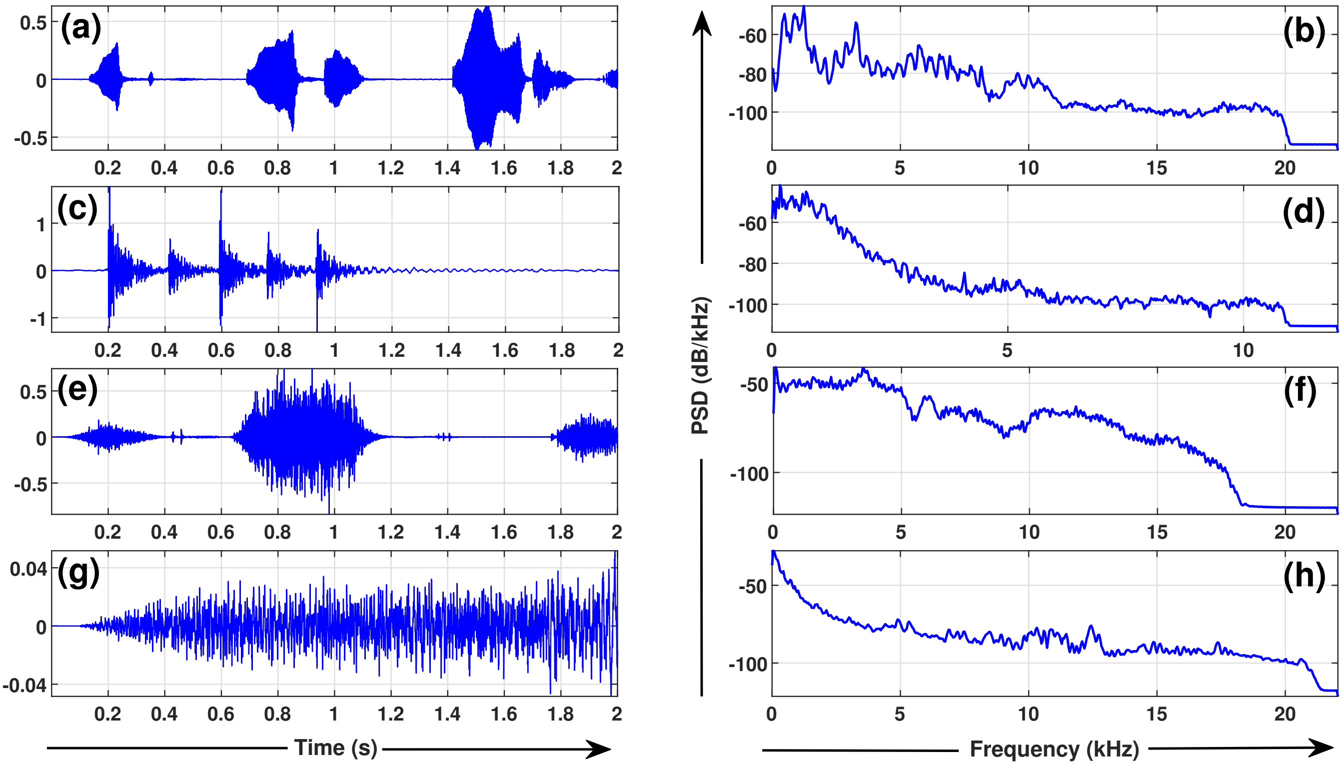

Figure 3.10: Different types of noises with their power spectral densities (PSDs). (a,b) Baby crying and its PSD, (c,d) Door knocking and its PSD, (e,f) Human sniffing and its PSD, and (g,h) Urban square atmospheric noise and its PSD.

proposed models for classifying HMs.

The efficacy of the proposed method is verified by comparing it with the current existing methods. It can be seen from Table 3.5 that the proposed method outperforms the existing methods in terms of an OA of 99.10 %. The methods presented in Table 3.5 have used the HSM database. The improved performance indicates that the proposed feature fusion models using the HLSTM network capture distinct temporal information among classes to enhance the detection of HMs.

3.2.7 Performance of Proposed Models in Noisy Conditions

The proposed models' robustness is evaluated by randomly adding an environmental or body noise to the PCG signal. These eight noises include an urban square atmosphere, urban street kids playing, a sick man coughing, a sick man sneezing, a person nose-blowing, a human sniffling, door knocking, and a baby crying[1]. Figure 3.10 shows some of these noises with their respective power spectral densities. It can be observed that there is a significant overlap of the frequency

---

[1] [Online Available]: https://mixkit.co/free-sound-effects/static/

Table 3.6: The overall accuracy (OA) of the proposed SFF-HLSTM and PFF-HLSTM models under different SNR levels of environmental and body noises.

SFF-HLSTM	OA (%)	PFF-HLSTM	OA (%)
0 dB SNR	93.41	0 dB SNR	92.18
10 dB SNR	96.26	10 dB SNR	95.28
20 dB SNR	97.42	20 dB SNR	95.28
Without Noise	99.10	Without Noise	98.71

Table 3.7: Comparison of the proposed models' NMP, ART, and OA for CICC 2016 and HSM databases.

Database	Proposed Model	NMP	ART (ms)	OA (%)
CICC 2016	SFF-HLSTM	225803	2.39	98.76
	PFF-HLSTM	418604	3.84	98.29
HSM	SFF-HLSTM	226106	2.61	99.10
	PFF-HLSTM	419207	3.86	98.71

content of these noises with the PCG signal frequency band (0 Hz to 500 Hz). We have analyzed the performance of the proposed models by adding noises of different signal-to-noise ratio (SNR) levels to the PCG signal. Table 3.6 shows the OA (%) of the proposed SFF-HLSTM and PFF-HLSTM models using different SNR levels for the HSM database. It can be observed from Table 3.6 that OA (%) for both models increases with the increase in SNR (dB). For 20 dB SNR, it shows an OA of 97.42 % and 95.28 % for SFF-HLSTM and PFF-HLSTM models, respectively. This test reveals the proposed models' robustness for different types of noises at different SNR (dB) levels.

3.2.8 Complexity of the Proposed Models

Table 3.7 shows the comparison of the number of model parameters (NMP), the average run time (ART), and the OA of the proposed models on the CICC 2016 and the HSM databases. The ART is measured using a Tesla P100 16 GB graphics processor. It can be observed that the proposed SFF-HLSTM model consists of fewer NMP and less ART than the PFF-HLSTM model on both databases. The SFF-HLSTM model also provides slightly better OA than the PFF-HLSTM model for both databases. Thus, the acceptable ART (ms) and the impressive OA (%) of the SFF-HLSTM model make it suitable for rapid and improved preliminary diagnosis of HVDs.

3.3 Summary

In this chapter, a novel acoustic feature-fusion using an attention-based HLSTM network is presented to classify HMs. MFCC and LPCC features capture the complementary aspects of the PCG segment to distinguish HMs. Further, these two features' temporal and relative-temporal variations are exploited using the HLSTM network with an attention mechanism. The experimental results verify that the proposed method outperforms the existing baseline methods. It shows an OA of 98.76% and 99.10% using the CICC 2016 and HSM databases, respectively. The robustness of the proposed method is evaluated by adding different body and environmental noises to the PCG signal. The proposed SFF-HLSTM network has an NMP of 225803 and an ART of 2.39 ms using the CICC 2016 database for binary classification. Similarly, the proposed SFF-HLSTM network has an NMP of 226106 and an ART of 2.61 ms using the HSM database for multi-class classification of HMs. The proposed method's excellent OA and timely response make it reliable to use in primary healthcare units to detect HMs.

4

SWT and Attention-based HLSTM Networks for HMs Classification

Contents

In the last chapter 3, we have discussed an acoustic feature-fusion approach to capture the HM patterns from the PCG signal. Further, an attention-based HLSTM network exploits the temporal and relative-temporal information among these features to classify HMs. In this chapter 4, we explore another transfer-domain approach followed by the DL model to classify HMs.

Researchers have used wavelet-based features to classify HMs [59, 61, 62, 72]. The wavelet methods decompose the PCG signal and localize the signal both on a time-frequency scale. The significant energy concentration of fundamental HSs (S1 and S2 sound) lies in the low-frequency range of $(20 - 200)$ Hz or higher subbands of the wavelet decomposition. Similarly, the pathological HS energy concentration presents up to 900 Hz. In contrast, the classical wavelet-based methods lose resolution at higher subbands (or coarse scales) and lack shift-invariant properties. As a result, most of the energy content and morphological patterns of the PCG signal, which is crucial for anomaly detection, are mostly missing in the coarse scales of those wavelet transforms (WTs). Therefore, this chapter proposes a stationary wavelet transform (SWT) based PCG signal decomposition for adequately capturing PCG patterns in coarse scales. SWT is shift-invariant and maintains the same number of samples in all the subbands. Figure 4.1 compares conventional DWT and SWT for a five-level decomposition of an MVP PCG segment. It can be seen from Figure 4.1 that D_4, D_5, and A_5 contain useful information related to fundamental HSs and murmur patterns, which SWT captures. In contrast, DWT cannot capture it properly due to fewer samples at the coarse scales.

In the literature, after wavelet decomposition, various global features such as relative-subband energy, skewness, and kurtosis are extracted from these WT subbands and fed to a supervised ML classifier to classify HMs [61, 64, 65]. However, an HM shows different temporal and relative-temporal variations within and across different subbands. So, effectively handling these subbands' temporal information is essential for the robust detection and classification of HMs. Therefore, we have proposed subband-specific hierarchical long short-term memory (HLSTM) networks to utilize the temporal and relative-temporal information across each subband. Further, for selecting relevant clinical information from each subband and across the subbands, we have proposed intra-subband and inter-subband attention modules.

The rest of the chapter is organized as follows: Section 4.1 describes the proposed SWT and attention-based-HLSTM networks to classify HMs. Section 4.2 discusses the obtained experimental results. Section 4.3 ends with the summary.

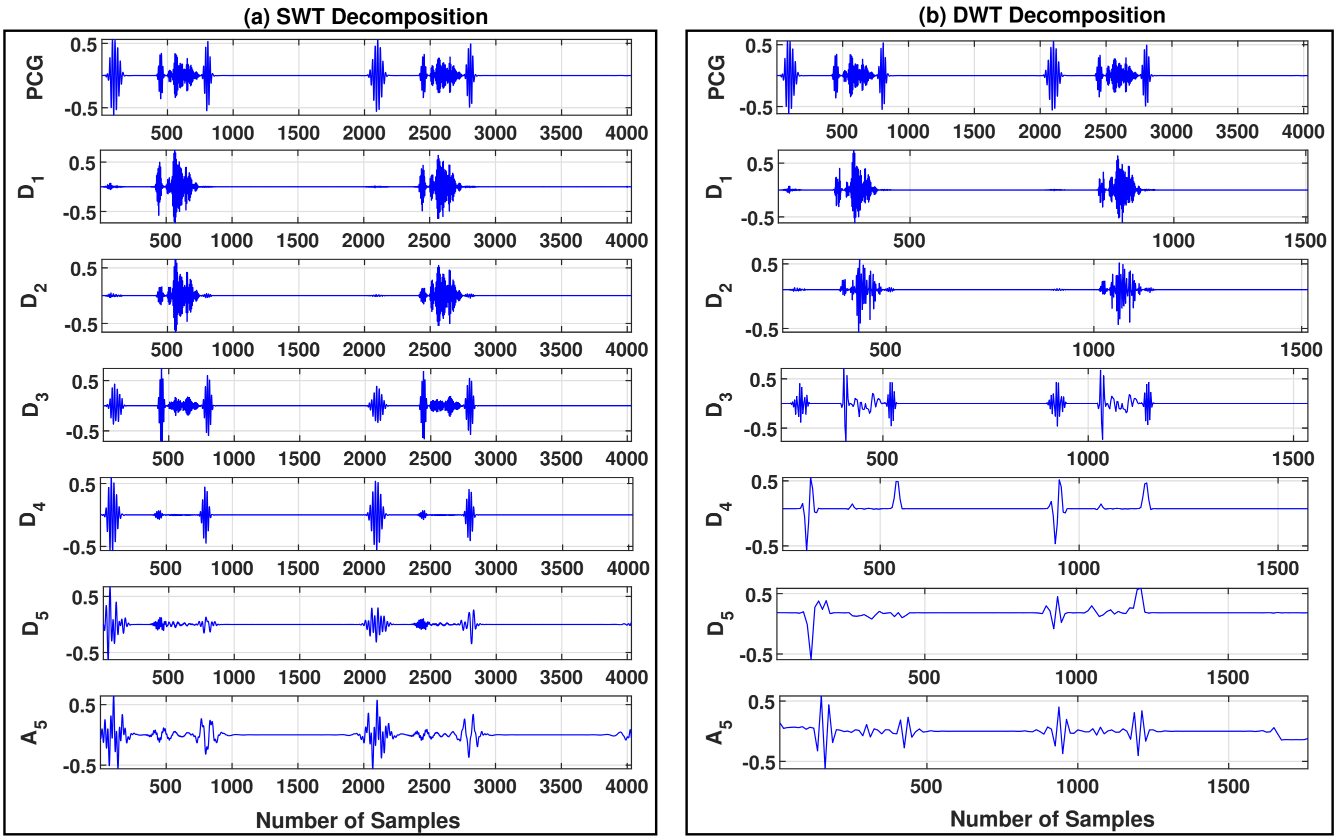

Figure 4.1: The decomposition of a mitral valve prolapse (MVP) PCG segment using (a) SWT and (b) DWT.

4.1 Proposed SWT and HLSTM Network for HMs Classification

The proposed method comprises five stages. First, preprocessing and segmentation are performed on the raw PCG signal, which we have discussed in section 3.1.1. Then, each PCG segment is decomposed using SWT. Third, deployment of subband-specific HLSTM networks for each decomposed subband. Fourth, intra-subband attention layers follow inter-subband attention layers to obtain the context vector for each PCG segment. Finally, the context vector is given to a classification module to detect HMs. Figure 4.2 shows the block diagram of the proposed method.

4.1.1 SWT-based decomposition of PCG segment

After preprocessing and segmentation, each 2 s PCG segment is decomposed using an SWT-based decomposition. For SWT decomposition, the input signal, $\mathbf{x}[n]$, the length must be multiples of 2^L (L is the order of decomposition level). Otherwise, the signal is zero-padded. So, the SWT

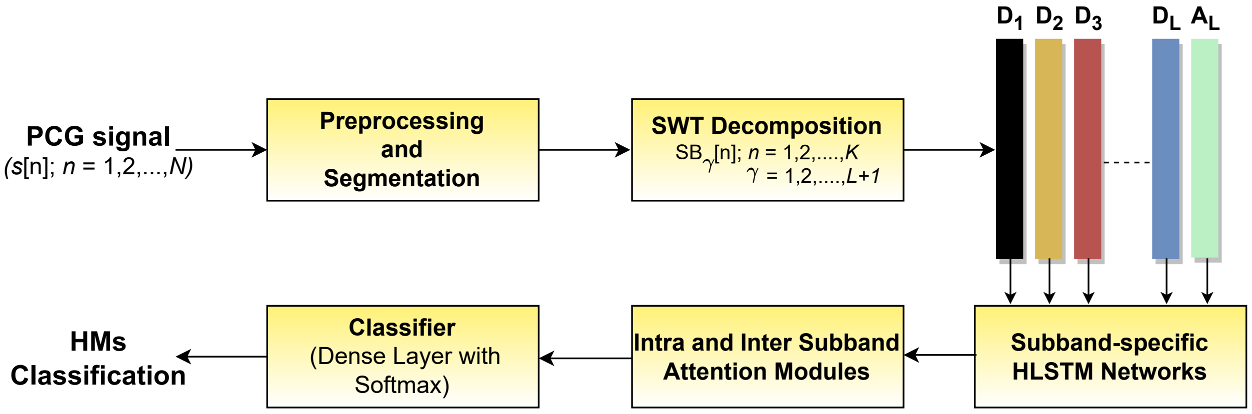

Figure 4.2: Block diagram of the proposed SWT and attention-based HLSTM networks to classify HMs.

decomposition for level γ is obtained as follows [158].

$$\mathbf{A}_{\gamma+1}[n] = \sum_k h_\gamma[n-k] \cdot \mathbf{A}_\gamma[k] \tag{4.1}$$

$$\mathbf{D}_{\gamma+1}[n] = \sum_k g_\gamma[n-k] \cdot \mathbf{A}_\gamma[k] \tag{4.2}$$

$$h_{\gamma+1} = \text{upsample}(h_\gamma, 2), \quad g_{\gamma+1} = \text{upsample}(g_\gamma, 2) \tag{4.3}$$

where $\mathbf{A}_{\gamma+1}[n]$ and $\mathbf{D}_{\gamma+1}[n]$ are the approximation and detail coefficients for γ^{th} level SWT decomposition. Where h_γ and g_γ are the low-pass and high-pass filters for level γ, respectively. SWT does not downsample its detail and approximation coefficients for the next level but rather upsample its respective filters by a factor of 2. Figure 4.3 shows a two-level decomposition of input signal $\mathbf{x}[n]$ using SWT. The subbands for an L- order decomposition is represented as follows:

$$\mathbf{SB}_\gamma[n] = \begin{cases} \mathbf{D}_\gamma[n], & \gamma = 1, 2, ..., L \\ \mathbf{A}_\gamma[n], & \gamma = L+1 \end{cases} \tag{4.4}$$

where $\mathbf{SB}_\gamma[n]$; $(\gamma = 1, 2, ..., L+1)$ are the subbands for L- order decomposition. $n = 1, 2, 3, ..., K$ is the number of coefficients for each subband.

