Imperial College London
Department of Electrical and Electronic Engineering

Automatic Analysis and Classification of Cardiac Acoustic Signals for Long Term Monitoring

Amit Krishna Dwivedi

September 2020

Supervised by Prof. Esther Rodriguez-Villegas

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronic Engineering of Imperial College London and the Diploma of Imperial College London

Abstract

Objective: Cardiovascular diseases are the leading cause of death worldwide resulting in over 17.9 million deaths each year. Most of these diseases are preventable and treatable, but their progression and outcomes are significantly more positive with early-stage diagnosis and proper disease management. Among the approaches available to assist with the task of early-stage diagnosis and management of cardiac conditions, automatic analysis of auscultatory recordings is one of the most promising ones, since it could be particularly suitable for ambulatory/wearable monitoring. Thus, proper investigation of abnormalities present in cardiac acoustic signals can provide vital clinical information to assist long term monitoring. Cardiac acoustic signals, however, are very susceptible to noise and artifacts, and their characteristics vary largely with the recording conditions which makes the analysis challenging. Additionally, there are challenges in the steps used for automatic analysis and classification of cardiac acoustic signals. Broadly, these steps are the segmentation, feature extraction and subsequent classification of recorded signals using selected features. This thesis presents approaches using novel features with the aim to assist the automatic early-stage detection of cardiovascular diseases with improved performance, using cardiac acoustic signals collected in real-world conditions.

Methods: Cardiac auscultatory recordings were studied to identify potential features to help in the classification of recordings from subjects with and without cardiac diseases. The diseases considered in this study for the identification of the symptoms and characteristics are the valvular heart diseases due to stenosis and regurgitation, atrial fibrillation, and splitting of fundamental heart sounds leading to additional lub/dub sounds in the systole or diastole interval of a cardiac cycle. The localisation of cardiac sounds of interest was performed using an adaptive wavelet-based filtering in combination with the Shannon energy envelope and prior information of fundamental heart sounds. This is a prerequisite step for the feature extraction and subsequent classification of recordings, leading to a more precise diagnosis. Localised segments of S_1 and S_2 sounds, and artifacts, were used to extract a set of perceptual and statistical features using wavelet transform, homomorphic filtering, Hilbert transform and mel-scale filtering, which were then fed to train an ensemble classifier to interpret S_1 and S_2 sounds. Once sound peaks of interest were identified, features extracted from these peaks, together with the features used for the identification of S_1 and S_2 sounds, were used to develop an algorithm

to classify recorded signals. Overall, 99 features were extracted and statistically analysed using neighborhood component analysis (NCA) to identify the features which showed the greatest ability in classifying recordings. Selected features were then fed to train an ensemble classifier to classify abnormal recordings, and hyperparameters were optimized to evaluate the performance of the trained classifier. Thus, a machine learning-based approach for the automatic identification and classification of S_1 and S_2, and normal and abnormal recordings, in real-world noisy recordings using a novel feature set is presented. The validity of the proposed algorithm was tested using acoustic signals recorded in real-world, non-controlled environments at four auscultation sites (aortic valve, tricuspid valve, mitral valve, and pulmonary valve), from the subjects with and without cardiac diseases; together with recordings from the three large public databases. The performance metrics of the methodology in relation to classification accuracy (C_A), sensitivity (S_E), precision (P_+), and F_1 score, were evaluated.

Results: This thesis proposes four different algorithms to automatically classify fundamental heart sounds – S_1 and S_2; normal fundamental sounds and abnormal additional lub/dub sounds recordings; normal and abnormal recordings; and recordings with heart valve disorders, namely the mitral stenosis (MS), mitral regurgitation (MR), mitral valve prolapse (MVP), aortic stenosis (AS) and murmurs, using cardiac acoustic signals. The results obtained from these algorithms were as follows:

- The algorithm to classify S_1 and S_2 sounds achieved an average S_E of 91.59% and 89.78%, and F_1 score of 90.65% and 89.42%, in classifying S_1 and S_2, respectively. 87 features were extracted and statistically studied to identify the top 14 features which showed the best capabilities in classifying S_1 and S_2, and artifacts. The analysis showed that the most relevant features were those extracted using Maximum Overlap Discrete Wavelet Transform (MODWT) and Hilbert transform.
- The algorithm to classify normal fundamental heart sounds and abnormal additional lub/dub sounds in the systole or diastole intervals of a cardiac cycle, achieved an average S_E of 89.15%, P_+ of 89.71%, F_1 of 89.41%, and C_A of 95.11% using the test dataset from the PASCAL database. The top 10 features that achieved the highest weights in classifying these recordings were also identified.
- Normal and abnormal classification of recordings using the proposed algorithm achieved a mean C_A of 94.172%, and S_E of 92.38%, in classifying recordings from the different databases. Among the top 10 acoustic features identified, the deterministic energy of the sound peaks of interest and the instantaneous frequency extracted using the Hilbert Huang-transform, achieved the highest weights.
- The machine learning-based approach proposed to classify recordings of heart valve disorders (AS, MS, MR, and MVP) achieved an average C_A of 98.26% and S_E of

95.83%. 99 acoustic features were extracted and their abilities to differentiate these abnormalities were examined using weights obtained from the neighborhood component analysis (NCA). The top 10 features which showed the greatest abilities in classifying these abnormalities using recordings from the different databases were also identified.

The achieved results demonstrate the ability of the algorithms to automatically identify and classify cardiac sounds. This work provides the basis for measurements of many useful clinical attributes of cardiac acoustic signals and can potentially help in monitoring the overall cardiac health for longer duration. The work presented in this thesis is the first-of-its-kind to validate the results using both, normal and pathological cardiac acoustic signals, recorded for a long continuous duration of 5 minutes at four different auscultation sites in non-controlled real-world conditions.

Contents

Abstract **2**

List of Figures **9**

List of Tables **13**

List of Appendices **15**

Acknowledgements **16**

Declaration of Originality **18**

Copyright Declaration **19**

Terms and Abbreviations **20**

Thesis Structure **26**

1 An Introduction to Algorithms for Automatic Analysis and Classification of Heart Sounds 31

1.1 Overview 31

1.2 Pathophysiology of Heart Sounds Signals 34

1.2.1 Fundamental heart sounds 35

1.2.2 Abnormal heart sounds 35

1.3 Current State-of-the-Art Algorithms 36

1.3.1 Heart Sounds Databases 37

1.3.2 Segmentation Approaches 41

1.3.2.1 Envelope-based methods 42

1.3.2.2 ECG and/or carotid pulse reference-based methods 44

1.3.2.3 Probabilistic models for the segmentation 44

1.3.2.4 Feature-based methods 46

1.3.2.5 Time-frequency/Wavelet analysis-based methods 47

1.3.2.6 Learning-based methods 48

1.3.3 Feature Extraction 49
1.3.4 Classification of Cardiac Acoustic Signals 52
1.3.4.1 Support Vector Machine 52
1.3.4.2 Hidden Markov Models 54
1.3.4.3 k-Nearest Neighbor (k-NN) 55
1.3.4.4 Neural Networks 56
1.4 State-of-the-Art Cardiac System Monitors 59
1.5 Discussion 62

2 Materials and Methods 68
2.1 Heart Sounds Datasets 68
2.2 Feature Extraction & Selection 73
2.2.1 Statistical time and frequency domain features 73
2.2.2 Deterministic and non-deterministic energies 75
2.2.3 Multiresolution analysis using a non-decimated transform 75
2.2.4 Homomorphic envelogram 77
2.2.5 Hilbert-Huang transform-based features 77
2.2.6 Perceptual features 80
2.2.7 Feature selection and reduction 81
2.3 Classifier Training and Hyperparameter Optimization 81
2.3.1 Classifier training 82
2.3.2 Hyperparameter optimization 82

3 Automatic Classification of S_1 and S_2 Sounds from Cardiac Auscultatory Recordings 83
3.1 Introduction 83
3.2 Methods 84
3.2.1 Pre-processing of recorded signals 84
3.2.2 Localization of sound peaks of interest 89
3.2.3 Annotation of S_1 and S_2 90
3.3 Feature Extraction & Selection 93
3.4 Classifier Training and Testing 96
3.4.1 Training and test datasets 97
3.4.2 Classifier training 98
3.4.3 Hyperparameter optimization 99
3.5 Experimental Results 99
3.6 Discussion 105

4 An Approach for Automatic Identification of Fundamental and Additional Sounds 108
4.1 Introduction 108
4.2 An Approach to Analyse Split of S_1 and S_2 Sounds 109
4.3 Segmentation of Heart Sounds Signals 112

4.3.1 Pre-processing 112
4.3.2 Identification of sound peaks of interest 113
4.4 Feature Extraction & Selection 116
4.4.1 Feature extraction 116
4.4.2 Feature selection and reduction 118
4.5 Classifier Training and Testing 122
4.5.1 Training and test datasets 122
4.5.2 Model training and parameter optimization 123
4.6 Results 124
4.6.1 Evaluation metrics 124
4.6.2 Results 124
4.7 Discussion 125

5 Classification of Normal and Abnormal Heart Sounds from Recorded Signals 128
5.1 Introduction 128
5.2 Methods 128
5.3 Pre-processing and Segmentation 129
5.4 Feature Extraction & Selection 131
5.5 Classification of Cardiac Acoustic Signals 133
5.5.1 Training and test datasets 134
5.5.2 Model training 135
5.5.3 Hyperparameter optimization 135
5.6 Experimental Results 136
5.7 Discussion 138

6 Non-invasive Approach for Abnormality Detection using Heart Sounds Recordings 143
6.1 Introduction 143
6.2 Methods 143
6.3 Feature Extraction & Selection 146
6.3.1 Feature extraction 146
6.3.2 Feature selection and reduction 148
6.4 Classification of Pathological Recordings 152
6.4.1 Training and test datasets 152
6.4.2 Model training 154
6.4.3 Hyperparameter optimization 155
6.5 Experimental Results 156
6.5.1 Evaluation metrics 156
6.5.2 Results 156
6.6 Discussion 160