This work uses a Biorthogonal wavelet 6.8 (bio6.8) as a basis function. This basis function is chosen because of its morphological similarity with the PCG signal. Furthermore, each PCG segment is decomposed into five levels. It is so because mostly the fundamental HSs and murmur

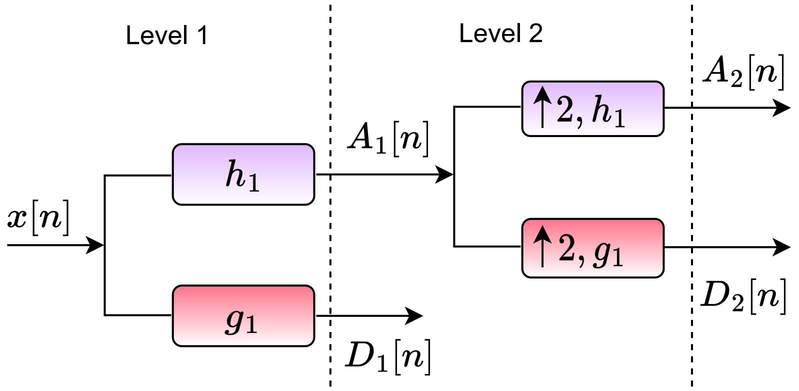

Figure 4.3: The two-level decomposition using SWT.

information are present in frequency bands (> 20Hz). There is no need for further decomposition at a coarse scale. It can be observed from Figure 4.1 that murmur information are present in D_1, D_2, and D_3 subbands. Whereas D_3, D_4, D_5, and A_5 mostly contain the fundamental HS components.

4.1.2 HLSTM Encoding of Decomposed Subband

The decomposed subbands of each PCG segment are fed as input to the proposed network to classify the HMs. Figure 4.4 shows the proposed attention-based HLSTM network architecture. The decomposed subbands are encoded by their respective subband-specific HLSTM networks. Further, the encoded vectors of each subband-specific HLSTM network are fed to their respective intra-subband attention modules, followed by an inter-subband attention module to weigh aggregate the encoded vectors based on their clinical relevance to classify HMs.

The advantage of using subband-specific HLSTM networks is that it exploits the temporal information in each subband. Further, it utilizes the relative temporal information present among the basic PCG waveforms (i.e., S1 sound, systole, S2 sound, and diastole) in each subband for detecting HMs. The HLSTM network is a variant of LSTM [154]. Two LSTM layers are stacked to design the HLSTM network. Figure 4.5 shows the network architecture of HLSTM for γ^{th}- subband. Let $\mathbf{X}^\gamma = \{x_1^\gamma, x_2^\gamma, ..., x_t^\gamma, ..., x_T^\gamma\}$ are the γ^{th}- subband input sequences that are fed to the first layer HLSTM ($\mathbf{L}_1$). Here, $\gamma \in \{1, 2, ..., L+1\}$ represents the number of subbands. $x_t^\gamma \in \mathbb{R}^{d_{in}}, t \in \{1, 2, ..., T\}$ represents the t^{th} time-stamp input for γ^{th}- subband. d_{in} is the dimension of each time-stamp, and T is the total number of time-stamps. The outputs of $\mathbf{L}_1$ are represented

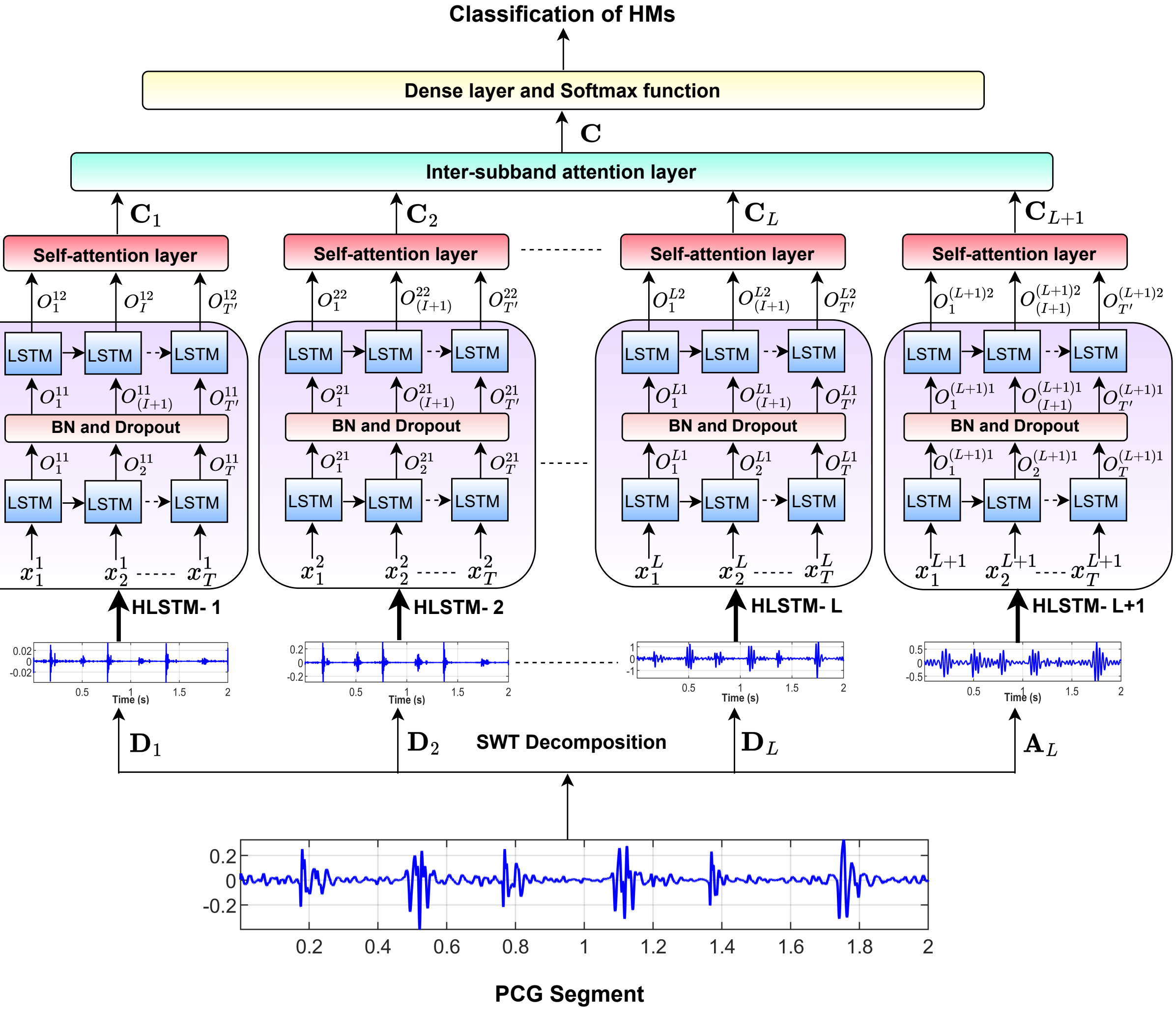

Figure 4.4: The proposed network architecture using subband-specific HLSTM networks with intra-subband and inter-subband attention modules for the automated classification of HMs.

as follows:

$$\mathbf{O}^{\gamma 1} \in \{O_1^{\gamma 1}, O_2^{\gamma 1}, ..., O_t^{\gamma 1}, ..., O_T^{\gamma 1}\} \tag{4.5}$$

where $\mathbf{O}^{\gamma 1} \in \mathbb{R}^{d_{L1}}$ are the outputs of $\mathbf{L}_1$ for γ^{th}- subband and d_{L1} is the dimension of each output. So, $\mathbf{L}_1$ exploits the temporal information content by the subband sequences.

The second layer of HLSTM ($\mathbf{L}_2$) utilizes the relative temporal information in the basic PCG waveforms for each subband. It is performed by feeding the output of $\mathbf{L}_1$ at an interval of I to the input of $\mathbf{L}_2$. So, the optimum value of I is computed experimentally. The input and output of $\mathbf{L}_2$ for γ^{th}- subband is represented as follows:

$$\mathbf{X}^{\gamma 2} \in \{O_1^{\gamma 1}, O_{(I+1)}^{\gamma 1}, O_{(2I+1)}^{\gamma 1}, ..., O_{(mI+1)}^{\gamma 1}, ..., O_{T'}^{\gamma 1}\} \tag{4.6}$$

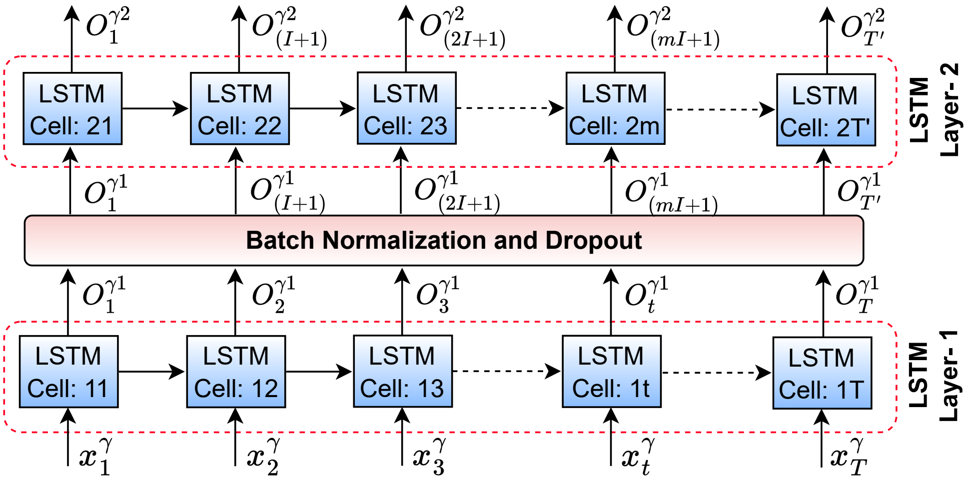

Figure 4.5: An HLSTM network for γ^{th}- subband.

$$\mathbf{O}^{\gamma 2} \in \{O_1^{\gamma 2}, O_{(I+1)}^{\gamma 2}, O_{(2I+1)}^{\gamma 2}, ..., O_{(mI+1)}^{\gamma 2}, ..., O_{T'}^{\gamma 2}\} \tag{4.7}$$

where $m = \{0, 1, 2, ..., [\frac{T}{I}] - 1\}$. $T' = [\frac{T}{I}] - 1\}$ is the last time stamp that is fed to the input of $\mathbf{L}_2$. $\mathbf{O}^{\gamma 2} \in \mathbb{R}^{d_{L2}}$ are the outputs of $\mathbf{L}_2$ for γ^{th}- subband and d_{L2} is the dimension of each output. For both the LSTM layers $\mathbf{L}_1$ and $\mathbf{L}_2$, the extraction of sequential information from the input timestamps is performed using LSTM cells.

Thus, the HLSTM utilizes each subband's temporal and relative-temporal information for detecting HMs. To further capture the diagnostic information and improve the representation power of the HLSTM encoded vectors, we have proposed intra-subband attention layers for each subband, followed by an inter-subband attention layer.

4.1.3 Intra-subband and Inter-subband Attention Modules

The attention mechanism has shown encouraging outcomes for recognition problems by providing more weights to the relevant portions of an input sequence [155, 159]. So, the obtained encoded vectors of HLSTM are fed to their respective intra-subband attention layers. The intra-subband attention layers weight aggregates the encoded vectors of each subband-specific HLSTM network to improve the HMs detection. Figure 4.6 shows the network architecture of a self-attention layer for γ^{th}- subband. The obtained outputs $\mathbf{O}^{\gamma 2}$ of the HLSTM network for γ^{th}- subband is given to its

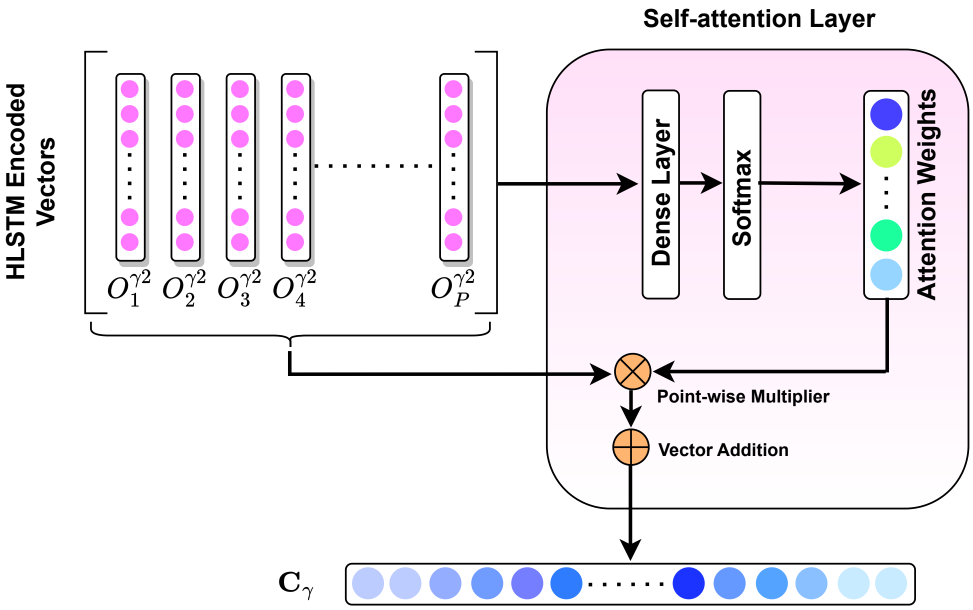

Figure 4.6: An intra-subband attention layer for γ^{th}- subband.

intra-subband attention layer to obtain scores. It is computed as follows:

$$S_p^\gamma = \tanh(\mathbf{W}^\gamma_{\text{Intra}} \cdot O_p^{\gamma 2} + \mathbf{b}^\gamma_{\text{Intra}}) \tag{4.8}$$

where $\mathbf{W}^\gamma_{\text{Intra}} \in \mathbb{R}^{d_{L2}}$ is the weight and $\mathbf{b}^\gamma_{\text{Intra}} \in \mathbb{R}$ is the bias of the intra-subband attention module for γ^{th}- subband. S_p^γ is the obtained score for output $O_p^{\gamma 2}$ at p^{th} time stamp of $\mathbf{L}_2$. Where $p = \{1, 2, ..., P\}$ and P is the total number of time stamps of $\mathbf{L}_2$. Further, the S_p^γ is input to a softmax function to get the attention weight for $O_p^{\gamma 2}$. It is formulated as follows:

$$\alpha_p^\gamma = \frac{\exp(S_p^\gamma)}{\sum_{p=1}^{P} \exp(S_p^\gamma)} \tag{4.9}$$

where α_p^γ is the attention weight for $O_p^{\gamma 2}$ for γ^{th}- subband. The obtained weight for an encoded vector is high if it preserves clinically relevant information and vice-versa. Finally, a context vector for γ^{th}- subband is obtained as follows:

$$\mathbf{C}_\gamma = \sum_{p=1}^{P} \alpha_p^\gamma \cdot O_p^\gamma \tag{4.10}$$

where $\mathbf{C}_\gamma \in \mathbb{R}^{d_{L2}}$ represents the context vector for γ^{th}- subband. There are six context vectors ($\gamma = L + 1 = 6$) which form an intra-subband attention matrix $\mathbf{A} \in \mathbb{R}^{d_{L2} \times 6}$ and is fed as input to the inter-subband attention module.

Further, the inter-subband attention module weight aggregates the context vectors of the respective subbands based on their importance in improving the performance to classify HMs. It resulted in a higher level context vector across all the subbands and formulated similar to the intra-subband attention layer, as discussed above.

$$S_\gamma = \tanh(\mathbf{W}_{\text{Inter}} \cdot \mathbf{C}_\gamma + \mathbf{b}_{\text{Inter}}) \tag{4.11}$$

$$\alpha_\gamma = \frac{\exp(S_\gamma)}{\sum_{\gamma=1}^{L+1} \exp(S_\gamma)} \tag{4.12}$$

$$\mathbf{C} = \sum_{\gamma=1}^{L+1} \alpha_\gamma \cdot \mathbf{C}_\gamma \tag{4.13}$$

where $\mathbf{C} \in \mathbb{R}^{d_{L2}}$ is the final context vector. It gives importance to specific subbands to improve classification performance for HMs detection. Where $\mathbf{W}_{\text{Inter}} \in \mathbb{R}^{d_{L2}}$ is the weight and $\mathbf{b}_{\text{Inter}} \in \mathbb{R}$ is the bias of the inter-subband attention module.

4.1.4 Classification of HMs

Finally, the $\mathbf{C}$ is given as input to a fully-connected dense layer followed by a softmax function to estimate the probability of each class. It is formulated as follows:

$$\mathbf{y}_{\text{p}} = \arg\max_k \{\text{softmax}(\mathbf{W}_d \cdot \mathbf{C} + \mathbf{b}_d)\} \tag{4.14}$$

where $\mathbf{W}_d \in \mathbb{R}^{K \times d_{L2}}$ is the weight and $\mathbf{b}_d \in \mathbb{R}^K$ is the bias of the dense layer. K is the total number of classes. Where $k = 1, 2, ..., K$ and $\mathbf{y}_{\text{p}}$ is the predicted class label.

The parameters of the proposed network are trained by computing the cross-entropy loss function and which is formulated as follows:

$$\mathcal{L} = -\sum_{n=1}^{N} \sum_{k=1}^{K} \mathbf{y}_{\text{t}}^{n,k} \cdot \log[\mathbf{y}_{\text{p}}^{n,k}] \tag{4.15}$$

where $\mathbf{y}_{\text{t}}^{n,k}$ and $\mathbf{y}_{\text{p}}^{n,k}$ are the true and predicated labels for n^{th} training data. N is the total number of training data. $\mathcal{L}$ is the multiclass cross-entropy loss function.

4.2 Experimental Results and Discussion

4.2.1 Databases

In this work, two publicly available PCG databases, PhysioNet/ Computing in Cardiology Challenge (CICC) 2016 [13, 144], and heart sound murmur (HSM) [59] are used for experimental purposes. The details of the CICC 2016 database is given in the chapter 2.2.1. The details of the HSM database is given in the previous chapter 3.2.1.1.

4.2.2 Performance Measures

The performance of the proposed method for the CICC 2016 database is evaluated by using four measures, i.e., overall sensitivity (OSe), overall specificity (OSp), overall F1- score (OF1), and overall accuracy (OA). The performance measures of the proposed method for the HSM database are evaluated using class-wise sensitivity (Se), class-wise specificity (Sp), class-wise F1- score (F1), and OA [156]. All the results for both databases are stratified five-fold cross-validated. The cross-validation provides generalizable results by preventing over-fitting to a specific subset of the database.