7 Conclusions 163

7.1 Contributions 163
7.2 Further Work 166

List of Publications 167

Appendices 168

Bibliography 192

List of Figures

Figure 1.1. Breakdown of leading causes of deaths worldwide. 32

Figure 1.2. Health care spending as a percentage of GDP [284]. 32

Figure 1.3. Block diagram of the system for heart sounds analysis and classification. 34

Figure 2.1. Location of heart valves and sites of auscultation [285]. 69

Figure 2.2. An electronic stethoscope from the ThinkLabs used in this study to record cardiac acoustic signals. 70

Figure 2.3. A 12-lead ECG system from Contec Medical Systems Co. Ltd., used in this study to record ECG signals. 70

Figure 2.4. Various statistical features extracted from the localized sound peaks of interest in a recorded heart sounds signal from normal subject, from the PASCAL database. 74

Figure 2.5. Flowchart for the non-deterministic energy calculation. 76

Figure 2.6. Multiresolution analysis of a recorded acoustic signal using the maximum overlap discrete wavelet transform (MODWT). Only meaningful multiresolution analysis (level 5 to 11) of the MODWT matrix coefficients are shown in (I)-(VII) for a recorded signal shown in (VIII). 78

Figure 2.7. Different envelopes extracted for feature extraction of S_1 and S_2 sounds. 79

Figure 2.8. (a) Extracted 4 out of 10 intrinsic mode functions (IMF) of a heart sounds signal using EMD. (b) Intrinsic frequencies of the different IMFs. (c) Instantaneous energy (IE) of the different IMFs. (d) IE of IMF 2 and 5 to demonstrate the variation of the energy content with respect to the IMFs. 80

Figure 2.9. MFCCs extraction from heart sounds signals using a mel-filter bank. 81

Figure 3.1. Fundamental heart sounds (S_1 and S_2) in a recording from healthy subject. 84

Figure 3.2. Flow-chart of the overall system to classify S_1 and S_2 sounds using features extracted from labelled segments of S_1 and S_2. 86

Figure 3.3. (I) Frequency response of a pre-processed recorded acoustic signal. (II) Time-frequency analysis of a recorded acoustic signal obtained using STFT. (III) Pre-processed and filtered recorded heart sounds. 87

Figure 3.4. The different steps used in the adaptive wavelet-based filtering and the localization of sound peaks of interest in recordings. (In this work n=5). 88

Figure 3.5. (I), (II) and (III) are the reconstructed recorded signals using adaptive sublevel filtering for different decomposition levels (DL) 5, 6 and 7, respectively. (IV) is the original recorded acoustic signal. 90

Figure 3.6. Localization of sound peaks of interest. (a) recorded heart sounds signal. (b) S_1 and S_2 sounds after discarding extra peaks based on priori information. 91

Figure 3.7. (I)-(IV) The various steps used in the identification of R-peaks in recorded ECG signal using the Pan-Tompkins algorithm, used as reference to mark S_1. (V) Identified T-waves in ECG signals used as reference to mark S_2. 92

Figure 3.8. S_1, systole, S_2, and diastole intervals in a filtered cardiac acoustic signal labelled using R-peaks and end-T-waves locations identified in an ECG recording. 93

Figure 3.9. Labelled S_1 and S_2 sounds in a recorded signal from the PASCAL database. 93

Figure 3.10. S_1 sound segments extracted from the signal shown in Figure 3.9. 94

Figure 3.11. S_2 sound segments extracted from the signal shown in Figure 3.9. 94

Figure 3.12. Normalized feature weights of the selected features obtained using the NCA based feature selection approach. Feature weights were evaluated for both the PASCAL database and the RHSS database. 96

Figure 3.13. Boxplot of the top 14 features selected to differentiate S_1 and S_2 sounds identified using NCA based feature selection approach. 97

Figure 3.14. Key features of S_1 and S_2 sounds identified using NCA based feature selection approach. Plotted features are energy of different levels for MODWT (MODWTEnergy), energy computed using MRA (MRAEnergy), energy of IMFs using HHT (HHTenergy), instantaneous energy (InstEnergy) and IE using k-means clustering (IEcluster). 98

Figure 3.15. (a) Feature extraction from raw signals and (b) steps used for training and cross-validation of the classifier to extract hyperparameters. 99

Figure 3.16. Key hyperparameter optimization using misclassification rate (MCR). MNS represents the maximum number of splits in the tree. 100

Figure 3.17. $P(S_1)$ and $P(S_2)$ for all the observations in: (a) RHSS training set (86229 observations), (b) RHSS test set (21557 observations), (c) PASCAL training set (3032 observations) and (d) PASCAL test set (758 observations). Frequency represents the number of observations having a given probability $P(S_1)$ or $P(S_2)$. 104

Figure 3.18. Probabilities ($P(S_1)$ and $(1 - P(S_1))$) for each observation and the intervals used to estimate the classifier's performance in: (a) PASCAL test set (758 observations) and (b) PASCAL training set (3032 observations)..... 105

Figure 4.1. S_1 and S_2 sounds aligned to extract features based on the consistency of S_1 and S_2 sounds. (a) and (d) show 22 S_1 and S_2 sounds from the database, aligned and plotted on top of one another to extract S_1 and S_2 sounds as in (b) and (e), respectively. (c) and (f) show the extracted features of S_1 and S_2 sounds, respectively. 111

Figure 4.2. (a) Heart sounds signal from the PASCAL database. (b) Frequency spectrum of the signal using the Fourier analysis. This was used to estimate the cut-off frequency of the filter used to remove noise. (C) Spectrogram of the recorded signal. 113

Figure 4.3. (a) Shannon energy envelope of the signal shown in (c) (from the PASCAL database). (b) shows the filtered wavelet-based coefficients to extract envelope..... 114

Figure 4.4. Localization of sound peaks of interest and S_1 and S_2 sounds identification using the reference locations. 115

Figure 4.5. Identification of S_1 and S_2 sounds in the heart sounds signals with additional lub/dub sounds obtained from the database. (c) and (b) show the recorded signal and the identified sound lobes of interest, respectively. (a) shows the potential S_1 and S_2 sounds after discarding the identified abnormal sounds. 116

Figure 4.6. Identification of S_1 and S_2 sounds in a heart sounds signals recorded in an uncontrolled environment using an electronic stethoscope. (c) and (b) show the recorded signal and the identified sound lobes of interest, respectively. (a) shows the potential S_1 and S_2 sounds after discarding the identified abnormal sounds. 117

Figure 4.7. Envelopes extracted from a recorded signal. 118

Figure 4.8. The different intrinsic mode functions extracted from a heart sounds recording using EMD. IMF1 contains high-frequency components, hence, only IMFF2, IMF3, IMF4 and IMF5 were used to obtain the cluster of the instantaneous frequency and instantaneous energy of the signal. (Y-axis is in the normalised unit.) 119

Figure 4.9. (a) The instantaneous frequency (first figure), and (b) instantaneous energy (second figure), for the different IMFs selected using a signal, for feature extraction. 120

Figure 5.1. Block diagram of the overall heart sounds classification system. 129

Figure 5.2. Steps used for training and cross-validation of the classifier to extract optimized hyperparameters. 135

Figure 6.1. Different types of recordings used in this study. Murmur recording from the PASCAL database; and aortic stenosis (AS), mitral regurgitation (MR), mitral valve prolapse (MVP), mitral stenosis (MS), and normal recordings from a public database. 145

Figure 6.2. Block diagram of the overall heart valve disorders classification system. 146

Figure 6.3. (a) The instantaneous frequency, and (b) instantaneous energy for IMF 2, IMF 3, IMF 4 and IMF 5, selected for feature extraction. Murmur recording obtained from the PASCAL database is used here..... 148

Figure 6.4. Extracted 9 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with aortic stenosis (AS)..... 149

Figure 6.5. Extracted 10 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral valve prolapse (MVP). 149

Figure 6.6. Extracted 8 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral regurgitation (MR). 150

Figure 6.7. Extracted 10 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral stenosis (MS). 150

List of Tables

Table 1.1: Various pathological conditions of the cardiovascular system. 37

Table 3.1: Details of heart sounds signals used in this work. 85

Table 3.2: Various entropy estimation for the signal shown in Figure 3.5. 88

Table 3.3: Various features extracted and their description (F_1-F_{73}). 95

Table 3.4: Three-class confusion matrix. 101

Table 3.5: Confusion matrices of the trained classifier for different datasets. 102

Table 3.6: Confusion matrices of the trained classifier for three-classes from different datasets. 103

Table 3.7: Performance of the trained classifier using 10-fold cross validation. 104

Table 3.8: Performance comparison with approaches for S_1 and S_2 identification. 106

Table 4.1: Various features extracted and their description (F_1- F_{36}). 121

Table 4.2: Various features extracted and their description (F_{37}-F_{99}) 122

Table 4.3: Cardiac acoustic recordings used in this study. 123

Table 4.4: Three-class confusion matrix. 124

Table 4.5: Confusion matrices of the trained classifier for (a) training datasets and (b) test dataset. 125

Table 4.6: Performance of the classifier using 10-fold cross-validation. 125

Table 4.7: Performance comparison with approaches for normal, extra heart sound and artifacts classification. 127

Table 5.1: Cardiac acoustic recordings used in this work. 130

Table 5.2: The top 10 features selected from the different databases. 132

Table 5.3: Intra-class variations of the top 10 features selected using different databases. 133

Table 5.4: Confusion matrices of the trained classifier for different datasets. 137

Table 5.5: Performance of the classifier using 10-fold cross-validation. 138

Table 5.6: Classification performance of the classifier model after excluding observations within ±1% of the mean probability. 138

Table 5.7: Classification performance of the classifier model after excluding observations within ±5% of the mean probability. 139
Table 5.8: Performance comparison with approaches for normal and abnormal classification of recordings..... 141
Table 6.1: Cardiac acoustic recordings used in this work..... 144
Table 6.2: The top 10 features selected from the different databases..... 151
Table 6.3: Intra-class variations of features selected using different databases..... 153
Table 6.4: Recordings used to create training and test datasets..... 155
Table 6.5: Five-class confusion matrix..... 157
Table 6.6: Confusion matrices of the trained classifier for different datasets..... 158
Table 6.7: Performance of the trained classifier using 10-fold cross-validation. 159
Table 6.8: Performance comparison with approaches for pathological classification using the databases used in this study 161

List of Appendices

Appendix A: Pathophysiology of Normal and Abnormal Heart Sounds 168

Appendix B: Databases used for Validation of Algorithms for Heart Sounds Analysis 170

Appendix C: Summary of Heart Sounds Segmentation Methods and their Performance Comparison 174

Appendix D: Heart Sounds Feature Extraction Approaches for Analysis 177

Appendix E: Summary of Heart Sounds Classifiers and their Performance Comparison 180

Appendix F: List of Features Extracted in this Work and their Description 186

Appendix G: Summary of Clinical Study Protocol for the Collection of Signals 189

Acknowledgements

It gives me immense pleasure to express my deep sense of gratitude and indebtedness to my supervisor Prof. Esther Rodriguez-Villegas for providing me the opportunity to pursue my PhD research. I am blessed to have her as my supervisor, with amazing support, motivation and all the resources that she provided me throughout my PhD. She is a fantastic supervisor, innovator and always ready to guide me whenever I need her help, at both professional and personal level. I am thankful for her invaluable guidance, encouragement, and patient reviews and feedbacks, throughout my PhD. This is her continuous support that has made me to complete this thesis, without her help and guidance, this dissertation would have been impossible.

My special thanks to Anas and Piyush for their continuous support during my PhD. Anas has been very helpful and provided constructive feedbacks during all the discussion related to my PhD project. It has been a wonderful experience working with Anas, especially while publishing my research work. He provided his valuable inputs throughout the duration. I am also thankful for all the memories that I shared with Piyush during this time.

The facilities and knowledge of the Wearable Technologies Lab were very helpful for the successful completion of my PhD project. My time as a PhD student would not have been the same without my lab members. I would like to thank all the members (Irene, Renard, Zhou, Michal, Xuen, Sukhi, Stuart, Saam, and Majd) of the lab for all the good times and for being there in times of need. It has been a great pleasure working with you all, and I am thankful for all the moments that made my experience enjoyable and amazing.

I would also like to thank Wiesia for her incredible administrative support and personal suggestions that she provided. From the very first day of my life at Imperial, she has been very helpful and supportive. Also, thanks to Zohiab Akthar for his guidance during the lab works.

My sincere thanks to Dr. Ajeet Kumar, Dr. Abhyuday Kumar, and Dr. Shaheen Ahmad, from All India Institute of Medical Sciences (AIIMS), Patna, India, for providing cardiac acoustic recordings and ECG signals for this study. The signals were useful in validating algorithms developed in this work.

I would also like to extend my thanks to the Imperial College London, U.K., for supporting this research work through the President's PhD Scholarship, and in part, by the European Research Council (ERC). Also, I would like to extend my sincere gratitude to the examiners of

this thesis, Dr. Kia Nazarpour, University of Edinburgh, and Dr. Ravi Vaidyanathan, Imperial College London, for reviewing and providing valuable inputs in improving this thesis.

I would also like to thank all my best friends for their time-to-time encouragement and appreciation. My apologies and heartfelt gratitude to all those who have assisted me yet have not being acknowledged by name.

I am grateful for the support and encouragement from my parents, and all my family members. My deepest gratitude goes to the sacrifices they went through to see me complete my PhD research. Finally, I would like to thank God for giving me power to believe in myself and pursue my dreams.

Declaration of Originality

I hereby confirm that the work presented in this thesis is my own, and appropriate references have been used to denote the work of others.

Amit Krishna Dwivedi
September 2020

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Terms and Abbreviations

AD: Aortic disease

AEC: Aortic ejection click

AF: Atrial fibrillation

AHV: Aortic heart valve

AI: Aortic insufficiency

AIS: Artificial immune system

AN: Abnormal heart sounds

ANFIS: Adaptive-neuro fuzzy inference system

AR: Aortic regurgitation

AS: Aortic stenosis

ASD: Atrial septal defect

ASSA: Adaptive singular spectral analysis

AWN: Additive white noise

BAV: Bicuspid aortic valve

BMV: Bioprosthetic mitral valves

BP-ANN: Back-propagation artificial neural network

BPSO: Binary particle swarm optimization

C_A: Classification accuracy

CAD: Coronary artery disease

CART: Classification and regression trees

CFS: Correlation-based feature selection

CHD: Congenital heart disease

CHF: Chronic heart failure

CM: Continuous murmurs

CNN: Convolutional neural network

CQA: Cycle quality assessment

CSCW: Cardiac sound characteristic waveforms

CTLDB: Cumulant-based trapezoidal local discriminant basis

CWD: Choi–Williams distribution

CWT: Continuous wavelet transform

DA: Discriminant analysis

DC: Dilated cardiomyopathy

DHMM: Duration-dependent hidden Markov model

DHS: Doppler heart sounds

DM: Diastolic murmurs

DNN: Deep neural network

DT: Decision trees

DWD: Discrete wavelet decomposition

DWT: Discrete wavelet transform

EAS: Early aortic stenosis

EBP: Error back-propagation

EDC: Energy distribution coefficients

EEMD: Ensemble Empirical mode decomposition

EFSE: Energy fraction and sample entropy

EHS: Extra heart sounds

EKF: Extended Kalman filter

ELM: Extreme learning machine

ERBF: Exponential radial basis function

ETD: Exponential *T*-distribution

FCM: Fuzzy *C*-means

FD: Frequency domain

FFM: Frequency feature matrix

FFNN: Feed-forward neural network

FHS: Fundamental heart sounds

FM: Functional murmurs

FWT: Fast wavelet transform

GRKF: Gaussian radial basis kernel function

GSVM: Genetic-SVM

GTSVM: Growing time support vector machine

HE: Homomorphic envelogram

HF: Homomorphic filtering

HMM: Hidden Markov model

HOC: Higher order cumulants

HOCM: Hypertrophic obstructive cardiomyopathy

HSA: Heart sounds analysis

HT: Hilbert transform

HTD: Hyperbolic *T*-distribution

ICS: Intercostal space

IM: Innocent murmurs

IMF: Intrinsic mode functions

ISOM: Incremental self-organizing map

LAS: Late aortic stenosis

LBBB: Left bundle branch block

LICS: Left intercostal space

LKF: Linear kernel function

LLE: Largest Lyapunov exponent

LLSB: Left lower sternal border

LMSB: Left middle sternal border

LPC: Linear predictive coefficient

LPCC: Linear predictive coding coefficients

LUSB: Left upper sternal border

LVI: Left ventricular impairment

MCS: Modified Cuckoo search

MESE: Maximum entropy spectra estimation

MFCC: Mel-frequency cepstral coefficient

MF-DFA: Multifractal detrended fluctuation analysis

MFSC: Mel-frequency spectral coefficients

MHMM: Modified HMM

MHV: Mitral heart valve

MI: Mitral insufficiency

MLBS: Multi-level bias selection

MLP-BP: Multi-layer perceptron back-propagation

MMP: Multi-variate matching pursuit

MP: Matching pursuit

MPC: Miscellaneous pathological conditions

MODWT: Maximum overlap discrete wavelet transform

MR: Mitral regurgitation

MS: Mitral stenosis

MVP: Mitral valve prolapse

MWKF: Morlet wavelet kernel function

NAR-PSD: Normalized auto-regressive power spectral density

NPV: Negative predictive value

NRMR: Non-rheumatic mitral regurgitation

OMS-WPD: Optimum multi-scale wavelet packet decomposition

P_+: Precision

PDA: Patent ductus arteriosus

PI: Pulmonary insufficiency

PKF: Polynomial kernel function

PLSR: Partial least squares regression

PLSR: Partial least squares regression

PM: Pathological murmurs

PMD: Papillary muscle dysfunction

PMV: Prolapsed mitral valve

PNN: Probabilistic neural network

PS: Pulmonary stenosis

PSM: Parametric spectral modelling

QRD: *QR*-decomposition

RBF: Radial basis function

RBF-NN: Radial basis function neural network

RHD: Rheumatic heart disease

RME: Rényi marginal entropy

RQA: Recurrence quantification analysis

RUSB: Right upper sternal border

RWNN: Radial wavelet neural network

S_1: First heart sound (or lub sound)

S_2: Second heart sound (or dub sound)

SAMDF: Sum of average magnitude difference function

SCAS: Severe calciphic aortic stenosis

SD: Standard deviation

S_E: Sensitivity

SEC: Systolic ejection click

SFFS: Sequential forward floating selection

SG: Summation gallop

SHM: Systolic heart murmurs

SKF: Sigmoid kernel function

SM: Systolic murmurs

SPWVD: Smoothed pseudo WVD

SR: Segmentation rate

SS-PLSR: Scaled spectrogram and partial least squares regression

SS-TD: Scaled spectrogram and tensor decomposition

STMHT: Short-time modified Hilbert transform

SVD: Singular value decomposition

SVM-DM: SVM-diffusion maps

SVM-MCS: SVM-modified Cuckoo search

TD: Time-domain

TF: Time-frequency

TFD: Time-frequency domain

TGNN: Time-growing neural network

TI: Tricuspid insufficiency

TOF: Tetralogy of Fallot

TQWT: Tunable-Q wavelet transform

TRF: Time-frequency representation

TS: Tricuspid stenosis

TVAR: Time-varying auto-regression

VHD: Valvular heart disease

VIM: Viola integral method

VSD: Ventricular septal defect

WD: Wavelet decomposition

WFSS: Wrapper feature selection scheme

WHSMD: Washington heart sounds and murmurs database

WPD: Wavelet packet decomposition

WPNN: Wavelet packet neural network

WVD: Wigner-Ville distribution

Thesis Structure

This thesis is organised into six main chapters with five additional appendices. The details of each chapter are summarised below.