4.2.3 Network Architecture

The subbands $\mathbf{SB}_\gamma(\gamma = 6)$ are fed as input to their respective HLSTM networks. Each PCG subband with 4000 samples is split into 39 timestamps ($T = 39$) using a rectangular window with 50% overlap. So, each time-stamp is having 200 samples are fed to $\mathbf{L}_1$ of HLSTM network of size 100 ($d_{L1} = 100$) to obtain the output vectors $\mathbf{O}^{\gamma 1}$. Then, these $\mathbf{O}^{\gamma 1}$ are fed to $\mathbf{L}_2$ of HLSTM network of size 50 ($d_{L2} = 50$) at an interval of I. It results encoded vectors $\mathbf{O}^{\gamma 2}$. These encoded vectors are given as input to their corresponding intra-subband attention layer having a dimension of 50 ($d_{\text{Intra}} = d_{L2} = 50$) to obtain the context vectors ($\mathbf{C}_\gamma$) for each subband. Further, these context vectors are given as input to the inter-subband attention layer with a dimension of $d_{\text{Inter}} = d_{L2} = 50$ to generate the final context vector ($\mathbf{C}$). Finally, the 50- dimensional $\mathbf{C}$ is given as input to a softmax function preceded by the dense layer for detecting HMs. The batch normalization (BN) is applied between $\mathbf{L}_1$ and $\mathbf{L}_2$ and also after the intra-subband and inter-subband attention layer to increase the model's speed of convergence. A dropout layer with a dropout rate of 0.2 is deployed after each layer to prevent the proposed model from over-fitting issues. The optimum values for

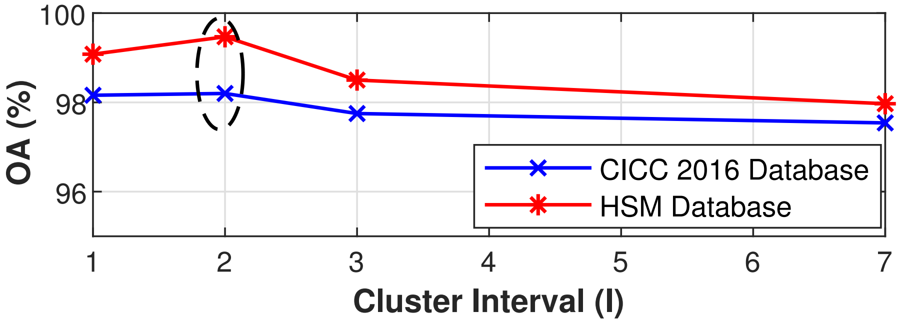

Figure 4.7: The OA (%) values for different cluster intervals using the CICC 2016 and HSM databases.

the hyper-parameters of the proposed model are tuned through experiments.

In this study, we have deployed an Adam optimizer with a learning rate of 0.005. The model's parameters are trained using back-propagation through time (BPTT) [160]. The batch size is set to 256, and the model is trained for 350 epochs. The deep learning models are implemented using Python programming language with the PyTorch library. The experiments are performed by using a Titan P100 GPU server.

4.2.4 Fixing the Interval Value of HLSTM Network

The optimum value of interval I at which the output of $\mathbf{L}_1$ is fed as input to the $\mathbf{L}_2$ of the subband-specific HLSTM networks are fixed through experimentation. Figure 4.7 shows the proposed model's OA(%) for different intervals of I. It can be seen that the proposed model shows the highest OA(%) for both the CICC 2016 and HSM databases at I of 2. It implies that rather than taking all the timestamp outputs of $\mathbf{L}_1$, the outputs at an interval of 2 to the input of $\mathbf{L}_2$ provide better co-relation among the timestamps to improve the HMs classification.

4.2.5 Efficacy of the Proposed Model

The performance of the proposed model, i.e., the attentional HLSTM model (M4), is analyzed with the other three models, M1, M2, and M3. M1 uses subband-specific LSTM networks instead of HLSTM networks and without any intra-subband and inter-subband attention modules. M2 uses subband-specific HLSTM networks without any intra-subband and inter-subband attention modules. M3 uses subband-specific HLSTM networks with only intra-subband attention modules. These models (M1, M2, and M3) have the same network architecture as the M4 (Figure 4.5), with a slight variation in generating the final representation vector before feeding to the classification

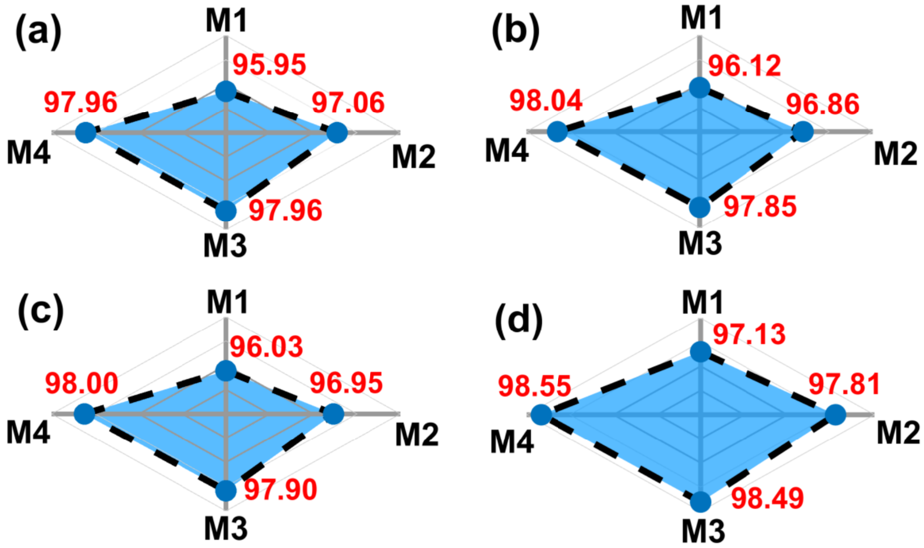

Figure 4.8: Comparison of the performance measures for the proposed model (M4) with three other models (M1, M2, and M3). (a) OPr, (b) OSe, (c) OF1, and (d) OA.

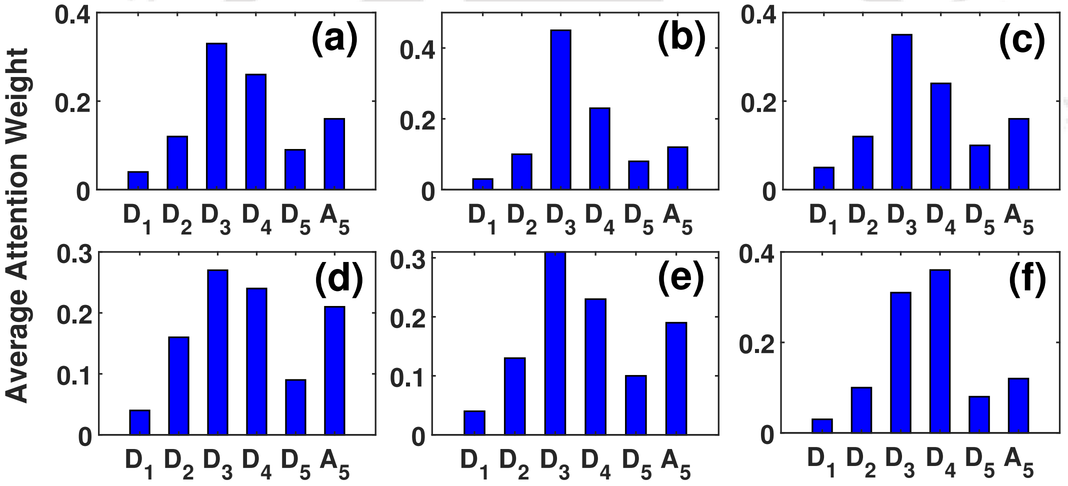

Figure 4.9: Subband-wise average attention weights for the HSM database. (a) For the whole database, (b) AS class, (c) MR class, (d) MS class, (e) MVP class, and (f) Normal class.

block. The comparison of these four models is tested using CICC 2016 database. Figure 4.8 shows the comparison results for all models in overall precision (OPr), OSe, OF1, and OA. It can be seen from Figure 4.8 that M2 shows better performance results than M1. It indicates that exploiting both temporal and relative-temporal information capture more distinctive features related to HMs than exploiting only temporal information. Further, M3 shows better performance than M2. It implies that providing a higher weight to the informative timestamp of the subband improves HVDs detection. Finally, providing intra-subband and inter-subband attention to the HLSTM model (M4), i.e., the proposed model, provides better discrimination capability among classes. It shows the best performance results among all four models.

Figure 4.9 shows the inter-subband average attention weights for the individual and all classes using the HSM database. It can be observed that for all six cases, D_3 and D_4 subbands have learned more weights than the other subbands. It may be because the fundamental HSs and murmur-related temporal information are primarily content in these subbands for detecting HVDs. Among all subbands, D_1 shows less average attention weights for all the classes. It may be because the critical information for diagnosing HMs is primarily present in the frequency range below 500 Hz [13]. Thus, it concludes that the proposed model assigns attention weights to each subband by their clinical utility for identifying HMs.

Further, intra-subband attention weights of D_2, D_3, D_4, and D_5 subbands for a normal and an AS PCG segment is shown in Figure 4.10. It can be seen that the attention mechanism shows more weight to that particular portion of the subbands, which are sensitive to detect normal or HVD pathology. For the normal PCG segment (Figure 4.10(a)), it can be observed that S1 and S2 sound portions show more attention importance in the subbands D_2 to D_5. Similarly, AS PCG segment (Figure 4.10(b)) shows more attention weights in the specific systolic portion where the murmur has occurred and also the fundamental HSs portions in the respective subbands. So, it is evident from inter and intra-subband attention plots that the proposed network provides more attention weight to the portions of the PCG subbands that are clinically relevant to classify HMs.

4.2.6 SWT and DWT Performance Comparison

We have compared the performance of SWT and DWT-based decomposition using the proposed attentional HLSTM network to detect HVDs. The HSM database is used to compare these two decomposition methods. It can be observed from Table 4.1 that our proposed method using SWT shows impressive results in terms of an OPr (%) of 99.43, an ORe (%) of 99.46, an OF1 (%) of 99.44, and an OA (%) of 99.47. All the performance measures show an improvement of almost 13 % using SWT-based decomposition to the conventional DWT-based decomposition. It indicates that SWT captures HM-related information in the coarse scales more efficiently than DWT.

We also quantitatively compare the performance of the proposed method using SWT and DWT decomposition. The feature visualization of these two decompositions using the proposed attention-based HSLTM network is shown in Figure 4.11. For both approaches, first, the learned features for each class are computed before being fed to the dense layer. Further, the dimension

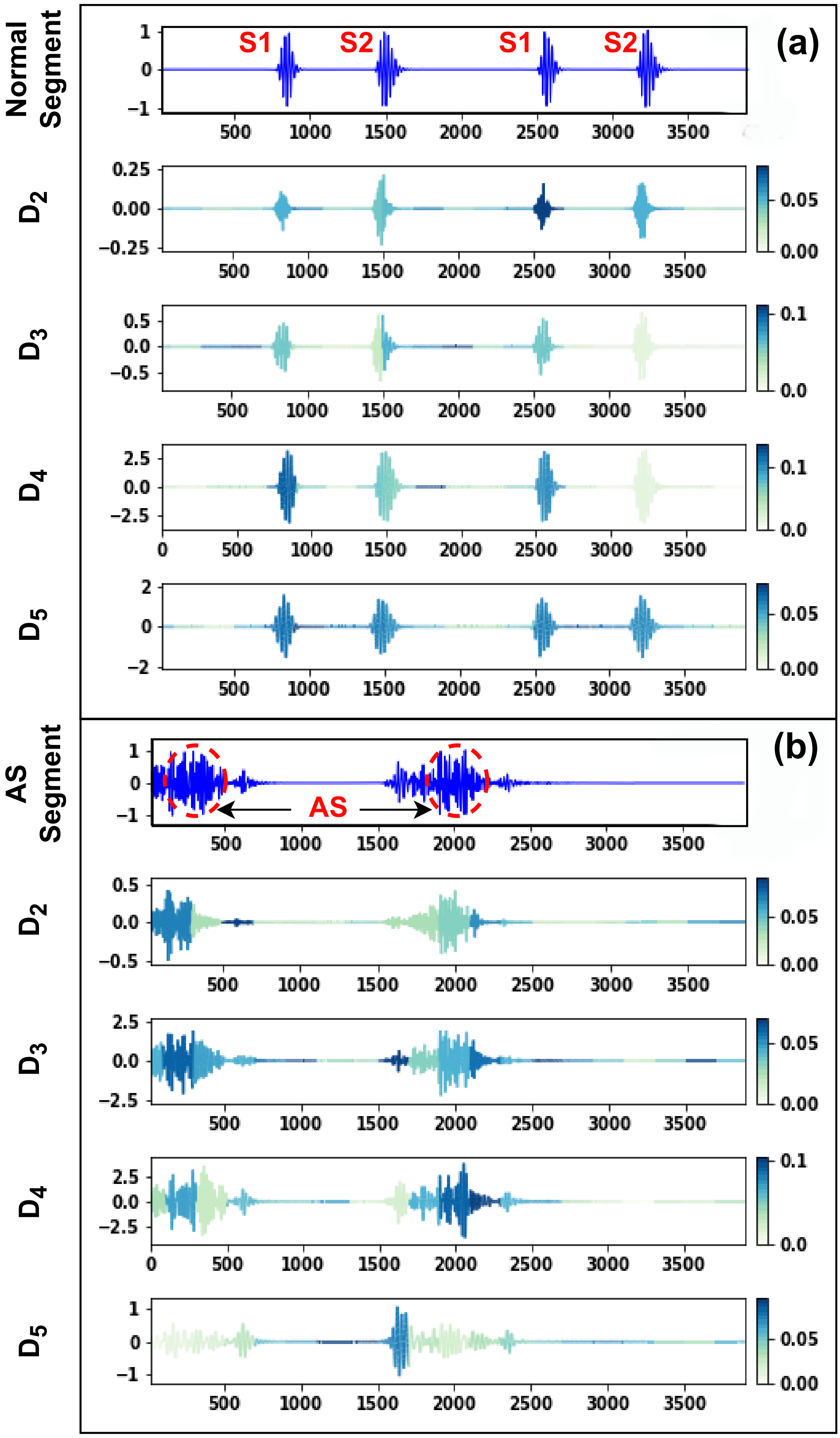

Figure 4.10: Intra-subband attention mapping of PCG segments for D$_2$, D$_3$, D$_4$, and D$_5$ subbands. (a) Normal PCG segment and (b) AS PCG segment

Table 4.1: The DWT and SWT performance comparison using the proposed attentional HLSTM network.

Wavelet Transform	OPr (%)	ORe (%)	OF1 (%)	OA (%)
DWT	86.76 ± 0.93	86.66 ± 0.81	86.54 ± 0.87	86.58 ± 0.89
Proposed SWT	99.43 ± 0.47	99.46 ± 0.42	99.44 ± 0.45	99.47 ± 0.42

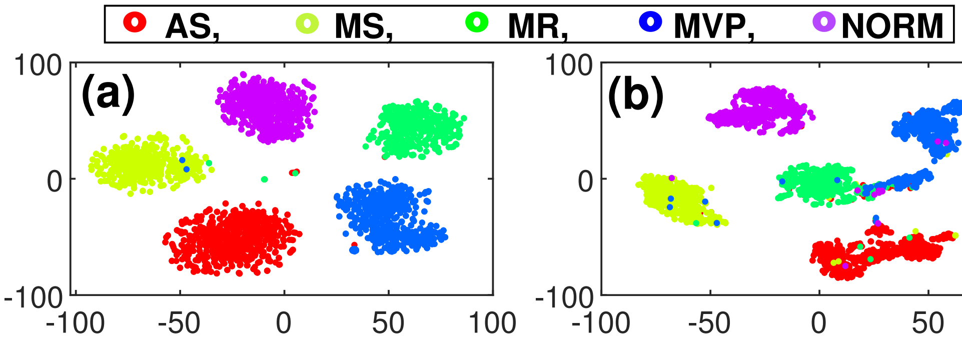

Figure 4.11: Feature visualization of the proposed method. (a) Using SWT decomposition and (b) Using DWT decomposition.

Table 4.2: Confusion matrix of the proposed method using CICC 2016 database.

Actual/ Predicated	Healthy	Pathological
Healthy	101699	1012
Pathological	933	30825

reduction is performed on the learned features using the t-distributed stochastic neighbor embedding (t-SNE) for visualization [161]. It can be observed from Figure 4.11(a) that the proposed method using SWT shows better feature discrimination between the classes. However, the proposed method using DWT decomposition shows the overlap of features between MR and MVP classes (Figure 4.11(b)).

4.2.7 Results using CICC 2016 Database

Table 4.2 shows the confusion matrix of the proposed method using CICC 2016 database across five folds. It shows TP and TN values of 30825 and 101699. It also shows FP and FN values of 1012 and 933. So, high values of TP and TN and low values of FP and FN indicate the reliability of the proposed method to classify healthy and pathological classes. Table 4.3 compares the proposed and state-of-the-art methods in the literature using the CICC 2016 database. The proposed method shows impressive performance results in terms of OSe of 98.04 %, OSp of 99.02 %, and OA of 98.55 %. All the results are five-fold cross-validated. It also shows a minimal variation in results for all the performance metrics across the folds. It can be observed

Table 4.3: The Comparison of Performance Measures for the Proposed and state-of-the-art methods in the literature using the CICC 2016 Database.