Chapter 1: An Introduction to Algorithms for Automatic Analysis and Classification of Heart Sounds

In this chapter, an introduction to the cardiovascular diseases is briefly discussed followed by their impact on society. The current clinical practice in the diagnosis of these diseases, and the limitation and drawbacks of this are discussed. This discussion highlights the importance of automatic detection of cardiovascular diseases in their early stage using long term monitoring. It is justified that automatic analysis and classification of heart sounds signals can provide critical information about these diseases and contribute to their early diagnosis. A comprehensive literature review and critical analysis of all the existing state-of-the-art approaches for heart sounds segmentation, feature extraction and subsequent classification, is presented to explain the challenges, limitation and drawbacks of the methods for the automatic analysis and classification of heart sounds. This also contains a brief introduction to the available heart sounds databases and methods used to acquire these signals. A discussion on existing relevant commercial devices for monitoring the cardiovascular system is also presented. From the review, it is clear that, although, a lot of research has been done in the field of automated analysis, there is still some work to be done to develop robust methods for the identification and classification of various events in the cardiac cycle so that this could be effectively used to improve the diagnosis and management of cardiovascular diseases in combination with wearable mobile technologies. Based on the findings from the review, potential solutions and the challenges are discussed in detail.

Chapter 2: Materials and Methods

This chapter provides a detailed description of the datasets and methods used to facilitate and evaluate the performance of the approaches presented in Chapters 3, 4, 5 and 6. The four different studies in chapters 3, 4, 5 and 6 utilised the features described in this chapter to develop machine-learning based approaches with the aim to address the challenges in the automatic classification of cardiac recordings from different conditions.

Chapter 3: Automatic Classification of S_1 and S_2 Sounds from Cardiac Auscultatory Recordings

The importance of segmentation of heart sounds signals leading to the identification of fundamental heart sounds – first heart sound (S_1) and second heart sounds (S_2)- is discussed in this chapter. It is found from the review presented in Chapter 1 that the accurate localization of heart sounds assists the classification of pathological events in cardiac acoustic recordings, leading to a more precise diagnosis. It is a prerequisite to identify systole and diastole intervals in a cardiac cycle to assist the feature extraction for the classification. However, there are challenges involved in the localization and segmentation of heart sounds in recorded signals, especially when signals are collected in real-world scenarios where they are often contaminated with noise and artifacts. This chapter presents an algorithm to classify S_1 and S_2 sounds, and artifacts, in real-noisy cardiac auscultatory recordings using a novel feature vector. An adaptive wavelet-based filtering approach was used to isolate components of S_1 and S_2 sounds. Subsequently, heart sounds of interest are localized using the Shannon energy envelope in combination with prior information of fundamental heart sounds. Features were extracted from S_1 and S_2 sounds, as well as artifacts, and statistically analysed to investigate the top features that demonstrated the greatest ability to differentiate S_1 and S_2 sounds. Finally, selected features were used to train a classifier model to identify peaks of S_1 and S_2 sounds, and artifacts. For the first time, the segmentation algorithm presented was tested using acoustic signals recorded for 5 minutes from four auscultation sites (aortic valve, tricuspid valve, mitral valve, and pulmonary valve), from both subjects with, and without, cardiac disease; together with signals from a public database.

Chapter 4: An Approach for Automatic Identification of Fundamental and Additional Sounds

Chapter 4 presented an approach for the automatic identification of recordings with normal and abnormal additional sounds; where the latter included both splitting of fundamental, S_1 and S_2, sounds and additional lub/dub sounds in the diastolic and systolic periods. A discussion on abnormal additional sounds due to the sub-components of S_1 and S_2 sounds is presented. The sub-components of S_1 are M_1 (due to the closure of the mitral valve) and T_1 (due to the closure of the tricuspid valve), while A_2 (due to closure of aortic valve) and P_2 (due to the closure of pulmonary valve) are of S_2. The detection of abnormal fundamental heart sounds is based on the feature extraction from those segments that were previously identified as normal fundamental heart sounds. The steps described in Chapter 3 were utilized to localize sound

peaks of interest, which were subsequently used to extract features. The methods used to extract the features are discussed in detail. Finally, the extracted features were used to train an ensemble classifier to interpret the class of recorded signals. The performance metrics such as classification accuracy, sensitivity, and F_1 score, were evaluated and discussed in detail for proper comparisons to the recent works.

Chapter 5: Classification of Normal and Abnormal Heart Sounds from Recorded Signals

In this chapter, automatic classification of normal and abnormal recordings is presented. Heart sounds of interest were localized using the steps described in Chapter 3. 99 features (described in Chapter 2) were extracted, and corresponding weights were evaluated using neighborhood component analysis (NCA) to identify the best performing features. An ensemble classifier model was trained to classify recordings using the selected features. To demonstrate the robustness of the algorithm in long term continuous monitoring, the trained classifier model was tested using data collected in this study for a duration of 5 minutes from different auscultation sites; together with the data obtained from the two large public databases. Finally, evaluation metrics were computed and compared with recent works to demonstrate the improvement achieved in the performance parameters of the approach described in this thesis.

Chapter 6: Non-invasive Approach for Abnormality Detection using Heart Sounds

Chapter 6 presents an approach to automatically detect cardiac abnormalities, namely the mitral stenosis (MS), mitral regurgitation (MR), mitral valve prolapse (MVP), aortic stenosis (AS) and murmurs, using cardiac auscultatory recordings. To achieve this, steps described in Chapter 3 were used to localize heart sounds followed by the feature extraction for the classification. 99 acoustic features were extracted and their abilities to differentiate different cardiac abnormalities were examined. The neighborhood component analysis was used to obtain weights corresponding to each feature and to subsequently select the best features for the classification. Selected features were used to propose a machine learning-based method that involved acoustic feature extraction and subsequent classification of recordings using an ensemble classifier with optimized hyperparameters. The proposed algorithm was validated using signals obtained from two large public databases. The achieved results demonstrate the improvement in the overall accuracy of the approach proposed for automatic identification of pathological events and suggest features that can potentially help in long term monitoring of cardiac abnormalities.

Appendix A: Pathophysiology of Normal and Abnormal Heart Sounds

Characteristics of adventitious heart sounds are provided in Appendix A.

Appendix B: Databases used for Validation of Algorithms for Heart Sounds Analysis

A list of existing public databases as well as data collected by researchers in the reviewed studies is provided in Appendix B. This also contains information related to the type of sounds, number of participants, subject demographics (a breakdown of the subjects used in the study by age, gender and cardiac conditions), auscultation sites, sensor used to record signals, sampling frequency and duration of recordings. Environmental conditions such as controlled and uncontrolled conditions, during the data collection are also provided.

Appendix C: Summary of Heart Sounds Segmentation Methods and their Performance Comparison

The approach to abnormal heart sounds classification is framed as a three-step process. This involves the heart sounds segmentation, feature extraction and subsequent classification of heart sounds using the extracted features. Approaches for the automatic heart sounds segmentation are reviewed in Chapter 1. In addition to the performance achieved by the approaches reviewed in the chapter, information related to the type of signals, subject demographics, number of recordings/cardiac cycles, and the type of sounds (S_1, S_2, S_3, and S_4) localized during the segmentation process, is provided in Appendix C.

Appendix D: Heart Sounds Feature Extraction Approaches for Analysis

Various physiological and pathological characteristics of cardiac acoustic signals were evaluated to assist the classification of recordings. Details of feature extraction methods and the types of features extracted in articles reviewed in this thesis are provided in Appendix D. This also contains information related to the type of signal representation used to extract features, number of features, and approaches used to select the best features.

Appendix E: Summary of Heart Sounds Classifiers and their Performance Comparison

The final step of the automatic heart sounds analysis algorithm is to train a classifier to interpret the class of recordings, using the extracted features. The performance (in terms of sensitivity (S_E), specificity (S_P), precision (P_+) and classification accuracy (C_A)) achieved by the approaches reviewed in Chapter 1 are discussed. Additionally, information related to the type of signals, subject demographics, number of recordings/cardiac cycles, size of feature vector, type of classifier and cardiac abnormality investigated in the study are provided in Appendix E.

Appendix F: List of Features Extracted in this Work and their Description

In this thesis, 99 features were extracted and statistically analysed to investigate the top features that demonstrated the greatest ability to differentiate normal recordings from the different types of abnormal recordings. Details of all these features is provided in Appendix F.

Appendix G: Summary of Clinical Study Protocol for the Collection of Signals

The study designed to record cardiac auscultatory signals during this work is presented in Appendix G. Information related to the recording devices, selection criteria of subjects used in the study, study groups and study methods are described in detail.

1 An Introduction to Algorithms for Automatic Analysis and Classification of Heart Sounds

The research presented within this chapter is an edited version of research article previously published as:

Amit Krishna Dwivedi, Syed Anas Imtiaz, and Esther Rodriguez-Villegas, "Algorithms for Automatic Analysis and Classification of Heart Sounds–A Systematic Review", IEEE Access 7 (2018), pp. 8316-8345.

1.1 Overview

Cardiovascular diseases (CVDs) kill over 17.5 million people a year and are the leading cause of death worldwide [1]. Furthermore, this number is predicted to increase to approximately 23 million per year by 2030 [2]. Apart from the personal consequences, the high prevalence and cost of cardiovascular diseases constitute a serious social and financial burden. As an illustration, 85 million Americans suffer from cardiovascular diseases. This results in an approximate health care cost of $320 billion annually, with a projected increase to nearly $1 trillion [3]. While the estimated number of cardiac patients and healthcare costs are too high, an important thing to consider is that most cardiovascular diseases are curable and preventable. However, this requires a proper timely diagnosis and disease management. Effective intervention cannot be independent of early-stage diagnosis [4]. Consequently, there is an urgent need to improve technologies to intensively monitor and analyze physiological parameters associated with a cardiac function in a timely and cost-effective manner. With recent evolution of mobile technologies, there is a growing, justified, interest on finding ways to continuously track the cardiovascular system for long periods of time, as a potentially more effective way to both diagnose and manage cardiac conditions.