Related Work	Cyclewise Segment	Used Features	Model	OSe (%)	OSp (%)	OA (%)
Potes *et al.*, 2016 [100]	Yes	Time and Frequency	AdaBoost, CNN	94.24	77.81	86.02
Zhang *et al.*, 2019 [96]	No	Temporal, Quasi-periodic	LSTM	94.22 ± 0.02	90.48 ± 0.02	92.35 ± 0.01
Chen *et al.*, 2020 [113]	Yes	Modified Frequency Slice	CNN	95 ± 2.11	93 ± 1.86	94 ± 1.34
Dhar *et al.*, 2021 [98]	No	Cross-wavelet Transform	AlexNet	98	98	98
Li *et al.*, 2021 [14]	No	Raw PCG Segment	CNN-GRU	94.41	96.98	95.50
Proposed Work	No	SWT	Attentional HLSTM	**98.04 ± 0.10**	**99.02 ± 0.10**	**98.55 ± 0.05**

Table 4.4: The Comparison of Performance measures for the proposed and state-of-the-art methods using the HSM database.

Related Work	Cycewise Segment	Used Features	Model	Class	Se (%)	Sp (%)	F1 (%)	OA (%)
Son *et al.*, 2018 [59]	No	DWT	SVM	All	92.3	98.4	99.1	92.3
Ghosh *et al.*, 2019 [63]	Yes	WSST	Random Forest	Except MVP	—	—	—	95.13
Son *et al.*, 2018 [59]	No	DWT and MFCC	SVM	All	98.2	99.4	99.7	97.9
Ghosh *et al.*, 2020 [64]	Yes	Chirplet Transform	Composite Classifier	N MR AS MS	98.49 ± 1.45 96.33 ± 3.58 99.66 ± 0.70 98.83 ± 1.12	99.94 ± 0.17 99.94 ± 0.71 99.04 ± 0.75 99.26 ± 0.37	99.32 ± 1.04 97.99 ± 1.60 98.36 ± 1.32 98.60 ± 1.62	98.54 ± 1.11
Proposed Work	No	SWT	Attentional HLSTM	N MVP MR AS MS	100 98.26 ± 1.73 99.27 ± 0.67 100 99.78 ± 0.50	100 100 99.81 ± 0.27 99.89 ± 0.15 99.75 ± 0.18	100 99.12 ± 0.88 99.15 ± 0.70 99.84 ± 0.22 99.11 ± 0.84	**99.47 ± 0.42**

from Table 4.3 that the proposed method performs better than the state-of-the-art methods across all the performance results. Specifically, compared to the time and frequency domain approaches [96, 100], the proposed method shows an improvement of almost 4% in OSe, 9% in OSp, and 7% in OA. The proposed method also shows an improvement of almost 6% in OA than the LSTM-based approach [96].

4.2.8 Results using HSM Database

Table 4.4 shows the multi-class classification comparison of the proposed method and the state-of-the-art methods using the HSM database. It can be observed that the proposed method exhibits outstanding results for both class-wise Se, Sp, and F1 measures as well as the OA

Table 4.5: Confusion matrix of the proposed method using HSM database.

Actual/ Predicated	N	MVP	MR	AS	MS
N	488	0	0	0	0
MVP	0	507	3	1	5
MR	0	0	406	1	2
AS	0	0	0	609	0
MS	0	0	1	0	444

measure. The proposed method achieves an OA of 99.47 % with a small standard deviation of 0.42. The confusion matrix verifies the results as shown in Table 4.5. It shows promising results for all five classes across five folds. The efficacy of the proposed method is verified by comparing it with the currently existing methods. Table 4.4 shows that the proposed method performs better than the existing methods in terms of class-wise Se, Sp, F1, and an excellent OA. Specifically, compared to the conventional DWT-based approach [59], the proposed method shows an improvement of almost 7% in both Se and OA measures. The improved results show that the proposed attention-based HLSTM network effectively captures the temporal information at different abstractions to improve HMs classification.

4.2.9 Efficacy of the proposed method under noisy conditions

The robustness of the proposed method is evaluated by adding different types of body and environmental noises. In this work, we have used eight noises: a sick man sneezing, a human sniffling, a sick man coughing, a person's nose blowing, urban square noise, kids playing, a baby crying, and door knocking [1]. A detailed discussion about the noises is given in 3.2.7. A particular noise (among these eight noises) was selected randomly and added to each PCG recording. We have experimented with the performance of the proposed method for different signal-to-noise ratio (SNR) levels using the HSM database. Table 4.6 shows the OA results for different SNR levels. We have obtained an OA of 98.27%, 98.96%, 99.26%, and 99.47% for an SNR of 0 dB, 10 dB, 20 dB, and without noise, respectively. The results indicate that the proposed method is robust enough to handle environmental and body-generated noises.

---

[1] https://mixkit.co/free-sound-effects/static/

Table 4.6: The Performance of the proposed models under different SNR levels of environmental and body noises.

SNR	OA (%)
0 dB	98.27
10 dB	98.96
20 dB	99.26
Without Noise	99.47

Table 4.7: Comparison of the NMP, ART, and OA for the CICC 2016 and HSM databases.

Database	NMP	ART (ms)	OA (%)
CICC 2016	909559	9.96	98.55
HSM	909712	10.14	99.47

4.2.10 Complexity of the Proposed Model

Table 4.7 shows the complexity of the proposed attention-based HLSTM model for both the CICC 2016 and HSM databases. The proposed model shows 909559 model parameters (NMP), an average run time (ART) of 9.96 ms, and an OA of 98.55% using the CICC 2016 database. Similarly, using the HSM database, the proposed model shows 909712 NMP, an ART of 10.14 ms, and an OA of 99.47%. So, the reasonable ART (ms) and the excellent OA (%) of the proposed model for both databases make it suitable for faster and better diagnosis of HMs in primary healthcare units.

4.3 Summary

In this chapter, we have proposed a novel method for the detection (normal or abnormal) and classification (normal, AS, MVP, MR, and MS) of HMs using two open-access databases. The proposed method is an SWT-based HLSTM network with intra-subband and inter-subband attention modules. SWT captures the fundamental HSs and HVDs-related information adequately in its subbands. However, the subband-specific HLSTM networks exploit each subband's temporal and relative-temporal variation to detect HVDs. Finally, the attentional modules represent the encoded vectors of HLSTM networks to a discriminative representation by emphasizing clinically relevant information. The proposed method yields competitive results for detecting and classifying HMs. It shows an OA of 98.55% and 99.47% using the CICC 2016 and HSM databases, respectively. The robustness of the proposed method is evaluated by adding different body and environmental noises to the PCG signal. The proposed network architecture has an NMP of 909559 and an ART of 9.96 ms using the CICC 2016 database for binary classification. Similarly, the proposed network architecture has an NMP of 909712 and an ART of 10.14 ms using the HSM database for multi-class classification of HMs. The proposed method's excellent OA and speedy response help clinicians during preliminary check-ups to detect HMs.

5

HM Severity Stages Classification Using Multi-kernel Residual CNN

Contents

In literature, most existing automated methods focus only on classifying PCG signals as healthy or pathological. In chapter 1, we proposed an oscillatory modeling method to classify pathological HSs. A few works are there in the literature where they have performed a multi-class analysis of HVDs. We also proposed two DL-based transfer-domain approaches in chapter 3 and chapter 4 for multi-class classification of HMs. However, they have limited to fewer data for a specific HVD analysis [59]. To fill this research gap, we have proposed an automated method to classify three HMs severity stages, i.e., absent HMs (AHMs) or healthy, soft heart murmurs (SHMs), and loud heart murmurs (LHMs), along with the low-quality PCGs (LQP). To our knowledge, we have used the CirCor Digiscope Phonocardiogram (CCDP) database [144, 162] for the first time in the literature to classify the HM severity stages. The CCDP database was used for the George B. Moody PhysioNet challenge 2022 to identify the presence, absence, or unclear cases of HMs. In the case of HMs, the loudness and quality of the sound can indicate the severity of the heart condition. SHM may indicate a less severe heart condition, while LHM may indicate a more severe one. Additionally, an LQP may be less reliable for diagnosing the disease. In cardiac physiology, experts have used the Levine scale to evaluate the severity of an HM based on its intensity or loudness. Levine scale is a numeric grading system ranging from Grade I to Grade VI [22]. The details about the HMs grading system are provided in chapter 1.2.1. If the HM grading is equal to or less than Grade II, it is defined as an SHM. Similarly, if the HM grading is greater than or equal to Grade III, it is defined as an LHM. Therefore, designing a robust automated HM grading system using PCG can provide a reliable solution for primary healthcare units to detect HVDs.

Most existing deep CNN-based models have used transfer-domain approaches to represent PCG segments as 2D images [71]. These images are then fed to a 2D- CNN model to classify HMs. However, representing the time-series PCG signal in an image form may not capture the relevant HM information. A few methods have used single-kernel 1D- CNN models [130], where local spatial features are extracted using the CNN model, followed by classification using dense and softmax layers. However, considering the morphological variability across the HMs and sometimes the weak HMs are often masked by other unwanted sounds in the PCG signal. These approaches may not be adequate for effectively capturing the relevant diagnostic features in such cases. Also, deep neural networks have the problem of vanishing gradients. Therefore, the multi-kernel deep CNN with residual learning can extract deep scale-specific features. A fusion of these features can provide better performance results.

This chapter proposes a multi-kernel residual CNN (MK-RCNN) model for more robust feature extraction and efficient classification of HM severity stages. The multi-kernel CNNs with different filter sizes help capture a broad range of features from the PCG signal. Further, employing residual learning can help to extract deep features from CNN layers without degrading the performance accuracy.

The rest of the chapter is organized as follows. Section 5.1 and 5.2 present the proposed framework and the experimental results of the MK-RCNN model, respectively. The chapter summary is presented in section 5.3.

5.1 Proposed MK-RCNN Model for HM Severity Stage Classification

In the proposed MK-RCNN model, first, the raw PCG signal $\mathbf{S_o}[r]; (r = 1, 2, ..., R)$ is preprocessed and segmented as discussed in the chapter 3.1.1. Then, each 2s PCG segment $\mathbf{C}[r]; (r = 1, 2, 3, ..., Q)$ is fed to the MK-RCNN model to classify the HM severity stages.

5.1.1 MK-RCNN Model Architecture

The MK-RCNN model extracts better diagnostic information from the PCG segment $\mathbf{C}[r]$ to detect HMs by leveraging multi-kernel CNNs. The multiple kernels can capture different aspects of the PCG segment, i.e., the low and high-frequency content of the PCG segment, to improve the detection of HM severity stages. Figure 5.1 shows the network architecture of the proposed MK-RCNN model. It can be seen from Figure 5.1 that the PCG segment is fed to the corresponding residual CNN (RCNN) block to extract multi-kernel disease features. The number of RCNN blocks for feature extraction is set through experimentation. In this work, four RCNNs blocks, i.e., $\mathbf{RCNN}_0$, $\mathbf{RCNN}_1$, $\mathbf{RCNN}_2$, and $\mathbf{RCNN}_3$ having kernel sizes $k = 3, 5, 7, 9;$ are used. The residual learning in each RCNN block helps to extend the CNN network to build a deeper structure with no degradation in detection performance [163, 164]. It is performed by using identity mapping.

Figure 5.2 shows a residual network designed by stacking two basic CNNs, i.e., CNN- 1 and CNN- 2. A basic CNN is a particular type of neural network (NN) [165]. It comprises convolutional layers followed by activation functions and the pooling layer. The CNN extracts hierarchical features from the input PCG segment. The first CNN layer extracts the low-level representation of the input PCG segment. Increasing the CNN layers extracts the more complex deep features from the PCG segment. The kernel size k, the number of filters n, and the stride s are used to do

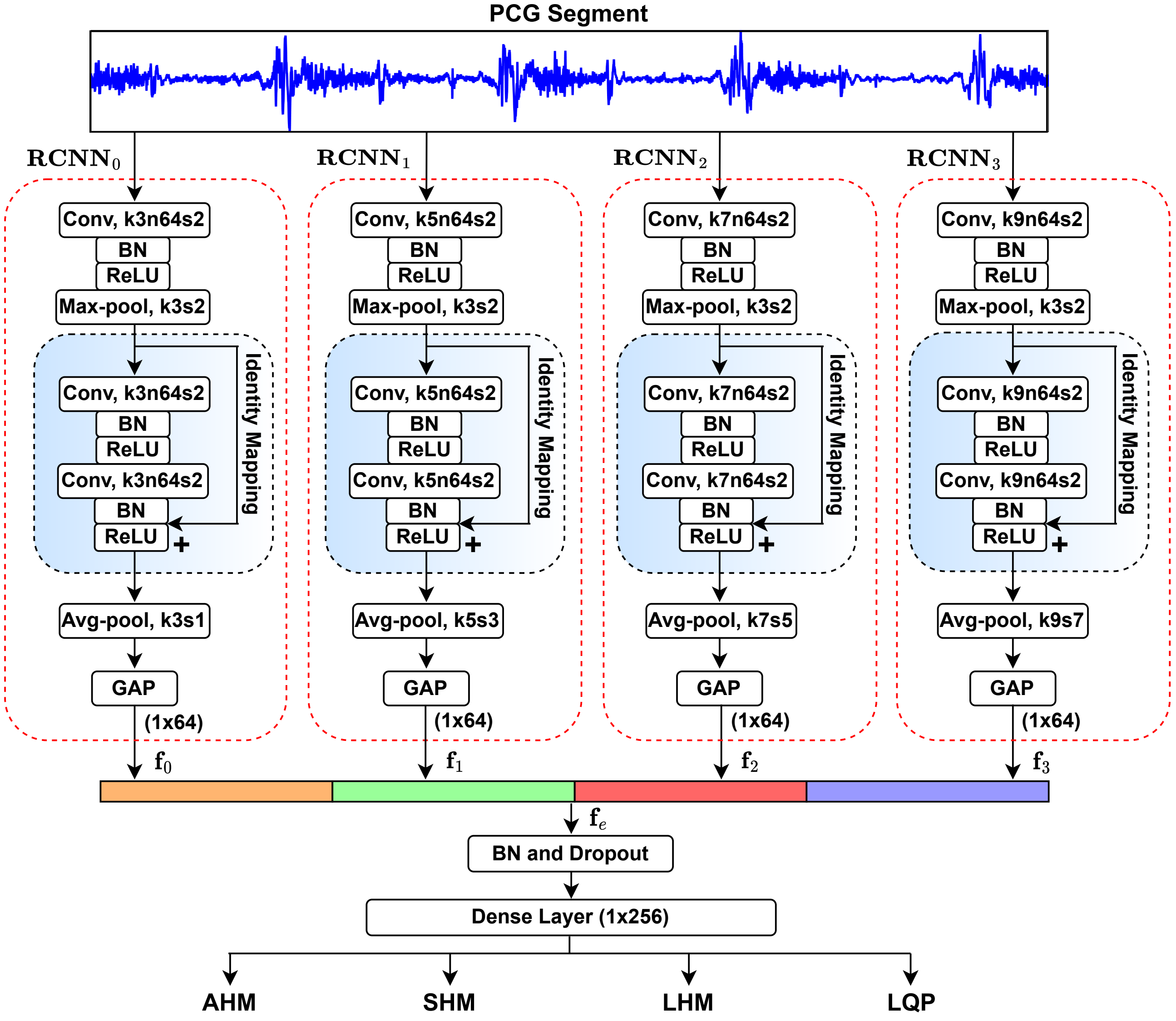

Figure 5.1: Illustration of the proposed multi-kernel residual CNN (MK-RCNN) model to detect heart murmur (HM) severity stages. Where k, n, and s represent kernel size, number of filters, and stride for the convolution operations.

the linear transformation of the local regions. It extracts complex features from the non-stationary PCG segment to enhance the detection of HM severity stages. Let the input signal to the residual learning network is $\mathbf{x}_i$. Initially, it is passed through CNN-1, which performs three sequential operations. First, $\mathbf{x}_i$ is passed through a convolutional layer, then through a batch normalization (BN) layer, and followed by a rectified linear unit (ReLU) activation layer. BN is used to improve generalization and provide higher learning rates. The ReLU activation function improves the training speed of the proposed network. The output of CNN-1 is h_{1i} and fed as input to CNN- 2. In CNN-2, it is passed through a convolution operation followed by the BN layer. The output of the

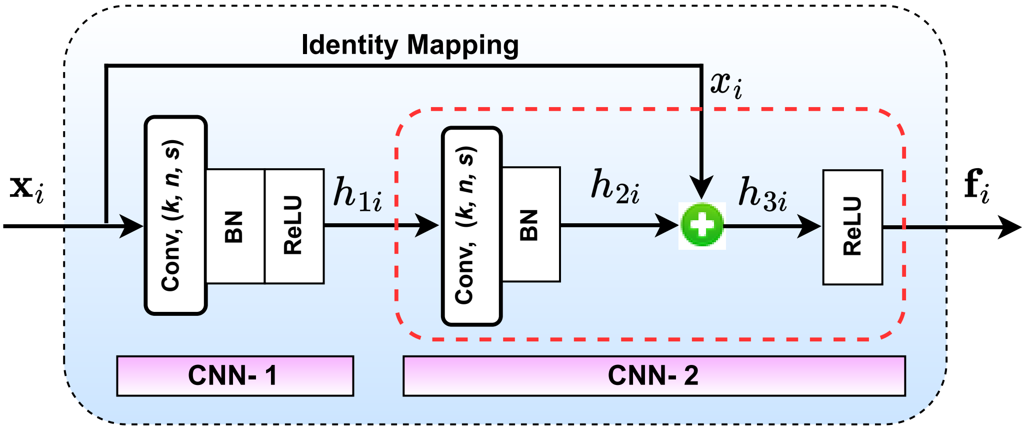

Figure 5.2: Illustration of a residual network having two basic CNN networks (CNN-1 and CNN-2). An identity mapping is used after the batch normalization (BN) layer of CNN- 2.

BN layer is h_{2i}. To deal with the degradation problem, h_{2i} is added with the input $\mathbf{x}_i$, known as identity mapping. So, the output h_{3i} is defined as follows.

$$h_{3i} = h_{2i} + \mathbf{x}_i$$

Further, the output h_{3i} is passed through a ReLU activation function to give the final output $\mathbf{f}_i$ of the residual network and is defined as follows.

$$\mathbf{f}_i = \text{ReLU}(h_{3i})$$

Then, $\mathbf{f}_i$ is fed as input to the average pooling layer followed by global average pooling (GAP). GAP reduces the number of weights and helps the network to avoid over-fitting.