In literature, both invasive and non-invasive approaches for monitoring cardiovascular system using different sensing schemes have been reported. Approaches to diagnose cardiovascular diseases that require continuous monitoring of several vital parameters are cumbersome and expensive. The performance of these approaches is restricted due to the limited battery life, mainly because of the power-hungry algorithms and grueling circuit

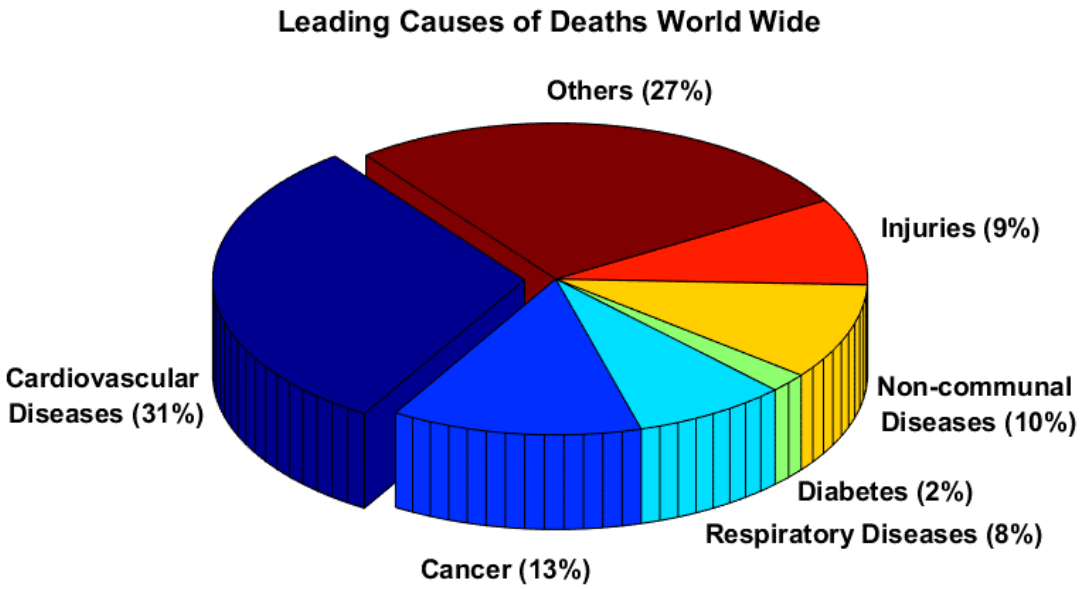


Figure 1.1. Breakdown of leading causes of deaths worldwide.

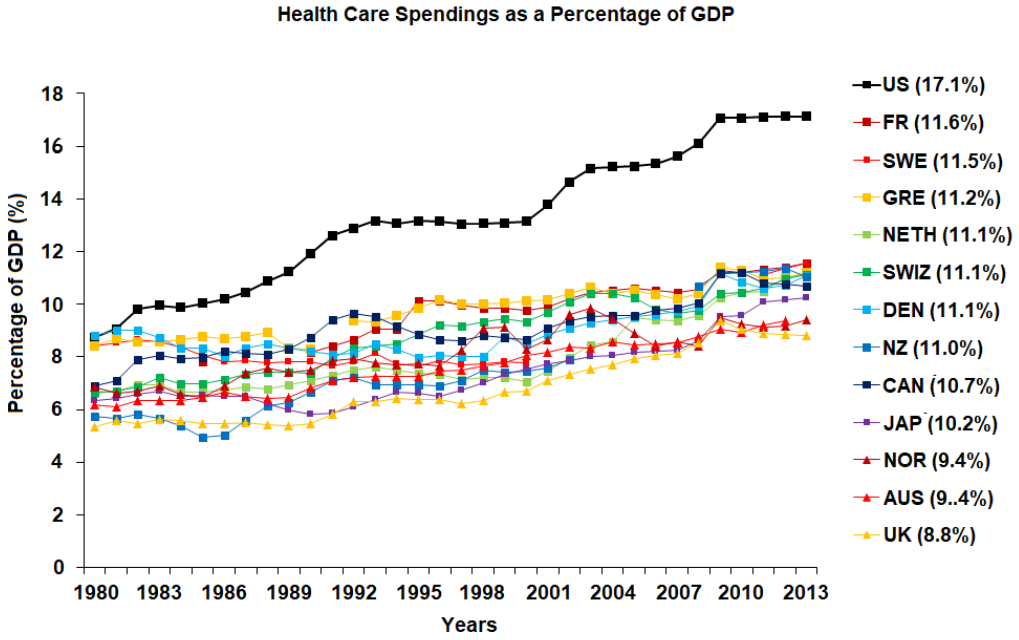


Figure 1.2. Health care spending as a percentage of GDP [284].

implementations. Also, some of these approaches are not suitable for long term continuous monitoring of cardiac signals in unsupervised environments, which would however be the optimum way of monitoring/managing some cardiac conditions. An example is atrial fibrillation, in which events do happen scattered in time and hence might not be caught in short monitoring sessions [5].

Recent advancements in computing together with the ever-shrinking size of electronic

devices have enabled the design of wearables loaded with sensors that can perform the task of long term continuous monitoring. Wearables have the potential to facilitate early-stage diagnosis of cardiovascular diseases and improve their management since they could provide continuous monitoring and enable timely intervention. Furthermore, they have the advantage of usability. Thus, wearables could allow self-health monitoring and save the time required for clinical appointments. This is why wearables have attracted a lot of attention from scientists in this field. Though, potentially, available cardiac wearables can assist in real time monitoring, it is challenging to obtain a high degree of accuracy, especially under varying environmental conditions.

The sensing modality, and hence the measured physiological signal, used by different kind of wearables varies and which one to choose depends on a number of trade-offs that need to be made between the particular clinical application, usability aspects and accuracy, amongst others. In the case of wearables for cardiac applications, one of the physiological signals that can potentially provide a lot of information is the sounds generated by the heart. Cardiac acoustic signals caused by the mechanical vibrations of the heart or turbulent blood flow, can be recorded using an electronic stethoscope or a microphone from the surface of the chest. These acoustic signals are indicators of a wide number of cardiovascular abnormalities and suitable for long term monitoring since they can be recorded non-invasively for longer duration. Thus, wearable monitors that can automatically analyse and interpret the recorded signal will be of great importance. These monitoring devices can assist physicians in early-stage diagnosis and continuous monitoring of cardiovascular diseases in an accurate and automated manner.

Heart sounds auscultation is a simple, convenient, cheap and non-invasive approach that has been used for over a century by physicians to diagnose cardiac diseases. However, manual signal interpretation is not easy, mostly in noisy environments, demanding high levels of skills. More recently human-only stethoscope-based interpretation is being complemented by computer-aided heart sounds analysis. Thus, cardiac acoustic signals can be recorded during auscultation and information can be presented graphically in what is known as a phonocardiogram (PCG). Recorded heart sounds signals can be subsequently visualized and stored for further interpretation. Computer-aided auscultation has as a potential advantage that the interpretation of heart sounds is not as subjectively dependent on factors such as ear sensitivity, skills, and the experience of the individual physicians [6], [7].

A wearable automated system for analysis of heart sounds could potentially be used for early cost-effective screening of cardiovascular diseases, as well as to manage the progression of the condition. However, in order for this to practically happen, algorithms are required that can shift the signal interpretation load from the clinician to the technology since otherwise the amount of information generated would be unmanageable in practice. This is a reason why

automated analysis and interpretation of the heart sounds is a prolific area of research, with an also rapidly increasing interest. Though computerized analysis of heart sounds has been the focus of increasing number of studies recently, a consistent approach to develop these methods have not been established.

Though, potentially, heart sounds analysis and classification can assist in real time monitoring, it is challenging to obtain a high degree of accuracy, especially under the varying environmental conditions. Hence, despite significant advances in the automated heart sounds analysis and classification, the clinical reliability and diagnostic accuracy of algorithms cannot be extrapolated for real clinical applications. In some cases, the algorithms developed for the automated signal interpretation have not been validated with a large datasets of heart sounds signals obtained from different subjects, age groups, auscultation sites and body positions.

A block diagram of the system for the automated heart sounds analysis consists of segmentation, feature extraction and classification as shown in Figure 1.3. This chapter presents a brief discussion on the pathophysiology of heart sounds signals followed by a comprehensive review of the steps used in the automated analysis and classification of heart sounds signals recorded under different environmental conditions.