Finally, the outputs of the RCNNs ($\mathbf{RCNN}_i$, $i \in \{0, 1, 2, 3\}$) are fused through concatenation to have the encoded multi-kernel feature vector and can be expressed as $\mathbf{f}_e = [\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3]$. The encoded vector captures the variations of the PCG segment. It transfers the feature representation to higher dimensions by increasing the non-linearity, which assures better discrimination capability of the classifier. In the end, $\mathbf{f}_e$ is given as input to a dense layer with a softmax activation function to obtain the class probabilities for the input PCG segment. It is formulated as follows;

$$\hat{\mathbf{y}} = \text{softmax}\left(\mathbf{W}_o \cdot \mathbf{f}_e + \mathbf{b}_o\right) \tag{5.1}$$

where [$\mathbf{W}_o \in \mathbb{R}^{M \times d}$, $\mathbf{b}_o \in \mathbb{R}^{M}$] are the learnable weight matrix and bias vector of the dense layer. M is the number of classes. The predicted class label is $\hat{\mathbf{y}} \in \mathbb{R}^{M}$.

Training loss optimization: Conventionally, the model parameters are trained by using a cross-entropy loss function ($\mathcal{L}$) which is formulated as follows [160]:

$$\mathcal{L} = -\sum_{n=1}^{N}\sum_{m=1}^{M}(\mathbf{y}_{n,m}) \cdot \log\left[(\hat{\mathbf{y}}_{n,m})\right] \tag{5.2}$$

where $\mathbf{y}_{n,m}$, $\hat{\mathbf{y}}_{n,m}$ are the true and predicated labels for the n^{th} sample. N is the total number of training samples. M is the number of classes. However, the existing PCG databases often have imbalanced classes. As a result, the classification results can be more biased towards the majority classes and provide poor results to the minority classes. To overcome this issue, we have proposed a cost-sensitive loss function to train the proposed MK-RCNN model [166]. A class-weighted categorical cross-entropy loss is deployed that penalizes the class errors differently. The minority classes are assigned higher weights than the majority classes. The class weighted cross-entropy loss function $\mathcal{L}_w$ is formulated as follows:

$$\mathcal{L}_w = -\sum_{n=1}^{N}\sum_{m=1}^{M} w_m \cdot (\mathbf{y}_{n,m}) \cdot \log\left[(\hat{\mathbf{y}}_{n,m})\right] \tag{5.3}$$

Where w_m is the weight assigned to the class m. The w_m is computed as follows:

$$w_m = \frac{\text{Number of training samples in majority class}}{\text{Number of training samples in class } m} \tag{5.4}$$

We can conclude from Eq. 5.4 that the weight assigned to a particular class is inversely proportional to the class distribution. $\mathcal{L}_w$ assigns higher weights to the low sample size training samples. Therefore, the obtained weights equalize the degree of impact for individual classes toward the total loss.

5.2 Experimental Results and Discussion

This section discusses the databases used to evaluate the proposed method, performance measures, network architecture, and ablation study. Further, we have discussed the performance results using different databases and the model complexity of the proposed MK-RCNN model.

5.2.1 Databases

In this work, we have used CirCor Digiscope Phonocardiogram (CCDP) database [144,162] for the HM severity stages classification. We have also evaluated the efficacy of the proposed MK-RCNN

Table 5.1: The gender and age distribution of the subjects among AHM, SHM, LHM, and LQP classes for the CCDP database. Acronym: male (M) and female (F).

Class	**AHM** (695)		**SHM** (132)		**LHM** (46)		**LQP** (68)	
Gender	M (340)	F (355)	M (61)	F (71)	M (25)	F (21)	M (29)	F (39)
Neonates	2	2	0	1	0	0	0	0
Infants	70	51	14	20	0	3	15	18
Child	231	219	39	42	20	17	12	17
Adolescents	35	18	7	5	4	1	1	2
Unknown	2	65	1	3	1	0	1	2

model by using PhysioNet/ Computing in Cardiology Challenge (CICC) 2016 database for binary classification of PCG segments [13, 144] and the HSM database for multi-class classification of PCG segments [59]. The details of the CICC 2016 and HSM databases are given in chapter 2.2.1 and 3.2.1.1, receptively. A brief detail of the CCDP database is as follows.

5.2.2 CCDP Database

The database was used for the George B. Moody PhysioNet challenge 2022. It is the largest publicly available pediatric murmur database. It was collected in Paraiba, Brazil, as two mass screening campaigns from July to August 2014 and June to July 2015. It contains a total of 3159 PCG recordings from 941 subjects. There are 2391- PCG records collected from 695 subjects in the absent heart murmur (AHM) or healthy (H) class, 427- PCG records collected from 132 subjects in the soft heart murmur (SHM) class, and 185- PCG records collected from 46 subjects in the loud heart murmur (LHM) class. Additionally, 156- PCG records were collected from 68 subjects in the low-quality PCG (LQP) class. Table 5.1 shows the gender and age distribution of the subjects for these four classes. The mean age of the subjects is 73.4 ± 0.1 months. Two independent expert cardiac physiologists verify the ground truth of the PCG recordings. Murmur gradings are based on the Levine scale [167]. Murmur grades I and II on the Levine scale correspond to SHM class. In comparison, murmur grades III or above in Levine class corresponds to LHM class. For most subjects, there are four PCG recordings from the four heart valve locations, i.e., aortic, pulmonary, tricuspid, and mitral valves. These recordings are recorded from the valve locations sequentially (not simultaneously). The PCG signals are acquired using a Littmann 3200 stethoscope with a sampling rate of 4 kHz. Further, they have normalized within the $[-1, 1]$ range. The recordings' duration varies between 4.8 s to 80.4 s ($\mu \pm \sigma$ = 22.9 ± 7.4 s). The data are

Table 5.2: The effect of the number of RCNNs for the classification performance of the proposed MK-RCNN model using the CCDP database.

Configuration	OPr (%)	ORe (%)	OF1 (%)	OA (%)
$RCNN_0$	83.54	59.50	65.66	85.44
$RCNN_0 - RCNN_1$	87.59	73.37	78.90	89.96
$RCNN_0 - RCNN_1 - RCNN_2$	96.11	92.31	94.07	97.07
$\mathbf{RCNN}_0 - \mathbf{RCNN}_1 - \mathbf{RCNN}_2 - \mathbf{RCNN}_3$	**96.98**	**97.36**	**97.05**	**98.33**
$RCNN_0 - RCNN_1 - RCNN_2 - RCNN_3 - RCNN_4$	95.55	91.47	93.38	96.61

collected in a real-world scenario and have stethoscope rubbing noise, speaking, crying, and laughing sounds in the background.

5.2.3 Evaluation Scheme and Performance Measures

The performance of the proposed method for the CCDP database is evaluated by using measures, i.e., classwise precision (Pr), classwise recall (Re), classwise F1- score (F1), overall precision (OPr), overall recall (ORe), overall F1- score (OF1), and overall accuracy (OA) [143]. The performance of the proposed method for the CICC 2016 and HSM database is evaluated by using three measures, i.e., overall sensitivity (OSe), overall specificity (OSp), and overall accuracy (OA). All the results are stratified five-fold cross-validated. The cross-validation provides generalizable results and prevents over-fitting to a specific subset of the database.

5.2.4 Network Architecture and Ablation Study

The preprocessed PCG segments are provided to the MK-RCNN model for extracting multi-kernel feature representations. Figure 5.1 shows the network architecture of the MK-RCNN model. The number of RCNN modules ($\mathbf{RCNN}_i, i \in \{0, 1, ..., Q\}$) used for the proposed method are set through experimentation. Table 5.2 shows the performance results by increasing the number of RCNNs using the CCDP database. It can be observed from Table 5.2 that the classification performance is improved by fusing RCNN modules up to four; beyond that, it is reduced. So, we consider the fusion of four kernels, i.e., $k = \{3, 5, 7, 9\}$. A dropout layer with a dropout rate of 0.2 is deployed before the dense layer to prevent the proposed model from over-fitting issues. The optimum values for the hyper-parameters of the proposed model are tuned through experiments.

In this study, we have deployed an Adam optimizer with a learning rate of 0.005. The model's parameters are trained using back-propagation through time (BPTT) [160]. The batch size is set

Table 5.3: The effect of residual learning to classify HM severity stages using the CCDP database.

Configuration	OPr (%)	ORe (%)	OF1 (%)	OA (%)
CNN_0	80.47	56.29	61.03	84.08
RCNN$_0$	**83.54**	**59.50**	**65.66**	**85.44**

Table 5.4: The performance measures for the proposed MK-RCNN model using CCDP database across five folds.

Class	Pr (%)	Re (%)	F1 (%)	OA (%)
AHM	99.11	98.85	98.98	**98.33** ± **0.12**
SHM	95.31	96.15	95.72	
LHM	98.77	98.71	98.74	
LQP	94.71	95.73	94.74	
Overall	96.98 ± 2.29	97.36 ± 1.65	97.05 ± 2.13	

to 128, and the model is trained for 350 epochs. The deep learning models are implemented using Python programming language with the PyTorch library. The experiments are performed by using a Titan P100 GPU server.

5.2.4.1 Significance of multi-kernel feature fusion:

It can be observed from Table 5.2 that a single kernel RCNN network, i.e., **RCNN**$_0$ having kernel size 3 shows an OA of 85.44%. However, adding multi-kernel RCNN networks, i.e., **RCNN**$_0$ $-$ **RCNN**$_1$ $-$ **RCNN**$_2$ $-$ **RCNN**$_3$ having kernel sizes 3, 5, 7, and 9 show an OA of 98.33%. There is an improvement of 13% in OA by deploying a multi-kernel RCNN network than single-kernel RCNN. The proposed MK-RCNN model has also shown a better OPr of 96.98%, ORe of 97.36%, and OF1 of 97.05%. This verifies that the multi-kernel RCNNs improve the discrimination capability of the classifier, thereby providing better performance results.

5.2.4.2 Significance of residual learning:

We have verified the effectiveness of the residual learning for the proposed method to classify the HM severity stages. Table 5.3 shows the performance results for **RCNN**$_0$ having residual learning and CNN_0 have the same network architecture as that of **RCNN**$_0$ without having residual learning. The **RCNN**$_0$ shows a better OPr of 83.54%, ORe of 59.50%, OF1 of 65.66%, and OA of 85.44%. Thus, it indicates that residual learning improves the performance results by leveraging identity mapping. It helps to capture the deep features from deeper layers and reduces the risk of information loss.

Table 5.5: The comparison of performance measures for the proposed and the state-of-the-art methods using the HSM database.

Related Work	Cycewise Segment	Used Features	Model	OSe (%)	OSp (%)	OF1 (%)	OA (%)
Son *et al.*, 2018 [59]	No	DWT	SVM	92.3	98.4	99.1	92.3
Ghosh *et al.*, 2019 [63]	Yes	WSST	Random Forest	—	—	—	95.13
Son *et al.*, 2018 [59]	No	DWT and MFCC	SVM	98.2	99.4	99.7	97.9
Ghosh *et al.*, 2020 [64]	Yes	Chirplet Transform	Composite Classifier	98.32	99.54	98.57	98.54
Oh *et al.*, 2020 [129]	No	Raw PCG Segments	CNN with Residual Block	92.5	98	—	97
Alkhodari *et al.*, 2021 [127]	No	Raw PCG Segments	CNN with BLSTM	98.30	99.58	98.30	99.32
Proposed Work	No	Raw PCG Segments	**MK-RCNN**	100	100	100	100

5.2.5 Results on the CCDP Database

Table 5.4 shows the classwise and overall performance measures to classify the HM severity stages using the CCDP database. All the results are five-fold cross-validated. The proposed MK-RCNN model shows an OA (%) of 98.33, OPr (%) of 96.98, ORe (%) of 97.36, and OF1 (%) of 97.05. These results indicate that the proposed MK-RCNN model accurately classifies the four HM severity stages. The high Pr, Re, and F1 for all classes and the OA of 98.33% demonstrate the potential usefulness of the model for the clinical diagnosis of HM.

5.2.6 Results using the HSM Database

We have deployed the HSM database for multi-class classification of HMs using the proposed MK-RCNN model. Table 5.5 shows the comparison results of the proposed method and the state-of-the-art methods using the HSM database. It can be observed that the proposed method exhibits outstanding 100% results for Se, Sp, F1, and OA measures. The proposed method shows an improvement of almost 8% in both Se and OA measures concerning time-frequency-based approaches [59, 63]. The improved results show that the proposed MK-RCNN model effectively captures the spatial information from raw PCG segments to improve HMs classification. The proposed method also shows competitive results with CNN-based approaches [127, 129].

Table 5.6: Confusion matrix of the proposed method using CICC 2016 database.

Actual/ Predicated	**Healthy**	**Pathological**
Healthy	34541	689
Pathological	290	31040

5.2.7 Results using the CICC 2016 Database

We have deployed the proposed MK-RCNN model for classifying PCG segments as healthy or pathological using the CICC 2016 database. The database contains 2575 healthy and 665 pathological PCG signals. After preprocessing, we segmented each PCG signal using a 2s rectangular window with 25% overlap for healthy and 75% overlap for pathological PCGs. The difference in overlap % for both classes is to have a balanced dataset. So, we have 35230 healthy and 31757 pathological PCG segments. Table 5.6 shows the confusion matrix of the proposed method using CICC 2016 database across five folds. The proposed method achieves TP and TN values of 31040 and 34541, respectively, correctly identifying true positives and negatives. The method also reports low FP and FN values of 689 and 290, respectively, demonstrating its ability to minimize false positives and false negatives. Table 5.7 compares the proposed method with state-of-the-art methods using the CICC 2016 database to classify PCG segments as healthy or pathological. The proposed method demonstrates impressive performance with OSe of 99.08%, OSp of 98.04%, and OA of 98.53%, while exhibiting minimal variation in performance metrics across the folds. Our proposed method outperforms all baseline methods across all performance results, with a noticeable improvement over time, frequency, and time-frequency domain approaches [96, 100] by almost 5% in OSe, 9% in OSp, and 7% in OA. Additionally, compared to the CNN-GRU-based approach [14], our proposed method demonstrates an improvement of almost 3% in OA. These results suggest the efficacy of the proposed method in classifying the pathological PCG segments.

5.2.8 Model Complexity of the Proposed MK-RCNN Network

Table 5.8 shows the complexity of the proposed MK-RCNN model for the CCDP, CICC 2016, and HSM databases. The proposed model shows 217604 model parameters (NMP), an average run time (ART) of 2.36 ms, and an OA of 98.33% using the CCDP database. The proposed model shows 217090 NMP, 2.82 ms ART, and an OA of 98.53% using the CICC 2016 database. Similarly,

Table 5.7: The comparison of performance measures for the proposed method and the state-of-the-art methods using the CICC 2016 Database.

Related Work	Cyclewise Segment	Used Features	Model	OSe (%)	OSp (%)	OA (%)
Potes *et al.*, 2016 [100]	Yes	Time and Frequency	AdaBoost, CNN	94.24	77.81	86.02
Zhang *et al.*, 2019 [96]	No	Temporal, Quasi-periodic	LSTM	94.22 ± 0.02	90.48 ± 0.02	92.35 ± 0.01
Chen *et al.*, 2020 [113]	Yes	Modified Frequency Slice	CNN	95 ± 2.11	93 ± 1.86	94 ± 1.34
Dhar *et al.*, 2021 [98]	No	Cross-wavelet Transform	AlexNet	98	98	98
Li *et al.*, 2021 [14]	No	Raw PCG Segment	CNN-GRU	94.41	96.98	95.50
Proposed Work	No	Raw PCG Segment	MK-RCNN	**99.08 ± 0.30**	**98.04 ± 0.15**	**98.53 ± 0.16**

Table 5.8: Comparison of the NMP, ART, and OA for the CCDP, CICC 2016, and HSM database.

Database	NMP	ART (ms)	OA (%)
CCDP	217604	2.36	98.33
CICC 2016	217090	2.82	98.53
HSM	217861	2.47	100

using the HSM database, the proposed model shows 217861 NMP, an ART of 2.47 ms, and an OA of 100%. So, the reasonable ART (ms) and the excellent OA (%) of the proposed model for all the databases make it suitable for faster and better diagnosis of HMs in primary healthcare units.

5.3 Summary

In this chapter, we have proposed an MK-RCNN model for the automated classification of HM severity stages. The proposed CNN architecture's residual learning helps avoid performance degradation and extract informative features from deep CNN layers. Further, the fusion of multi-kernel RCNN networks improves the classification performance of HMs by exploiting the pathological variations of the PCG segments in different scales. The performance results indicate that the proposed model is competitive enough with the baseline methods in the literature. The proposed model shows an end-to-end framework that extracts features from the raw PCG signals and is robust enough to classify the HMs. The accurate prediction performance and low ART (ms) for the proposed MK-RCNN model are well-suited for the reliable and fast diagnosis of HMs using PCG signals.

6

Conclusions

Contents

Cardiovascular diseases (CVDs) are a major cause of death worldwide, accounting for approximately 32% of all deaths [168]. Among different types of CVDs, heart valve diseases (HVDs) are highly prevalent in developing and underdeveloped countries. HVDs result in about 1.4 million deaths annually [5]. These countries' lack of healthcare infrastructure and skilled cardiologists exacerbates the problem. Although HVDs are not always lethal initially, failure to detect them early can lead to cardiac arrest or heart failure. Therefore, early detection of HVDs is crucial to reduce the mortality rate and associated healthcare costs. This thesis proposes four automated PCG-based diagnosis methods to detect HVDs. The thesis exploits the pathological variations in the PCG signal and develops reliable and robust DL-based approaches. The following paragraphs provide a summary of each chapter of the thesis.