1.2 Pathophysiology of Heart Sounds Signals

The electrical activity of the cardiovascular system causes atrial and ventricular contractions that assist in blood circulation between the chambers of the heart and around the body. Mechanical interactions between the blood flow and the different valves that operate to regulate the circulation of blood, contribute to the rhythmic heart sounds. Heart sounds are audible on

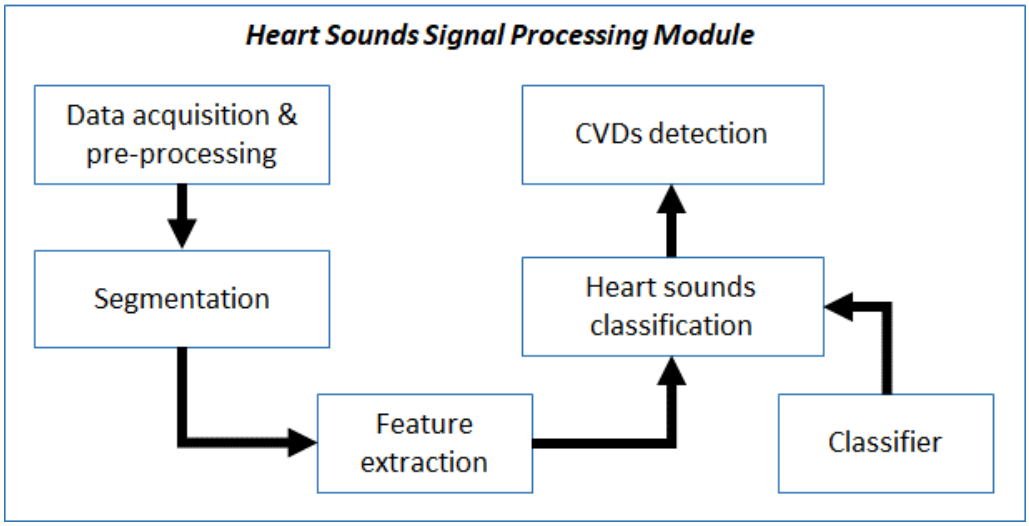


Figure 1.3. Block diagram of the system for heart sounds analysis and classification.

the chest wall and can be captured using an electronic microphone/stethoscope from different auscultation areas/sites associated with the valve locations [8], [9]. Abnormal heart sounds may relate to the structural malfunctioning and pathological conditions of the cardiovascular system and are considered vital clues to diagnose various cardiovascular diseases. Heart sounds can also be graphically represented as a phonocardiogram, in which pathological signs are used as diagnostic features. However, the correct interpretation of phonocardiograms is challenging because of the overlapping of normal and abnormal heart sounds in the cardiac cycle. This section briefly summarizes different types of heart sounds that may be observed in a cardiac cycle. Characteristics of adventitious heart sounds are also tabulated in Appendix A.

1.2.1 Fundamental heart sounds

Mechanical actions of heart valves produce heart sounds including fundamental heart sounds (FHSs), S_1 followed by S_2. [10]. While the exact physiological origins of heart sounds are not clear, recent studies confirmed that these are generated during heart valves' closures [11], [12]. The first heart sound (S_1) is heard at the onset of the systolic phase. This sound results from the sequential closure of the atrioventricular (AV) valves comprised of mitral and tricuspid valves. These valves produce two internal components of S_1: M_1 and T_1, respectively [13]. S_1 is characterized as a lower-pitch sound than S_2 with a frequency range between 10 and 200 Hz, and a more prolonged duration. The amplitude of S_1 has a great correlation with cardiac output [14]. Normally, S_1 is heard as a single sound with internal components M_1 and T_1 are separated by a very small gap of nearly 20-30 milliseconds (ms) [15], however, during some cardiac abnormalities (such as right bundle branch block), splitting of S_1 can be observed.

The second heart sound (S_2) occurs at the beginning of the diastolic phase, being caused by the closure of the aortic and pulmonic valves. S_2 is characterized as a higher-pitch sound than S_1 with a frequency range between 20 and 250 Hz, and shorter in duration as compared to S_1. S_2 is heard as a single sound with internal components: A_2 and P_2. However, during cardiac abnormalities, S_2 may be observed as two split beats of A_2 and P_2, because of a noticeable time gap existing between the closure of A_2 and P_2. This gap may vary between 30 to 80 ms during inhalation and may reduce to 15 ms during exhalation [16], [17]. Wide and paradoxical/reverse splitting of S_2 is observed during the right bundle branch block (RBBB) and left bundle branch block (LBBB), respectively.

1.2.2 Abnormal heart sounds

During normal cardiac operation, a clear S_1-S_2 pattern with a systolic period (S_1 to S_2) and diastolic period (S_2 to S_1) is observed. However, in the case of abnormalities being present,

apart from S_1 and S_2, other sounds, such as a third heart sound (S_3), fourth heart sound (S_4), gallops, clicks, opening snaps (OS), and murmurs might occur.

Early diastolic filling of the ventricle, caused by blood rushing in from the atria, produces S_3 shortly after S_2. This is due to vibrations caused by blood going backwards and forwards between the walls of the ventricles. S_3 is noted as a benign sound in the case of young people, athletes, and during pregnancy. In other cases, however, it is considered an important indicator of reduced systolic function. Diastolic dysfunction because of a stiff ventricle gives rise to an audible S_4 happening shortly before S_1 that contributes to the late diastolic filling. The occurrence of S_4 is considered a very significant indicator of cardiac abnormalities.

Clicks and snaps are also important evidence of abnormalities related to the operation of the valves. Systolic clicks are brief and high-pitch sounds, usually noticed during the opening of the semilunar valves. These occur shortly after S_1. Opening snaps may be observed shortly after S_2, with the opening of the mitral and tricuspid valves. These abnormal sounds are indicators of mitral valve prolapse (MVP), mitral regurgitation (MR), and other pathological conditions. Gallops are sounds that resemble galloping rhythms. These signpost serious myocardial dysfunction because of noncompliance of one or both ventricles. Gallop sounds may be observed during S_3 or S_4 or both [13], [18].

Turbulence due to accelerations and de-accelerations of blood in chambers of the heart, stiffening/narrowing or incompetence of the heart valves because of regurgitation, produce mechanical vibrations that propagate to the surface and give rise to audible whooshing sounds called murmurs. Most murmurs are intra-cardiac events observed in the frequency range between 20-600 Hz. The frequency spectrum of murmurs, artifacts, fundamental heart sounds and other heart sounds present in the cardiac cycle, overlap significantly. However, murmurs are more chaotic in nature. Murmurs can be broadly classified based on their characteristics: timings (systolic murmurs, diastolic murmurs, or maybe both); shape (crescendo, decrescendo or crescendo-decrescendo) and location in the cardiac cycle (early, mid or late or continuous) [19]. They may be also classified as stenosis (such as aortic stenosis (AS)) or regurgitation (such as mitral regurgitation (MR)) murmurs. Murmurs may be innocent or else may indicate pathological conditions depending on their characteristics. These pathological conditions are also listed in Table 1.1.

1.3 Current State-of-the-Art Algorithms

This section of the chapter presents a literature review of heart sounds databases and the algorithms reported for automatic analysis and classification of heart sounds signals. The different steps: segmentation, feature extraction and classification; involved in the analysis as

well as various approaches that range from traditional thresholding methods to recent machine learning based ones, are reviewed.

1.3.1 Heart Sounds Databases

Non-availability of standardized, good-quality, thoroughly validated, and documented datasets hinder the development of algorithms for heart sounds analysis. Currently, the most extensive database of heart sounds recordings is PhysioNet [8], [20]–[22]. Other databases used for the validation of algorithms in the reviewed papers included the PASCAL database [23], the Open Michigan Heart Sound & Murmur Library (OMHSML) [24], the Cardiac Auscultation of Heart Murmurs database (eGeneralMedical) [25], the heart sounds library by Thinklabs [26], the heart sounds Podcast Series by Robert J. Hall Heart Sounds Laboratory, Texas Heart Institute, Texas [27], Bioscience normal and abnormal heart sounds database (BHSD) [28], and the Cardiac Auscultatory Recording Database (CARD) [29]. In addition, a book by D. Mason comprises a CD with a limited number of heart sounds and murmurs [13]. Similarly, heart sounds signals

Table 1.1: Various pathological conditions of the cardiovascular system.

Name	**Possible Pathology**
Split in S_1	RBBB
Split in S_2	LBBB
Third heart sound	Mitral regurgitation
Fourth heart sound	Aortic stenosis/pulmonary arterial hypertension
Opening snap	Mitral/tricuspid stenosis
Ejection Sound	Aortic/pulmonary stenosis
Click Mid-systolic	Mitral valve prolapse
Holosystolic Murmur	Mitral/Tricuspid regurgitation
Early systolic Murmur	Severe pulmonary stenosis
Late systolic Murmur	Pulmonary stenosis
Mid systolic murmur	Mitral regurgitation
Early diastolic murmur	Aortic/Pulmonary regurgitation
Mid Diastolic murmur	Aortic/Mitral regurgitation
Presystolic murmur	Mitral stenosis and/or Tricuspid regurgitation
Continuous murmur	Pulmonary embolism

from an audio-visual presentation by Tavel *et al* was also used as a database in some of the reviewed articles [30].

PhysioNet provides a large database gathered from various key databases and is generally considered as a reliable source of the data by many researchers [8], [21], [22], [31]. This is an open-access database assembled for an international competition the PhysioNet/Computing in Cardiology (CinC) Challenge 2016. In total, it comprises nine databases and recordings from 1297 subjects, collected in a variety of clinical and non-clinical conditions by different research groups around the world. The details of assembled databases are summarised here.

- Heart sounds databases by Syed [32] and Syed *et al* [33] included in the PhysioNet database, are referred as the Massachusetts Institute of Technology heart sounds database (MITHSDB). Syed [32] recorded heart sounds from 51 subjects (30 subjects with normal heart conditions or murmurs, and 20 patients suffering from mitral valve prolapse (MVP)), at the Massachusetts General Hospital, Boston, USA, and sampled at 44 kHz with 16-bit quantization. Information regarding the type of sensor employed for recording sounds was not reported [32]. In another study, Syed *et al* [33] recorded heart sounds from 39 subjects (15 with normal murmurs, 11 patients suffering from mitral regurgitation (MR), and 13 subjects were having normal heart conditions), using a Welch Allyn Meditron electronic stethoscope (Skaneateles Falls, USA) and sampled at 44.1 kHz with 16-bit quantization. Collectively, the Massachusetts Institute of Technology heart sounds database (MITHSDB) contains 409 heart sounds signals recorded at a sampling frequency of 44.1 kHz and 16-bit quantization from 121 subjects [8].
- Heart sounds recorded by Schmidt *et al* [34], [35], assembled in the PhysioNet database are referred as the Aalborg University heart sound database (AADHSDB). Schmidt *et al* [34] recorded 73 heart sounds signals using an electronic stethoscope (3M Littmann E4000) from 113 subjects (40 subjects for training dataset, and 73 subjects for test dataset) at the Department of Cardiology, Aalborg Hospital, Denmark, and sampled at 4 kHz with 16-bit quantization. In another study, Schmidt *et al* [35] recorded 435 heart sounds signals (231 normal and 204 pathological) using the same electronic stethoscope from 133 subjects (63 subjects identified with coronary artery disease and 70 subjects with non-coronary artery disease) at the Department of Cardiology, Aalborg Hospital, Denmark. Collectively, the Aalborg University heart sound database (AADHSDB) comprises 695 recordings sampled at 4 kHz with 16-bit quantization from 151 subjects [8].
- The Aristotle University of Thessaloniki heart sounds database (AUTHHSDB) was

obtained from Papadaniil and Hadjileontiadis [36]. Heart sounds were recorded using a custom-made electronic stethoscope "*AUDIOSCOPE"* in the Cardiac Clinic of Papanikolaou General Hospital in Thessaloniki, Greece. This database contains 45 recordings sampled at 4 kHz with 16-bit quantization from 45 subjects (11 normal subjects, 17 subjects identified with aortic stenosis, and 17 subjects identified with mitral regurgitation) [8].