Chapter 1 of this dissertation provides an overview of the role of the heart and heart valves, as well as the clinical importance of the PCG signal. Additionally, it highlights the prevalence of HVDs and the challenges associated with PCG-based diagnosis. A comprehensive literature review of the automated diagnosis of HVDs using PCG is also presented, where it is identified that some research gaps still need to be addressed in previous studies. The chapter concludes by discussing the motivation for the current research, highlighting the significant findings related to each motivation. The chapter outlines the importance of the automated diagnosis of HVDs and the need for accurate and reliable diagnostic methods. It also addresses the challenges associated with detecting HMs in PCG signals and highlights the gaps in the current literature. These gaps include limited data availability, lack of focus on the classification of multiple HVD severity stages, and the limited exploration of deep learning approaches for the automated diagnosis of HVDs. The chapter concludes by discussing the significance of addressing these research gaps. The chapter also highlights the potential impact of the proposed research on developing reliable and effective diagnostic methods for HVDs.

Most of the existing PCG-based modeling methods [50,56,79,136] have difficulties in capturing the morphological variations of the PCG signal due to their non-stationary nature. To address this issue, existing models have focused on modeling short segments of the PCG cycle, which may not provide accurate diagnoses of HVDs. In Chapter 2, we proposed a novel multi-component oscillatory model to classify PCG cycles as healthy or pathological. We analyzed the proposed model in detail for the representation and recognition of PCG cycles and their respective HS states. The representation results show that the PCC value reaches almost 99% at model order two for healthy and pathological PCG cycles. For the classification of PCG cycles, the proposed method has shown an OA of 90.91% using the proposed model-based features with a DNN classifier. The results demonstrate that the proposed method is competitive with existing state-of-the-art methods. The number of parameter-sets for a model order depends on the number of zero-crossing points of the given input signal. Further, the intensity of heart sounds can vary from person to person due to a combination of factors, both physiological (age, cardiac conditions, heart rate, and pregnancy) and non-physiological (body size and composition, environmental noises, stethoscope quality, and patient movement). These variations are essential when interpreting heart sounds and diagnosing cardiac conditions. Moreover, the proposed method is easy to compute and implement. It requires only the calculation of zero-crossing points and fitting a half-period sine wave function between two consecutive zero-crossing points. However, designing hand-crafted features requires domain expertise and can be subjective. Additionally, finding a fixed set of discriminative features for training the DNN model can be challenging due to the variability in HVD manifestations.

Various feature extraction methods [14, 70, 72, 150] reported in the literature have shown promising results in detecting HMs using supervised machine learning (ML) algorithms. However, the pathological variability and artifacts in the PCG signal make extracting meaningful features challenging. Additionally, classical ML and deep learning (DL) models cannot exploit the temporal information in the PCG signal, further limiting their diagnostic potential. To address these challenges, we proposed an acoustic-feature fusion approach in Chapter 3 that leverage the DL model to exploit the temporal information in the PCG signal to classify HMs. The proposed method extracts MFCC and LPCC features. Then, these two features are fused using an HLSTM network with a self-attention module to classify HMs. The method was evaluated using two publicly available databases, CICC 2016 and HSM. The proposed method performed better than the baseline methods reported in the literature. It achieved an OA of 98.76% and 99.10% for CICC 2016 and HSM databases, respectively. The robustness of the proposed method was also evaluated by adding noise of different SNR levels to the PCG signal. The proposed method had an ART of 2.61 ms and an NMP of 226106 for the HSM database. Similarly, the proposed method had an ART of 2.39 ms and an NMP of 225803 for the CICC 2016 database. The proposed method is computationally efficient for clinical deployment and will provide a promising solution to PCG-based HVD diagnosis challenges.

The existing WT-based methods to classify HMs, lack resolution at higher subbands (coarse scales) and shift-invariant properties [59, 61, 62, 72]. Further, the ML classifiers used with the wavelet features cannot exploit the temporal variation in the WT subbands to detect HMs. In Chapter 4, we proposed an SWT-based decomposition of PCG signals to address the issue. Then, to exploit the temporal variation of the SWT subbands, we deployed subband-specific HLSTM networks. Further, to enhance the detection of HMs, we proposed intra-subband and inter-subband attention modules. The proposed method was tested and verified using two publicly available databases, CICC 2016 and HSM. The results showed superior performance compared to baseline methods reported in the literature. The proposed method achieved an OA of 98.55% for the CICC 2016 and 99.47% for the HSM. The robustness of the proposed method is verified by adding different environmental and body-generated noises to the PCG signals with different SNR levels. The proposed method had an ART of 10.14 ms and an NMP of 909712 for the HSM database. Similarly, the proposed method had an ART of 9.96 ms and an NMP of 909559 for the CICC 2016 database. However, selecting the optimal number of SWT subbands can affect the diagnosis of HMs. The optimal number of subbands should ensure that the characteristic features of heart murmurs are effectively captured. More subbands might lead to smaller subband signals, making them more noise-resistant. Fewer subbands may provide more energy per subband but can result in spectral leakage and may not effectively isolate noise from the signal. Further, a higher number of subbands increases computational complexity. Therefore, balancing computational efficiency and detection accuracy is essential, especially in real-time applications. Thus, each PCG segment is decomposed into five levels in this work. It is so because the fundamental HSs and murmur information are primarily present in frequency bands (> 20Hz). There is no need for further decomposition at a coarse scale. Overall, the proposed method provides a promising solution to the challenges encountered in PCG-based HVDs diagnosis and is computationally efficient for clinical deployment.

Most existing automated methods for detecting HVDs focus on classifying PCG signals as healthy or pathological. However, multi-class analysis of HM severity stages is limited, mainly due to a lack of data for specific HVDs. In Chapter 5, we proposed an end-to-end framework to classify three HMs severity stages, namely absent HMs (AHM) or healthy, soft heart murmurs (SHM), and loud heart murmurs (LHM), along with the low-quality PCGs (LQP). Recent approaches in deep CNN models have used transfer-domain techniques, converting PCG segments into images and feeding them into a 2D-CNN model to detect HVDs. However, representing the time-series PCG signal as a 2D image may not capture the relevant heart sound (HS) information. Therefore, we proposed a multi-kernel residual CNN (MK-RCNN) model to extract deep features and efficiently classify the HM severity stages. The MK-RCNN model uses multi-kernel CNNs of different filter sizes to capture a wide range of features from the PCG signal. Further, it employed residual learning to extract deep features from CNN layers without degrading performance accuracy. The proposed model achieves an OA of 98.33%, 98.53%, and 100% using the CCDP, CICC 2016, and HSM databases, respectively. The proposed model also shows ART (ms) of 2.36, 2.82, and 2.47 using the CCDP, CICC 2016, and HSM databases, respectively. The proposed model's reasonable ART (ms) and excellent OA (%) make it suitable for faster and better diagnosis of HMs in primary cardiac care units.

The thesis proposed four automated methods for diagnosing HVDs using PCG signals. Chapter 2 proposed a multi-component oscillatory model to classify PCG cycles as healthy or pathological. In Chapter 3, an acoustic feature-fusion using the attention-based HLSTM network was proposed to classify HMs by exploiting the temporal information in the PCG signal. In Chapter 4, an SWT-based decomposition and subband-specific HLSTM networks were proposed to exploit the temporal variation in the SWT subbands to classify HMs. Further, intra-subband and inter-subband attention modules were introduced to enhance the HM classification. In Chapter 5, an automated method to classify three HMs severity stages, namely absent HMs (AHM) or healthy, soft heart murmurs (SHM), and loud heart murmurs (LHM), was proposed using a multi-kernel residual CNN (MK-RCNN) model. All proposed methods have shown superior performance compared to baseline methods reported in the literature.

The practical integration of our proposed diagnostic frameworks into real-life clinical practices is essential in translating our research into tangible healthcare benefits. We can extend the reach of cardiac assessments by implementing these frameworks on portable or mobile devices or integrating them with existing cardiac diagnostic systems. This expansion empowers early detection and intervention, which is crucial for effectively managing HVDs. In the industrial landscape, companies such as Eko Health [1] and AI Health Highway [2] are at the forefront of advancing digital stethoscopes by integrating artificial intelligence (AI) technology. In a clinical setting, the process begins with acquiring PCG data from patients using an electronic stethoscope, adhering to standardized recording protocols. The acquired data is then processed through our

---

[1] [Online Available]: https://www.ekohealth.com/

[2] [Online Available]: https://aisteth.com/

preprocessing and noise mitigation techniques. Clinicians can easily access and visualize PCG data while efficiently store patient-specific information. Furthermore, our proposed frameworks function as diagnostic decision-making tools, assisting healthcare providers in interpreting PCG data, delivering automated assessments, and promptly identifying potential HVDs.

The proposed frameworks may be used to detect other heart diseases using the PCG signal rather than that of only HVDs. The PCG signal contains information about the heart's mechanical sounds and surrounding structures, including the heart valves. PCG signals can help diagnose congenital heart defects (CHDs), which involve structural abnormalities in the heart. Similarly, PCG can help identify irregular heart rhythms, including arrhythmias, which can point to various cardiac conditions. Heart chamber size and function abnormalities can be detected through PCG, indicating conditions such as cardiomyopathy. PCG signals may offer insights into pericardial diseases, which affect the pericardium surrounding the heart. PCG can be used to assess overall cardiac function, including cardiac output and ejection fraction measures, which are essential for understanding heart health. However, at present, we do not have any publicly available PCG databases with annotations of other cardiac diseases than that of HVDs to perform the study.

6.1 Future Directions

Based on the work proposed in the thesis for automated diagnosis of HVDs from PCG signals using DL models, some future directions might further improve the clinical decision-making process for HVDs detection.

- Large-scale data collection and annotation are critical for developing accurate and reliable HVD detection systems. In order to train deep learning models for HVD detection, a large number of labeled data is required. The more data available, the more accurately the model can identify patterns and make predictions. Additionally, the diversity of data is also essential for more robust detection. HVDs present differently across individuals, and the disease manifestation can vary based on age, sex, and other factors. Therefore, data collection from diverse populations is necessary for training models that generalize to new patients.

- Multi-channel and multi-label PCG recording can provide additional information and help improve the accuracy of HVDs detection. Multi-channel recording involves using multiple sensors to capture the PCG signal from different locations on the chest. This can provide information on the spatial distribution of heart sounds and murmurs, which can be used to improve the accuracy of diagnosis. For example, in the case of mitral regurgitation, the location and timing of the murmur can provide information on the severity and cause of the condition. Multi-label recording involves labeling PCG recordings with multiple labels indicating the presence of different HVDs. This can help in the detection of co-occurring HVDs and improve the accuracy of diagnosis. For example, a patient with mitral stenosis and aortic regurgitation may have overlapping symptoms, making it difficult to diagnose the individual conditions accurately. However, with multi-label recording, the presence of both conditions can be identified and diagnosed more accurately.

- Model interpretability of automated systems can help clinicians understand the PCG features relevant to diagnose HVDs. This can also provide insights into the physiology of HVDs and improve medical understanding of the disease. Several methods exist for interpreting DL models, such as feature visualization, saliency maps, and activation maximization. These methods help visualize the regions of the PCG signal that are most relevant for the model's decision-making. Model interpretability can enhance DL models' trust, transparency, and clinical acceptance for HVDs detection.

- Integrating flexible analytic wavelet transform (FAWT) and empirical wavelet transform (EWT) into the automated diagnosis of HVDs from PCG signals can hold significant promise. These adaptive signal processing techniques can precisely capture and analyze the time-varying characteristics of PCG signals. FAWT's flexibility and adaptability to varying frequency components and EWT's ability to decompose signals into intrinsic mode functions (IMFs) make them ideal candidates for feature extraction, noise reduction, and time-frequency localization, all of which are critical in identifying pathological HVD patterns. Incorporating these wavelet transforms into the diagnostic framework makes it possible to enhance the accuracy and reliability of HVD detection. The synergistic combination of FAWT and EWT with deep learning models can pave the way for robust diagnostic tools for the early detection of HVDs in clinical practice.

- Continuous monitoring of PCG signals can provide several benefits for the early detection and management of HVDs. It can allow for the timely identification of abnormal HSs and enable clinicians to provide immediate interventions. Continuous PCG signal monitoring can also help track disease progression and treatment efficacy over time. Additionally, it can assist in predicting potential complications. Continuous monitoring of PCG signals can be achieved using wearable devices or implantable sensors, which can transmit data to healthcare professionals remotely. This can improve patient outcomes and reduce the need for frequent hospital visits, thereby improving the quality of life for patients with HVDs.

A

Deep Neural Network (DNN)

Deep Neural Network (DNN) is a type of deep learning (DL) model [169]. They are composed of multiple layers of interconnected neurons. Figure A.1 shows the model architecture of a DNN model. It consists of an input layer, two hidden layers, and one output layer. They are designed to mimic the structure and functionality of the human brain. It enables them to learn and make predictions from complex data. The detailed description of the different components of the DNN is as follows.

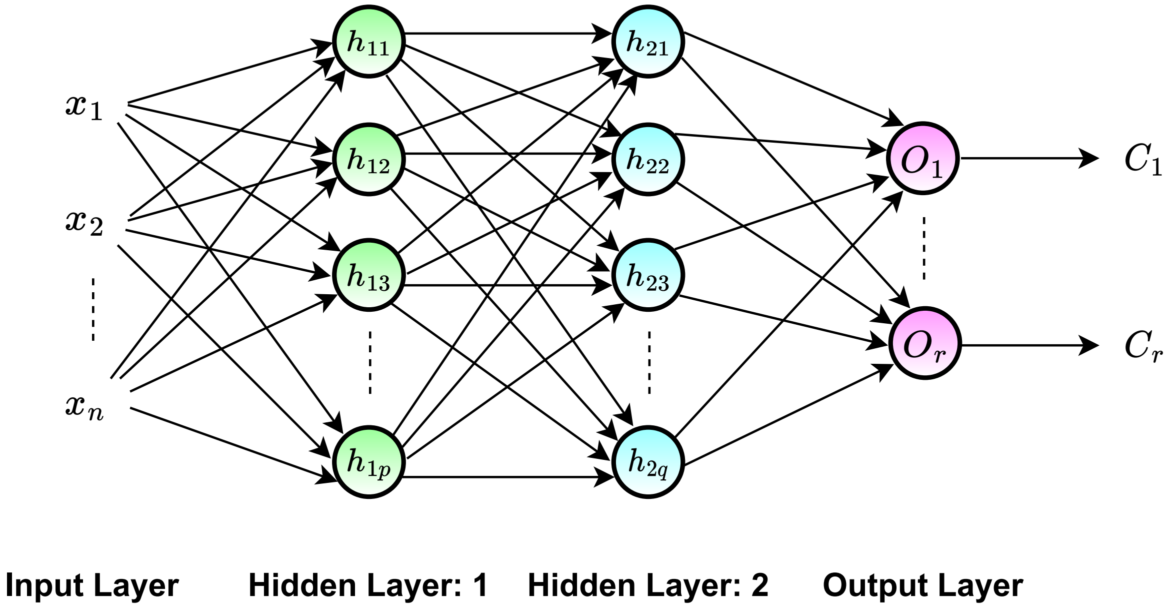

Figure A.1: The model architecture of a deep neural network (DNN) model.

- **Neurons:** They are the basic building blocks of a DNN. They receive input signals, perform computations, and produce output signals. Each neuron is associated with a weight vector, which determines the strength of the connections between neurons. The computation performed by a neuron can be represented as follows:

$$z = W \cdot X + b = \sum_{i=1}^{n} (w_i \cdot x_i) + b \tag{A.1}$$

$$a = f(z) \tag{A.2}$$

where $X = \{x_1, x_2, ..., x_n\}$; represents the input vector, $W = \{w_1, w_2, ..., w_n\}$; denotes the weight vector, b is the bias term, $f(.)$ is the activation function, z represents the weighted sum of inputs, and a denotes the output activation of the neuron.

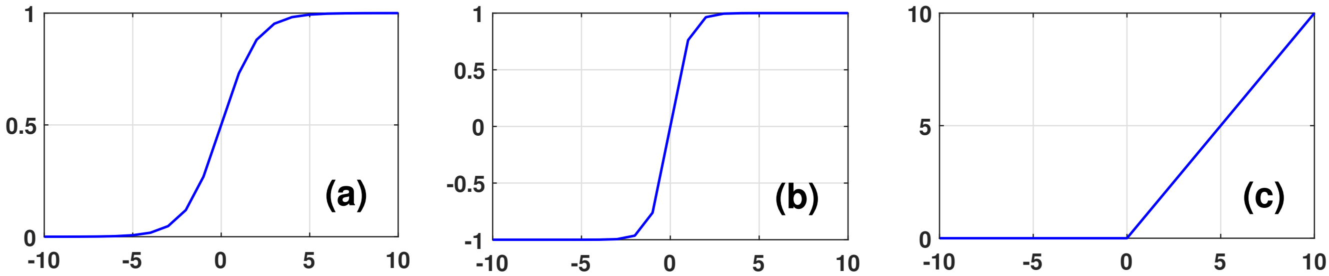

Figure A.2: Activation functions used in DNN models. (a) sigmoid activation function, (b) hyperbolic tangent activation function, and (c) rectified linear unit activation function.

- **DNN Layers:** Each layer of a DNN consists of multiple neurons. The neurons within a layer are connected with every neuron in the previous and subsequent layers. There are three types of layers in a DNN model:
  - (i) **Input Layer:** It receives input from the input signal fed to the DNN and passes it to subsequent layers. Figure A.1 shows the input layer having n inputs, i.e., $x_1, x_2, ..., x_n$.
  - (ii) **Hidden Layers:** They are intermediate layers between the input and output layers. They extract relevant features from the input data through non-linear transformations. Figure A.1 shows two hidden layers having p and q number of inputs.
  - (iii) **Output Layer:** They produce the final predictions or outputs of the DNN. The number of neurons in the output layer depends on the nature of the task, such as binary classification, multi-class classification, or regression. Figure A.1 shows the output layer having r outputs, i.e., $O_1, O_2, ..., O_r$.

- **Activation Functions:** They introduce non-linearity into the DNN. They enable the DNN to model complex relationships between inputs and outputs. Figure A.2(a), A.2(b), and A.2(c) show the sigmoid, hyperbolic tangent, and rectified linear unit activation functions, respectively. A brief description of these activation functions are as follows.
  - (i) **Sigmoid Function:** It squashes the input to a value between 0 and 1, making it suitable for binary classification tasks.

$$f(z) = \frac{1}{(1 + e^{-z})} \tag{A.3}$$

  - (ii) **Hyperbolic Tangent (tanh) Function:** It maps the input to a value between -1 and 1, making it useful for tasks with centered outputs.

$$f(z) = \frac{(e^z - e^{-z})}{(e^z + e^{-z})} \tag{A.4}$$

(iii) **Rectified Linear Unit (ReLU):** It sets negative inputs to zero and keeps positive inputs unchanged. It helps alleviate the vanishing gradient problem and speeds up training [170].

$$f(z) = \text{max}(0, z) \tag{A.5}$$

- **Loss functions:** They quantify the difference between the predicted outputs of the DNN and the true values, providing a measure of the model's performance. The choice of loss function depends on the nature of the task. Commonly used loss functions include:

  (i) **Mean Squared Error (MSE):** It calculates the average squared difference between predicted and true values, often used in regression problems. It is formulated as follows.

  $$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}(y_{t_n} - y_{p_n})^2 \tag{A.6}$$

  where N is the total number of data points. y_{t_n} and y_{p_n} are the true and predicated values of the n^{th} data point, respectively.

  (ii) **Cross-Entropy Loss:** It measures the dissimilarity between predicted probabilities and true class labels, typically used in classification tasks with multiple classes. It is formulated as follows [166].

  $$\mathbb{L}(y_t, y_p) = \sum_{n=1}^{N}(y_{t_n} \cdot \log y_{p_n}) \tag{A.7}$$

  where $\mathbb{L}$ is the cross-entropy loss function. Binary cross-entropy loss is used for binary classification problems. It is formulated as follows.

  $$\mathbb{L}(y_t, y_p) = -\sum_{n=1}^{N}[y_{t_n} \cdot \log(y_{p_n}) + (1 - y_{t_n}) \cdot \log(1 - y_{p_n})] \tag{A.8}$$

- **Optimization Algorithms:** These algorithms are responsible for adjusting the weights and biases of the DNN during the training process. They aim to minimize the loss function and find the optimal set of parameters. Some commonly used optimization algorithms include:

  (i) **Stochastic Gradient Descent (SGD):** It updates the weights by computing the gradient of the loss function with respect to the parameters for each training data or a batch-size.

$$W_{new} = W_{old} - \eta \cdot \nabla\{\mathbb{L}[W_{old}]\} \tag{A.9}$$

where W_{new} is the updated weighted vector, W_{old} is the existing weight vector, η is the learning rate, and ∇ is the gradient operator.

(ii) **Adam:** It combines the advantages of adaptive learning rates and momentum. It adjusts the learning rate for each parameter based on their past gradients and velocities [171].

$$m_{new} = \beta_1 \cdot m_{old} + (1 - \beta_1) \cdot \nabla\{\mathbb{L}[W_{old}]\} \tag{A.10}$$

$$v_{new} = \beta_2 \cdot v_{old} + (1 - \beta_2) \cdot \nabla\{\mathbb{L}[W_{old}^2]\} \tag{A.11}$$

$$W_{new} = W_{old} - \eta \cdot \frac{m_{new}}{\sqrt{v_{new}} + \epsilon} \tag{A.12}$$

References

[1] J. Hu, X. Cui, Y. Gong, X. Xu, B. Gao, T. Wen, T. J. Lu, and F. Xu, "Portable microfluidic and smartphone-based devices for monitoring of cardiovascular diseases at the point of care," *Biotechnology advances*, vol. 34, no. 3, pp. 305–320, 2016.

[2] W. H. O. (World Health Organization), "Fact Sheet: cardiovascular disease. Online Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)," [updated: 2021.6.11, accessed: 2022.6.14].

[3] R. S. Dilmaghani, H. Bobarshad, M. Ghavami, S. Choobkar, and C. Wolfe, "Wireless sensor networks for monitoring physiological signals of multiple patients," *IEEE Transactions on biomedical circuits and systems*, vol. 5, no. 4, pp. 347–356, 2011.

[4] A. Leatham, "Phonocardiography," *British medical bulletin*, vol. 8, no. 4, pp. 333–342, 1952.

[5] J. Leal, R. Luengo-Fernández, A. Gray, S. Petersen, and M. Rayner, "Economic burden of cardiovascular diseases in the enlarged european union," *European heart journal*, vol. 27, no. 13, pp. 1610–1619, 2006.

[6] J. Narula, Y. Chandrashekhar, and E. Braunwald, "Time to add a fifth pillar to bedside physical examination: inspection, palpation, percussion, auscultation, and insonation," *JAMA cardiology*, vol. 3, no. 4, pp. 346–350, 2018.

[7] S. Mendis, P. Puska, B. e. Norrving, W. H. Organization, *et al.*, *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

[8] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds–a systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2018.

[9] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of pcg signals," *IEEE Access*, vol. 8, pp. 160 882–160 890, 2020.

[10] R. Rangayan, "Biomedical signal analysis: A practical approach," 2007.

[11] N. M. Hoeting, C. E. McCracken, M. McConnell, D. Sallee, G. J. Iannucci, and M. E. Oster, "Systolic ejection click versus split first heart sound: Are our ears deceiving us?" *Congenital Heart Disease*, vol. 12, no. 4, pp. 417–420, 2017.

[12] D. Riknagel, H. Zimmermann, R. Farlie, D. Hammershøi, S. E. Schmidt, M. Hedegaard, P. Humaidan, and J. J. Struijk, "Separation and characterization of maternal cardiac and vascular sounds in the third trimester of pregnancy," *International Journal of Gynecology & Obstetrics*, vol. 137, no. 3, pp. 253–259, 2017.

[13] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson, *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological measurement*, vol. 37, no. 12, p. 2181, 2016.

[14] S. Li, F. Li, S. Tang, and F. Luo, "Heart sounds classification based on feature fusion using lightweight neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[15] R. M. Rangayyan, *Biomedical signal analysis*. John Wiley & Sons, 2015.

[16] J. D. Pollock and A. N. Makaryus, "Physiology, cardiac cycle," in *StatPearls [Internet]*. StatPearls Publishing, 2021.

[17] J. E. Frank and K. M. Jacobe, "Evaluation and management of heart murmurs in children," *American family physician*, vol. 84, no. 7, pp. 793–800, 2011.

[18] B. A. Carabello and W. J. Paulus, "Aortic stenosis," *The lancet*, vol. 373, no. 9667, pp. 956–966, 2009.

[19] E. Hayek, C. N. Gring, and B. P. Griffin, “Mitral valve prolapse,” *The Lancet*, vol. 365, no. 9458, pp. 507–518, 2005.

[20] J. K. Perloff and W. C. Roberts, “The mitral apparatus: functional anatomy of mitral regurgitation,” *Circulation*, vol. 46, no. 2, pp. 227–239, 1972.

[21] Y. Chandrashekhar, S. Westaby, and J. Narula, “Mitral stenosis,” *The Lancet*, vol. 374, no. 9697, pp. 1271–1283, 2009.

[22] L. Wierwille, “Pediatric heart murmurs: evaluation and management in primary care,” *The Nurse Practitioner*, vol. 36, no. 3, pp. 22–29, 2011.

[23] K. Basak, S. Mandal, M. Manjunatha, J. Chatterjee, and A. Ray, “Phonocardiogram signal analysis using adaptive line enhancer methods on mixed signal processor,” in *2010 International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2010, pp. 1–5.

[24] S. Mandal, K. Basak, K. Mandana, A. K. Ray, J. Chatterjee, and M. Mahadevappa, “Development of cardiac prescreening device for rural population using ultralow-power embedded system,” *IEEE transactions on biomedical engineering*, vol. 58, no. 3, pp. 745–749, 2010.

[25] S. R. Messer, J. Agzarian, and D. Abbott, “Optimal wavelet denoising for phonocardiograms,” *Microelectronics journal*, vol. 32, no. 12, pp. 931–941, 2001.

[26] D. Boutana, M. Benidir, and N. Kouras, “The empirical mode decomposition and the svd for abnormal heart sound signals detection and time-frequency analysis,” *IET*, 2015.

[27] A. Gavrovska, M. Slavković, I. Reljin, and B. Reljin, “Application of wavelet and emd-based denoising to phonocardiograms,” in *International Symposium on Signals, Circuits and Systems ISSCS2013*. IEEE, 2013, pp. 1–4.

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

[29] D. B. Springer, L. Tarassenko, and G. D. Clifford, “Logistic regression-hsmm-based heart sound segmentation,” *IEEE transactions on biomedical engineering*, vol. 63, no. 4, pp. 822–832, 2015.

[30] R. J. Lehner and R. M. Rangayyan, “A three-channel microcomputer system for segmentation and characterization of the phonocardiogram,” *IEEE Transactions on Biomedical Engineering*, no. 6, pp. 485–489, 1987.

[31] F. Plesinger, I. Viscor, J. Halamek, J. Jurco, and P. Jurak, “Heart sounds analysis using probability assessment,” *Physiological measurement*, vol. 38, no. 8, p. 1685, 2017.

[32] P. Sharma, S. A. Imtiaz, and E. Rodriguez-Villegas, “An algorithm for heart rate extraction from acoustic recordings at the neck,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 246–256, 2018.

[33] A. A. Sepehri, A. Gharehbaghi, T. Dutoit, A. Kocharian, and A. Kiani, “A novel method for pediatric heart sound segmentation without using the ecg,” *Computer methods and programs in biomedicine*, vol. 99, no. 1, pp. 43–48, 2010.

[34] V. N. Varghees and K. Ramachandran, “A novel heart sound activity detection framework for automated heart sound analysis,” *Biomedical Signal Processing and Control*, vol. 13, pp. 174–188, 2014.

[35] Y. Zheng, X. Guo, and X. Ding, “A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2710–2721, 2015.

[36] S. Patidar and R. B. Pachori, “Segmentation of cardiac sound signals by removing murmurs using constrained tunable-q wavelet transform,” *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 559–567, 2013.

[37] D. Boutana, M. Benidir, and B. Barkat, “Segmentation and time-frequency analysis of pathological heart sound signals using the emd method,” in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 1437–1441.

[38] E. Messner, M. Zöhrer, and F. Pernkopf, “Heart sound segmentation—an event detection approach using deep recurrent neural networks,” *IEEE transactions on biomedical engineering*, vol. 65, no. 9, pp. 1964–1974, 2018.

[39] T. Fernando, H. Ghaemmaghami, S. Denman, S. Sridharan, N. Hussain, and C. Fookes, “Heart sound segmentation using bidirectional lstms with attention,” *IEEE journal of biomedical and health informatics*, vol. 24, no. 6, pp. 1601–1609, 2019.

[40] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, “A robust interpretable deep learning classifier for heart anomaly detection without segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2162–2171, 2020.

[41] S. Chauhan, P. Wang, C. S. Lim, and V. Anantharaman, “A computer-aided mfcc-based hmm system for automatic auscultation,” *Computers in biology and medicine*, vol. 38, no. 2, pp. 221–233, 2008.

[42] E. Soares, P. Angelov, and X. Gu, “Autonomous learning multiple-model zero-order classifier for heart sound classification,” *Applied Soft Computing*, vol. 94, p. 106449, 2020.

[43] S. Gómez-Quintana, C. E. Schwarz, I. Shelevytsky, V. Shelevytska, O. Semenova, A. Factor, E. Popovici, and A. Temko, “A framework for ai-assisted detection of patent ductus arteriosus from neonatal phonocardiogram,” in *Healthcare*, vol. 9, no. 2. MDPI, 2021, p. 169.

[44] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, “Phonocardiographic signal analysis method using a modified hidden markov model,” *Annals of Biomedical Engineering*, vol. 35, no. 3, pp. 367–374, 2007.

[45] F. A. Khan, A. Abid, and M. S. Khan, “Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features,” *Physiological measurement*, vol. 41, no. 5, p. 055006, 2020.

[46] O. El Badlaoui, A. Benba, and A. Hammouch, “Novel pcg analysis method for discriminating between abnormal and normal heart sounds,” *Irbm*, vol. 41, no. 4, pp. 223–228, 2020.

[47] Y. Chen, S. Wang, C.-H. Shen, and F. K. Choy, “Matrix decomposition based feature extraction for murmur classification,” *Medical engineering & physics*, vol. 34, no. 6, pp. 756–761, 2012.

[48] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, “A framework for automatic heart sound analysis without segmentation,” *Biomedical engineering online*, vol. 10, no. 1, pp. 1–23, 2011.

[49] C. Kwak and O.-W. Kwon, “Cardiac disorder classification by heart sound signals using murmur likelihood and hidden markov model state likelihood,” *IET signal processing*, vol. 6, no. 4, pp. 326–334, 2012.

[50] R. G. Sæderup, P. Hoang, S. Winther, M. Bøttcher, J. Struijk, S. Schmidt, and J. Østergaard, “Estimation of the second heart sound split using windowed sinusoidal models,” *Biomedical Signal Processing and Control*, vol. 44, pp. 229–236, 2018.

[51] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, and D. V. Anderson, “Combining sparse coding and time-domain features for heart sound classification,” *Physiological measurement*, vol. 38, no. 8, p. 1701, 2017.

[52] S. Sun, H. Wang, Z. Chang, B. Mao, and Y. Liu, “On the mahalanobis distance classification criterion for a ventricular septal defect diagnosis system,” *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2665–2674, 2018.

[53] P. Samanta, A. Pathak, K. Mandana, and G. Saha, “Classification of coronary artery diseased and normal subjects using multi-channel phonocardiogram signal,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 2, pp. 426–443, 2019.

[54] F. Noman, S.-H. Salleh, C.-M. Ting, S. B. Samdin, H. Ombao, and H. Hussain, “A markov-switching model approach to heart sound segmentation and classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 705–716, 2019.

[55] A. Rujoie, A. Fallah, S. Rashidi, E. R. Khoshnood, and T. S. Ala, “Classification and evaluation of the severity of tricuspid regurgitation using phonocardiogram,” *Biomedical Signal Processing and Control*, vol. 57, p. 101688, 2020.

[56] G. Redlarski, D. Gradolewski, and A. Palkowski, “A system for heart sounds classification,” *PloS one*, vol. 9, no. 11, p. e112673, 2014.

[57] S. E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, and J. J. Struijk, “Acoustic features for the identification of coronary artery disease,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2611–2619, 2015.

[58] W. Zhang, J. Han, and S. Deng, “Heart sound classification based on scaled spectrogram and tensor decomposition,” *Expert Systems with Applications*, vol. 84, pp. 220–231, 2017.

[59] G.-Y. Son and S. Kwon, “Classification of heart sound signal using multiple features,” *Applied Sciences*, vol. 8, no. 12, p. 2344, 2018.

[60] F. Dong, K. Qian, Z. Ren, A. Baird, X. Li, Z. Dai, B. Dong, F. Metze, Y. Yamamoto, and B. W. Schuller, "Machine listening for heart status monitoring: Introducing and benchmarking hss—the heart sounds shenzhen corpus," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 2082–2092, 2019.

[61] Y. Wang, W. Li, J. Zhou, X. Li, and Y. Pu, "Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on oms-wpd," *Future Generation Computer Systems*, vol. 37, pp. 488–495, 2014.

[62] S. Patidar, R. B. Pachori, and N. Garg, "Automatic diagnosis of septal defects based on tunable-q wavelet transform of cardiac sound signals," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3315–3326, 2015.

[63] S. K. Ghosh, R. K. Tripathy, R. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the pcg signal using time-frequency magnitude and phase features," *IEEE Sensors Letters*, vol. 3, no. 12, pp. 1–4, 2019.

[64] S. K. Ghosh, R. Ponnalagu, R. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with pcg signals," *Computers in biology and medicine*, vol. 118, p. 103632, 2020.

[65] N. K. Sawant, S. Patidar, N. Nesaragi, and U. R. Acharya, "Automated detection of abnormal heart sound signals using fano-factor constrained tunable quality wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 111–126, 2021.

[66] J. S. Walker, *A primer on wavelets and their scientific applications*. CRC press, 2008.

[67] T. Kinnunen, R. Saeidi, F. Sedlák, K. A. Lee, J. Sandberg, M. Hansson-Sandsten, and H. Li, "Low-variance multitaper mfcc features: a case study in robust speaker verification," *IEEE transactions on audio, speech, and language processing*, vol. 20, no. 7, pp. 1990–2001, 2012.

[68] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9393–9400, 2018.

[69] W. Han, S. Xie, Z. Yang, S. Zhou, and H. Huang, "Heart sound classification using the snmfnet classifier," *Physiological measurement*, vol. 40, no. 10, p. 105003, 2019.

[70] M. Milani, P. E. Abas, L. C. De Silva, and N. D. Nanayakkara, "Abnormal heart sound classification using phonocardiography signals," *Smart Health*, vol. 21, p. 100194, 2021.

[71] V. Maknickas and A. Maknickas, "Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiological measurement*, vol. 38, no. 8, p. 1671, 2017.

[72] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection," *Computers in biology and medicine*, vol. 100, pp. 132–143, 2018.

[73] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition," *Speech communication*, vol. 54, no. 4, pp. 543–565, 2012.

[74] S. S. Stevens *et al.*, "A scale for the measurement of the psychological magnitude pitch," *The journal of the acoustical society of america*, vol. 8, no. 3, pp. 185–190, 1937.

[75] W.-W. Hung and H.-C. Wang, "On the use of weighted filter bank analysis for the derivation of robust mfccs," *IEEE signal processing letters*, vol. 8, no. 3, pp. 70–73, 2001.

[76] T. H. Joo, J. H. McClellan, R. A. Foale, G. S. Myers, and R. S. Lees, "Pole-zero modeling and classification of phonocardiograms," *IEEE Transactions on Biomedical Engineering*, no. 2, pp. 110–118, 1983.

[77] J. Xu, L.-G. Durand, and P. Pibarot, "Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, pp. 277–283, 2001.