- The K N Toosi University of Technology heart sounds database (TUTHSDB) contains recordings obtained from Naseri and Homaeinezhad [37] and Naseri *et al* [38]. Naseri and Homaeinezhad [37] recorded heart sounds from 50 subjects suffering from different valve diseases (aortic regurgitation, mitral regurgitation, aortic stenosis, and mitral regurgitation). In another study, Naseri *et al* [38] recorded 63 clean, and 63 noisy PCG signals from normal subjects; and subjects suffering from valvular diseases. The total number of subjects in the study was not found. In both the studies, heart sounds were recorded using an electronic stethoscope (3M Littmann 3200) at the Rajaei Cardiovascular, Medical & Research Center, Iran, and sampled at 4 kHz with 16-bit quantization. Overall, this database contained 174 recordings sampled at 4 kHz with 16-bit quantization from 44 subjects (28 normal subjects and 16 subjects identified with different types of valve diseases) [8].

- Moukadem *et al's* [39] database (University of Haute Alsace, France) contains 80 recordings that included 40 pathological cardiac signals sampled at 8 kHz with 16-bit quantization. These heart sounds were recorded using prototype electronic stethoscopes by different cardiologists in the Hospital of Strasbourg, France. The University of Haute Alsace heart sounds database (UHAHSDB) assembled in the PhysioNet database, contains 79 recordings (39 normal sounds and 40 pathological sounds) sampled at 8 kHz with 16-bit quantization from 55 subjects (25 normal and 30 cardiac patients) [8].

- Li *et al's* [40] and Tang *et al's* [41]–[43] databases assembled in the PhysioNet/CinC Challenge database was referred as the Dalian University of Technology heart sound database (DUTHSDB) [8]. This database comprises 673 heart sounds signals (338 normal and 335 pathological) recorded from 509 subjects (174 normal and 335 coronary artery disease (CAD) patients). The normal heart sounds were recorded using a microphone sensor (MLT201, ADinstrument, Australia) or a piezoelectric sensor (Xinhangxingye Technology Co. Ltd, China) at the Biomedical Engineering Lab in the Dalian University of Technology, China. The pathological sounds were recorded using an electronic stethoscope (3M Littmann) from patients diagnosed with coronary artery disease (CAD) in the Hospital of Dalian Medical University, China. This database was

sampled at different frequencies (800 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, 8 kHz or 22050 Hz) with 16-bit quantization.

- Heart sounds database created by Samieinasab and Sameni [44] was referred as the Shiraz University adult heart sounds database (SUAHSDB) in the PhysioNet database [8]. This database comprises 114 heart sounds signals (81 normal and 33 pathological sounds) recorded from 112 subjects (79 healthy and 33 patients) using a digital electronic stethoscope (GS Technology Co. Ltd, South Korea) and sampled at various frequencies (8 kHz, 44.1 kHz, and 384 kHz) with 16-bit quantization. Information regarding the place where these recordings were collected was not available.
- The Skejby Sygehus Hospital heart sounds database (SSHHSDB) comprises 35 heart sounds signals (12 normal and 23 pathological) recorded from 35 subjects (12 normal and 23 patients with heart valve defects) at the Skejby Sygehus Hospital, Denmark and sampled at 8 kHz. Information related to the type of sensor used for recordings and quantization rate was not found [8].
- The PhysioNet database also contains fetal heart sounds recordings contributed by Samieinasab and Sameni [44] (The Shiraz University fetal heart sounds database (SUFHSDB)). This database comprises 119 fetal heart sounds recordings from 116 fetal, and 92 maternal heart sounds recordings form 109 pregnant women. These sounds were recorded at the Hafez Hospital of Shiraz University of Medical Sciences, Iran, using an electronic stethoscope (GS Technology Co. Ltd, South Korea) and sampled at 8 kHz and 44.1 kHz with 16-bit quantization.

Apart from the PhysioNet database, other databases were also used for the validation of algorithms for heart sounds analysis and are briefly summarized here. The PASCAL database comprises two datasets, A and B, assembled for the Classifying Heart Sounds Challenge (CHSC) [23]. The dataset A contains 176 files recorded from the general public using the iStethoscope Pro iPhone app and the dataset B contains 656 files obtained from the clinical trials in the hospitals using a digital stethoscope (DigiScope) [23]. Recordings were collected in a real-world situation and the number of recordings was sufficient to validate the resutls, however, some important pathological information of heart sounds was suppressed due to preprocessing of data with a low-pass filter with cut-off frequency of 195 Hz. The type and the number of subjects used in the study were not available.

MP3 recordings with corresponding graphs of different heart sounds and murmurs are available from the University of Michigan Health System in Open Michigan Heart Sounds & Murmur Library (OMHSML) [24]. This database comprises 23 recordings of a total time duration of 1496.8s recorded using a stethoscope.

The Cardiac Auscultation of Heart Murmurs database (CAHMDB) provided by eGeneral Medical Inc. includes 64 heart sounds recordings of a total time duration of 338s [25]. These recordings are available online, but it requires payment to obtain user license. The information related to the type of subjects and methods to acquire heart sounds were not mentioned.

Other than these available databases, researchers have also collected their own data. Most of these recordings were obtained during clinical trials in the hospitals by auscultation using a digital stethoscope/microphone. A list of existing databases and their characteristics is provided in Appendix B. Most of the existing databases are restricted by the number of recordings, duration and sampling frequency. Also, other potentially important information such as gender, age and auscultation positions are not always specified, despite these being important for proper algorithm validation. In addition, in many cases, the signals had been pre-processed leading to the loss of both, pathological characteristics, as well as real-world artifacts which are nonetheless important to take into account when designing the algorithms/acquisition systems. Also, the length of individual recordings available is not sufficient to validate algorithms intended for continuous heart sounds analysis.

1.3.2 Segmentation Approaches

The purpose of heart sounds signals segmentation is to identify accurate locations of the fundamental heart sounds (S_1 and S_2). The peaks of S_1 and S_2 are required to determine the systolic and diastolic phases and to help in the subsequent estimation of cardiac cycles. This facilitates the identification and extraction of acoustic signals of interest in each cardiac cycle. Broadly, reported segmentation methods can be classified into: envelope based methods [39], [45]–[53], ECG and/or carotid pulse reference based methods [54]–[64], probabilistic models [34], [43], [65]–[74], feature-based methods [19], [36], [37], [75]–[79], time-frequency analysis based methods [80]–[84], and learning based methods [41], [85]–[87].

The works reported using the different segmentation methods have been classified in the following and summarised in Appendix C. In general, accurate segmentation is one of the most challenging tasks in heart sounds analysis, especially, when the signals are corrupted by real-world artifacts. Also, most of the available algorithms are designed to segment fundamental heart sounds. But other abnormal peaks and irregularities with low amplitudes also need to be investigated. In addition, existing segmentation algorithms mostly depend on absolute measures like time or frequency distributions which exhibit large disparity within subjects and hence result in poor segmentation accuracy. The limitations in the segmentation methods consequently impact the overall classification accuracy of the PCG signals.

1.3.2.1 Envelope-based methods

The envelope of heart sounds is used to identify S_1 and S_2 in the cardiac cycle using different approaches. Typical methods used for the envelope extraction are the normalized average Shannon energy, homomorphic filtering, Hilbert transform, moving window Hilbert transform, and short-time modified Hilbert transform.

In the time domain, the Shannon energy envelope attenuates low-intensity signal components and allows high and moderate intensity signal components to contribute to the envelope. Meanwhile, in the frequency domain, Shannon energy can represent the pressure difference across the heart valves and extract significant frequency components representative of their functioning. Shannon energy is the most common approach for envelope extraction of heart sounds. Liang *et al* [53] performed segmentation of heart sounds signals into four parts S_1, systole, S_2 and diastole, with the normalized average Shannon energy method and achieved segmentation accuracy of 93%. Moukadem *et al* [39] also proposed a method to construct the Shannon energy envelope based on the S-transform and achieved 97% and 95% of sensitivity and positive predictive values on pathological sounds. Similarly, the Hilbert transform based methods to extract the energy envelope of heart sounds were also reported. For instance, Sun *et al* [48] proposed a short-time modified Hilbert transform (STMHT) method to identify peaks in heart sounds signals. Hilbert transform was utilized to derive the instantaneous frequency and the Viola integral waveform method was applied to construct the envelope. Then the STMHT method automatically located the peak points by identifying the zero cross-over points. Authors calculated the accuracy of the proposed algorithm in two forms: one when S_1 and S_2 were detected separately, yielding an average accuracy of 98.53% and 98.31%, respectively; and the other one when S_1 and S_2 were not separated, which achieved 100% of accuracy [48]. These envelope-based methods suffer from the limitation that the envelopes are superimposed with murmurs and other artifacts, and were not always well correlated with cardiac events [88].