[78] X. Zhang, L. Durand, L. Senhadji, H. C. Lee, and J.-L. Coatrieux, "Analysis-synthesis of the phonocardiogram based on the matching pursuit method," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 8, pp. 962–971, 1998.

[79] S. Jabbari and H. Ghassemian, "Modeling of heart systolic murmurs based on multivariate matching pursuit for diagnosis of valvular disorders," *Computers in biology and medicine*, vol. 41, no. 9, pp. 802–811, 2011.

[80] J. Martínez-Alajarín and R. Ruiz-Merino, “Wavelet and wavelet packet compression of phonocardiograms,” *Electronics Letters*, vol. 40, no. 17, p. 1, 2004.

[81] M. S. Manikandan and S. Dandapat, “Wavelet energy based compression of phonocardiogram (pcg) signal for telecardiology,” *IET*, 2007.

[82] A. Baykal, Y. Z. Ider, and H. Koymen, “Distribution of aortic mechanical prosthetic valve closure sound model parameters on the surface of the chest,” *IEEE transactions on biomedical engineering*, vol. 42, no. 4, pp. 358–370, 1995.

[83] H. Tang, J. Zhang, J. Sun, T. Qiu, and Y. Park, “Phonocardiogram signal compression using sound repetition and vector quantization,” *Computers in biology and medicine*, vol. 71, pp. 24–34, 2016.

[84] R. C. King, E. Villeneuve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin, “Application of data fusion techniques and technologies for wearable health monitoring,” *Medical engineering & physics*, vol. 42, pp. 1–12, 2017.

[85] A. Sengur and I. Turkoglu, “A hybrid method based on artificial immune system and fuzzy k-nn algorithm for diagnosis of heart valve diseases,” *Expert Systems with Applications*, vol. 35, no. 3, pp. 1011–1020, 2008.

[86] A. Quiceno-Manrique, J. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, “Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic signals,” *Annals of biomedical engineering*, vol. 38, no. 1, pp. 118–137, 2010.

[87] S. Choi and Z. Jiang, “Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique,” *Computers in biology and medicine*, vol. 40, no. 1, pp. 8–20, 2010.

[88] E. Avci, “A new intelligent diagnosis system for the heart valve diseases by using genetic-svm classifier,” *Expert Systems with Applications*, vol. 36, no. 7, pp. 10 618–10 626, 2009.

[89] S. A. Pavlopoulos, A. C. Stasis, and E. N. Loukis, “A decision tree–based method for the differential diagnosis of aortic stenosis from mitral regurgitation using heart sounds,” *Biomedical engineering online*, vol. 3, no. 1, pp. 1–15, 2004.

[90] L. Atallah, B. Lo, R. King, and G.-Z. Yang, “Sensor positioning for activity recognition using wearable accelerometers,” *IEEE transactions on biomedical circuits and systems*, vol. 5, no. 4, pp. 320–329, 2011.

[91] A. Mannini and A. M. Sabatini, “Machine learning methods for classifying human physical activity from on-body accelerometers,” *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.

[92] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, “Neural network classification of homomorphic segmented heart sounds,” *Applied soft computing*, vol. 7, no. 1, pp. 286–297, 2007.

[93] S. Kang, R. Doroshow, J. McConnaughey, and R. Shekhar, “Automated identification of innocent still’s murmur in children,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1326–1334, 2016.

[94] T.-E. Chen, S.-I. Yang, L.-T. Ho, K.-H. Tsai, Y.-H. Chen, Y.-F. Chang, Y.-H. Lai, S.-S. Wang, Y. Tsao, and C.-C. Wu, “S1 and s2 heart sound recognition using deep neural networks,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 372–380, 2016.

[95] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, “Heart sound anomaly and quality detection using ensemble of neural networks without segmentation,” in *2016 computing in cardiology conference (CinC)*. IEEE, 2016, pp. 613–616.

[96] W. Zhang, J. Han, and S. Deng, “Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation,” *Biomedical Signal Processing and Control*, vol. 53, p. 101560, 2019.

[97] O. Deperlioglu, “Classification of phonocardiograms with convolutional neural networks,” *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 9, no. 2, pp. 22–33, 2018.

[98] P. Dhar, S. Dutta, and V. Mukherjee, “Cross-wavelet assisted convolution neural network (alexnet) approach for phonocardiogram signals classification,” *Biomedical Signal Processing and Control*, vol. 63, p. 102142, 2021.

[99] A. Pathak, K. Mandana, and G. Saha, “Ensembled transfer learning and multiple kernel learning for phonocardiogram based atherosclerotic coronary artery disease detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2804–2813, 2022.

[100] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *2016 computing in cardiology conference (CinC)*. IEEE, 2016, pp. 621–624.

[101] H. Li, X. Wang, C. Liu, Y. Wang, P. Li, H. Tang, L. Yao, and H. Zhang, "Dual-input neural network integrating feature extraction and deep learning for coronary artery disease detection using electrocardiogram and phonocardiogram," *IEEE Access*, vol. 7, pp. 146 457–146 469, 2019.

[102] A. Gharehbaghi and M. Lindén, "A deep machine learning method for classifying cyclic time series of biological signals using time-growing neural network," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 9, pp. 4102–4115, 2017.

[103] W. Zeng, J. Yuan, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "A new approach for the detection of abnormal heart sound signals using tqwt, vmd and neural networks," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1613–1647, 2021.

[104] S. Das, S. Pal, and M. Mitra, "Deep learning approach of murmur detection using cochleagram," *Biomedical Signal Processing and Control*, vol. 77, p. 103747, 2022.

[105] E. Kay and A. Agarwal, "Dropconnected neural network trained with diverse features for classifying heart sounds," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 617–620.

[106] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, and R. Getz, "The pascal classifying heart sounds challenge 2011 (chsc2011) results," *See http://www. peterjbentley. com/heartchallenge/index. html*, 2011.

[107] O. Deperlioglu, "Classification of segmented phonocardiograms by convolutional neural networks," *BRAIN. Broad research in artificial intelligence and neuroscience*, vol. 10, no. 2, pp. 5–13, 2019.

[108] F. Li, M. Liu, Y. Zhao, L. Kong, L. Dong, X. Liu, and M. Hui, "Feature extraction and classification of heart sound using 1d convolutional neural networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2019, no. 1, pp. 1–11, 2019.

[109] X. Cheng, J. Huang, Y. Li, and G. Gui, "Design and application of a laconic heart sound neural network," *IEEE Access*, vol. 7, pp. 124 417–124 425, 2019.

[110] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE journal of biomedical and health informatics*, vol. 24, no. 8, pp. 2189–2198, 2020.

[111] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, "Classification of heart sounds using convolutional neural network," *Applied Sciences*, vol. 10, no. 11, p. 3956, 2020.

[112] A. M. Alqudah, H. Alquran, and I. A. Qasmieh, "Classification of heart sound short records using bispectrum analysis approach images and deep learning," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, pp. 1–16, 2020.

[113] Y. Chen, S. Wei, and Y. Zhang, "Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network," *Medical & Biological Engineering & Computing*, vol. 58, pp. 2039–2047, 2020.

[114] H. Kui, J. Pan, R. Zong, H. Yang, and W. Wang, "Heart sound classification based on log mel-frequency spectral coefficients features and convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 69, p. 102893, 2021.

[115] K. N. Khan, F. A. Khan, A. Abid, T. Olmez, Z. Dokur, A. Khandakar, M. E. Chowdhury, and M. S. Khan, "Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning," *Physiological measurement*, vol. 42, no. 9, p. 095003, 2021.

[116] T. Li, Y. Yin, K. Ma, S. Zhang, and M. Liu, "Lightweight end-to-end neural network model for automatic heart sound classification," *Information*, vol. 12, no. 2, p. 54, 2021.

[117] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, "Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors," *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 1, pp. 24–34, 2017.

[118] J. M.-T. Wu, M.-H. Tsai, Y. Z. Huang, S. H. Islam, M. M. Hassan, A. Alelaiwi, and G. Fortino, "Applying an ensemble convolutional neural network with savitzky–golay filter to construct a phonocardiogram prediction model," *Applied Soft Computing*, vol. 78, pp. 29–40, 2019.

[119] H. Alaskar, N. Alzhrani, A. Hussain, and F. Almarshed, “The implementation of pretrained alexnet on pcg classification,” in *Intelligent Computing Methodologies: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part III 15*. Springer, 2019, pp. 784–794.

[120] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, “Heart sound classification based on improved mfcc features and convolutional recurrent neural networks,” *Neural Networks*, vol. 130, pp. 22–32, 2020.

[121] H. Li, X. Wang, C. Liu, Q. Zeng, Y. Zheng, X. Chu, L. Yao, J. Wang, Y. Jiao, and C. Karmakar, “A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection,” *Computers in biology and medicine*, vol. 120, p. 103733, 2020.

[122] M. Gjoreski, A. Gradišek, B. Budna, M. Gams, and G. Poglajen, “Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds,” *Ieee Access*, vol. 8, pp. 20 313–20 324, 2020.

[123] O. Deperlioglu, “Heart sound classification with signal instant energy and stacked autoencoder network,” *Biomedical Signal Processing and Control*, vol. 64, p. 102211, 2021.

[124] Y. Huang, H. Li, R. Tao, W. Han, P. Zhang, X. Yu, and R. Wu, “A customized framework for coronary artery disease detection using phonocardiogram signals,” *Biomedical Signal Processing and Control*, vol. 78, p. 103982, 2022.

[125] Q. Li, Y. Yang, T. Lan, H. Zhu, Q. Wei, F. Qiao, X. Liu, and H. Yang, “Msp-mfcc: Energy-efficient mfcc feature extraction method with mixed-signal processing architecture for wearable speech recognition applications,” *IEEE Access*, vol. 8, pp. 48 720–48 730, 2020.

[126] N. Baghel, M. K. Dutta, and R. Burget, “Automatic diagnosis of multiple cardiac diseases from pcg signals using convolutional neural network,” *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105750, 2020.

[127] M. Alkhodari and L. Fraiwan, “Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings,” *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105940, 2021.

[128] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, “Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings,” *IEEE Access*, vol. 9, pp. 36 955–36 967, 2021.

[129] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, and U. R. Acharya, “Classification of heart sound signals using a novel deep wavenet model,” *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105604, 2020.

[130] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, “Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption,” *Neurocomputing*, vol. 392, pp. 153–159, 2020.

[131] B. Xiao, Y. Xu, X. Bi, W. Li, Z. Ma, J. Zhang, and X. Ma, “Follow the sound of children’s heart: a deep-learning-based computer-aided pediatric chds diagnosis system,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1994–2004, 2019.

[132] R. Avanzato and F. Beritelli, “Heart sound multiclass analysis based on raw data and convolutional neural network,” *IEEE Sensors Letters*, vol. 4, no. 12, pp. 1–4, 2020.

[133] P. T. Krishnan, P. Balasubramanian, and S. Umapathy, “Automated heart sound classification system from unsegmented phonocardiogram (pcg) using deep neural network,” *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 505–515, 2020.

[134] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, “Heartbeat sound signal classification using deep learning,” *Sensors*, vol. 19, no. 21, p. 4819, 2019.

[135] J. Liu, H. Wang, Z. Yang, J. Quan, L. Liu, and J. Tian, “Deep learning-based computer-aided heart sound analysis in children with left-to-right shunt congenital heart disease,” *International journal of cardiology*, vol. 348, pp. 58–64, 2022.

[136] R. F. Ibarra-Hernández, M. A. Alonso-Arévalo, A. Cruz-Gutiérrez, A. L. Licona-Chávez, and S. Villarreal-Reyes, “Design and evaluation of a parametric model for cardiac sounds,” *Computers in biology and medicine*, vol. 89, pp. 170–180, 2017.

[137] T. Leung, P. White, J. Cook, W. Collis, E. Brown, and A. Salmon, "Analysis of the second heart sound for diagnosis of paediatric heart disease," *IEE Proceedings-Science, measurement and technology*, vol. 145, no. 6, pp. 285–290, 1998.

[138] H. Tang, T. Li, Y. Park, and T. Qiu, "Separation of heart sound signal from noise in joint cycle frequency–time–frequency domains based on fuzzy detection," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 10, pp. 2438–2447, 2010.

[139] N. P. Hughes and L. Tarassenko, "Probabilistic models for automated ecg interval analysis in phase 1 studies," Technical Report BSP 08-01, Available at http://www. robots. ox. ac. uk . . . , Tech. Rep., 2008.

[140] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[141] G. André, V. Kostrubiec, J.-C. Buisson, J.-M. Albaret, and P.-G. Zanone, "A parsimonious oscillatory model of handwriting," *Biological cybernetics*, vol. 108, no. 3, pp. 321–336, 2014.

[142] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.

[143] T. Kautz, B. M. Eskofier, and C. F. Pasluosta, "Generic performance measure for multiclass-classifiers," *Pattern Recognition*, vol. 68, pp. 111–125, 2017.

[144] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[145] R. Zhang, Y. Li, and X. Li, "Topology inference with network tomography based on t-test," *IEEE communications letters*, vol. 18, no. 6, pp. 921–924, 2014.

[146] I. J. D. Bobillo, "A tensor approach to heart sound classification," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 629–632.

[147] M. N. Homsi, N. Medina, M. Hernandez, N. Quintero, G. Perpiñan, A. Quintana, and P. Warrick, "Automatic heart sound recording classification using a nested set of ensemble algorithms," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 817–820.

[148] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[149] A. Chowdhury and A. Ross, "Fusing mfcc and lpc features using 1d triplet cnn for speaker recognition in severely degraded audio signals," *IEEE transactions on information forensics and security*, vol. 15, pp. 1616–1629, 2019.

[150] K. Iqtidar, U. Qamar, S. Aziz, and M. U. Khan, "Phonocardiogram signal analysis for classification of coronary artery diseases using mfcc and 1d adaptive local ternary patterns," *Computers in Biology and Medicine*, vol. 138, p. 104926, 2021.

[151] T. Barnwell, "Recursive windowing for generating autocorrelation coefficients for lpc analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 5, pp. 1062–1066, 1981.

[152] P. E. Papamichalis, *Practical approaches to speech coding*. Prentice-Hall, Inc., 1987.

[153] K. Greff *et al.*, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[154] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[155] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[156] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1806–1819, 2017.

[157] N. Mei, H. Wang, Y. Zhang, F. Liu, X. Jiang, and S. Wei, "Classification of heart sounds based on quality assessment and wavelet scattering transform," *Computers in Biology and Medicine*, vol. 137, p. 104814, 2021.

[158] M. Merah, T. Abdelmalik, and B. Larbi, “R-peaks detection based on stationary wavelet transform,” *Computer methods and programs in biomedicine*, vol. 121, no. 3, pp. 149–160, 2015.

[159] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, “Show, attend and tell: Neural image caption generation with visual attention,” in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.

[160] Y. LeCun *et al.*, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[161] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.” *Journal of machine learning research*, vol. 9, no. 11, 2008.

[162] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, *et al.*, “The circor digiscope dataset: from murmur detection to murmur classification,” *IEEE journal of biomedical and health informatics*, vol. 26, no. 6, pp. 2524–2535, 2021.

[163] R. K. Srivastava *et al.*, “Highway networks,” *arXiv preprint arXiv:1505.00387*, 2015.

[164] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[165] H. Wu, R. Lei, and Y. Peng, “Pcbnet: A lightweight convolutional neural network for defect inspection in surface mount technology,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[166] Y. Ho and S. Wookey, “The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling,” *IEEE Access*, vol. 8, pp. 4806–4813, 2019.

[167] A. Freeman and S. A. LEVINE, “The clinical significance of the systolic murmur: A study of 1000 consecutive” non-cardiac” cases,” *Annals of Internal Medicine*, vol. 6, no. 11, pp. 1371–1385, 1933.

[168] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, E. J. Benjamin, C. P. Benziger, *et al.*, “Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study,” *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020.

[169] G. Montavon, W. Samek, and K.-R. Müller, “Methods for interpreting and understanding deep neural networks,” *Digital signal processing*, vol. 73, pp. 1–15, 2018.

[170] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[171] Z. Zhang, “Improved adam optimizer for deep neural networks,” in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*. Ieee, 2018, pp. 1–2.

List of Publications

Journal Publications

- **S. Das**, D. Jyotishi, and S. Dandapat, “Automated Detection of Heart Valve Diseases Using Stationary Wavelet Transform and Attention Based Hierarchical LSTM Network,” **IEEE Transactions on Instrumentation and Measurement**, 2023.
- **S. Das**, D. Jyotishi, and S. Dandapat, “Heart Valve Diseases Detection Based on Feature-fusion and Hierarchical LSTM Network,” **IEEE Transactions on Instrumentation and Measurement**, 2022.
- **S. Das** and S. Dandapat, “Multicomponent Oscillatory Model based Heart Sound Classification,” **Journal of acoustic society of India (JASI)**, 2020.
- **S. Das** and S. Dandapat, “Heart Murmur Severity Stages Classification Using Multi-kernel Residual CNN,” **[Submitted, Under Review]**.

Conference Publications

- **S. Das** and S. Dandapat, “Synthesis and Classification of Heart Sounds Using Multi-component Oscillatory Model,” **2020 National Conference on Communications (NCC)**, Kharagpur, India, 2020.
- **S. Das** and S. Dandapat, “Automated Detection of Heart Murmurs From the PCG Signal Using Stationary Wavelet Transform,” **INDICON**, Kerala, India, 2022.
- **S. Das** and S. Dandapat, “Multiscale Kernel Residual Convolutional Neural Network to Detect Heart Valve Diseases,” **INDICON**, Kerala, India, 2022.
- S Kumari, D Jyotishi, **S. Das**, and S. Dandapat, “Analyzing The Effect of Segmentation on PCG Based Biometric System,” **INDICON**, Kerala, India, 2022.
- MJ Singh, **S. Das**, LN Sharma, and S. Dandapat, “Stationary Wavelet Transform Based Detection of Aortic Stenosis Using Seismocardiogram Signal,” **2023 National Conference on Communications (NCC)**, Guwahati, India, 2023.