Analytical models based on a single-degree-of-freedom were also proposed to extract characteristic waveforms from heart sounds. For instance, Jiang *et al* [6] proposed a method to obtain cardiac sounds characteristic waveforms, where adaptive threshold values were evaluated using the fuzzy *C*-means (FCM) clustering algorithm. Graphical results were presented to support the novelty achieved in the work. Analytical information was not available for comparison purposes. Similarly, Yan *et al* [49] proposed a characteristic moment waveform function based method for the segmentation and achieved an accuracy of 99% by applying the Viola integral waveform method in the time domain to overcome the unexpected noise interference in the cardiac signal characteristic waveform envelope. However, a dataset of only 9 recordings was evaluated. A subsequent study [46] also constructed a heart sounds signal envelope using cardiac sounds characteristic waveforms to identify the locations of S_1 and S_2

with improved performance compared to the Shannon energy or Hilbert transform based methods. However, the assumption made regarding the duration of systole and diastole period in the proposed method is not always valid. Similar consideration was made by Kang *et al* [89] for the segmentation, i.e. the time interval between the two consecutive sounds lobes in the Shannon energy envelope to be less than 50 ms and the root mean square energy of one sound lobe to be less than 40% of the other sound lobe. The assumption regarding the duration of systole and diastole period was made for temporal localization of S_1 and S_2. Further, analytical results for the performance comparison between cardiac sounds characteristic waveform method and the Hilbert and Shannon based methods can be found in Choi and Jiang [51]. Statistical results confirmed that the cardiac sounds characteristic waveform method achieved higher accuracy (100% and 88.2%) compared to the Shannon envelope (78.2% and 89.4%) and Hilbert envelope (51.4% and 47.3%) methods, for the segmentation of the normal and pathological sounds, respectively. These results considered the case when the fuzzy *C*-mean clustering algorithm was applied to 500 selected cardiac cycles. Though there was a significant improvement in the segmentation accuracy using the cardiac sounds characteristic waveform method, noise components directly affect the features of the characteristic waveform and may produce misleading results [51]. To mitigate these issues, Sun *et al* [47] proposed a moving window Hilbert transform (MWHT) for the segmentation that achieved a higher sensitivity of 98.8% than the threshold based segmentation method reported in [90].

Researchers also suggested that heart sounds spectra analysis in the frequency domain can be performed for the segmentation. However, the problem with such segmentation methods was mainly with the thresholds value selection and the unexpected noise in the frequency domain [46]. For instance, Sepehri *et al* [50] identified S_1 and S_2 using the short time spectral energy and autoregression characteristics of fundamental heart sounds, yielding an accuracy of 93.6% while segmenting pediatric heart sounds. Authors in [45] employed 6^{th}-order Daubechies discrete wavelet because of its resemblance with the heart sounds components, to estimate the heart cycle using decomposed and reconstructed normalized phonocardiograms. The extracted normalized Shannon energy envelope and short-time average magnitude difference functions were used to identify heart cycles.

In general, most of the envelope-based segmentation algorithms perform heart sounds segmentation with an assumption that the systolic period is shorter than the diastolic period. However, this may not be true in the case of infants and other cardiac patients having abnormal heart sounds [91]. Furthermore, the envelope extraction algorithm fails if the fundamental heart sounds were wrongly identified or when additional peaks appear superimposed to the fundamental heart sounds. [56]. Also, medium amplitude peaks including murmurs are attenuated in the envelope analysis while large and low peaks may appear as a single envelope

[79]. Thus, these methods fail to locate peaks of very low amplitude present in the cardiac cycle [78].

1.3.2.2 ECG and/or carotid pulse reference-based methods

A number of the reported segmentation approaches require an auxiliary signal (ECG signal and/or carotid pulse) as a reference to identify the locations of fundamental heart sounds in the cardiac cycle [54], [55], [64], [92], [56]–[63]. For instance, Gharehbaghi *et al* [55] presented a heart sounds segmentation method by assessing the beginning and end points of each cardiac cycle using simultaneously recorded ECG signals. This method achieved 97% and 94% of accuracy in identifying S_1 and S_2, respectively. In this study, results were validated using pediatric heart sounds signals only. Similarly, El-Segaier *et al* [60] recorded ECG signals simultaneously with heart sounds signals from children to detect S_1 and S_2 in cardiac cycles. The detection rate of S_1 and S_2 was 100% and 97%, respectively, when *R*-waves and *T*-waves in the ECG signals were used as a reference. However, authors excluded 20% of the starting and 30% of the late systole intervals for the analysis of systolic murmurs. In another study, Jabbari and Ghassemian [57] performed the segmentation of cardiac signals into separate heart cycles using simultaneously recorded ECG signals. The different intervals in the cardiac cycles were identified using *R*-peaks and *T*-peaks of ECG signals recorded. Statistical results achieved in the segmentation method were not reported. Similarly, Ahlstrom *et al* [59] identified S_1 and S_2 in a certain time window using the *R*-peaks and *T*-peaks of the ECG signals recorded simultaneously with the heart sounds signals. Segmentation results were not found in the study.

The general disadvantage of these methods is that a secondary signal is required, which is more complex both, from sensing point of view and synchronization. Also, these methods are affected by the mismatch in timing between the electrical and mechanical (E-M) activities of the cardiovascular system, which in turn depend on the pathological conditions of patients [88]. And methods that require the identification of *R*-peaks and *T*-peaks are more computationally hungry and demanding in processing power. In addition, accuracy also varies with low amplitude and abrupt changes in the QRS morphologies, which can make the identification of the *R*-peaks and *T*-peaks complex on its own.

1.3.2.3 Probabilistic models for the segmentation

As envelope-based methods have shown a modest success, many probabilistic models for the segmentation were reported in recent studies to try to overcome their shortcomings. The aim of probabilistic models is to characterize the fundamental heart sounds based on some discriminative features, such as temporal correlation, waveform function, time-frequency

energy, and other physiological information.

The hidden Markov models have proven suitable for the temporal localization of heart sounds. More recently, Schmidt *et al* [34] proposed a duration-dependent hidden Markov model that distinguished the different heart sounds based on the duration of the events in the cardiac cycle. Later, Springer *et al* [70] proposed a hidden semi-Markov model to try to boost the performance with the help of a priori information about the expected duration of the states. This model incorporated logistic regression and the extended Viterbi algorithm to improve the segmentation performance and was employed in various subsequent studies [67]–[69]. Further to identify wrongly segmented cycles, Abdollahpur *et al* [67] proposed an algorithm for the segmentation using modified Springer *et al's* [70] hidden semi-Markov model and a cycle quality assessment (CQA). Results were demonstrated in relevance to features extracted and noisy cycles were identified and rejected.

Gill *et al* [65] suggested an unsupervised identification method using a hidden Markov model and homomorphic filtering to extract smooth envelops of the heart sounds. In this method, event-based features were extracted to estimate the sequence of states and to detect peaks in the envelogram of heart sounds. Later, Gupta *et al* [74] reported a method for the segmentation using homomorphic filtering and *k*-means clustering. This method considered the fundamental heart sounds as an amplitude modulated waveforms whereas murmurs as an amplitude and frequency modulated waveforms. More recently, Hassani *et al* [71] also proposed a segmentation method using a homomorphic filtering that was applicable to heart sounds and murmurs. However, the sensitivity of the proposed method was not suited for the clinical applications.

Kwak and Kwon [73] proposed a hidden Markov model (HMM) based segmentation that was able to identify the starting points of heart sounds buried in murmurs that were missed by the conventional envelope-based segmentation method. Proposed HMM-based segmentation outperformed the envelope-based segmentation methods that identified S_1 and S_2 sounds as peaks exceeding a determined threshold; whereas rejecting other extra peaks. In another study, mel-frequency cepstral coefficients (MFCCs) along with the Shannon energy feature vectors and regression coefficients were utilized to develop a probabilistic model [66]. The Hilbert-Huang Transform proposed by Tseng *et al* [72] has proven suitable for detecting the presence of S_3 and S_4 in a cardiac cycle at the event level. However, the selection of intrinsic mode functions (IMFs) required in the model is challenging because of the varying temporal-spectral characteristics of heart sounds.

Though these models were efficient in improving the performance of the segmentation methods, the overall accuracy was more limited in some methods. Also, the characteristics of

the fundamental heart sounds which were used to develop various models, varied largely from infants to old people and from healthy to cardiac patients.

1.3.2.4 Feature-based methods

Feature-based methods are based on extracting certain features such as the energy fraction, sample entropy, total variation filtering, Shannon entropy, instantaneous phase boundary, boundary location identification, likelihood computation, etc.; to identify peaks present in the cardiac cycles. The main drawback of these methods is that the extracted features may vary with the signals they are tested on and hence they need to be verified using standardized databases.

Naseri and Homaeinezhad [37] applied an effective detection-delineation decision statistic derived after combining the frequency and amplitude based features, evaluated from the extracted segments. This work was extended to detect S_3, S_4, and the presence of murmurs based on regenerated decision statistic from located S_1 and S_2. The complexity-based segmentation method proposed by Nigam and Priemer [79] was insensitive to the amplitude and frequency variations and utilized the inherent complexity of the heart sounds signals. However, quantitative information was not available, and results were supported by the graphical representations.

Recent studies applied an Empirical mode decomposition (EMD) to obtain instantaneous frequency information and to decompose signals into intrinsic mode functions [16], [72], [93]. Papadaniil and Hadjileontiadis [36] reported heart sounds signals segmentation by employing an ensemble Empirical mode decomposition (EEMD) combined with the kurtosis features to detect S_1 and S_2. Though the algorithm achieved promising results in localizing S_1 and S_2 sounds, the Empirical mode decomposition has some limitation in identifying two closely spaced spectral components and mode mixing. In order to overcome these issues, a Hilbert vibration decomposition (HVD) and an ensemble Empirical mode decomposition (EEMD) were proposed, respectively [9], [94]. More recently, Varghees and Ramachandran [95] demonstrated the advantage of the Empirical wavelet transform over the Empirical mode decomposition as a consistent decomposition approach for the heart sounds segmentation. Authors considered low frequency (10-150 Hz) heart sounds signals for detecting S_1, S_2, S_3 and S_4; whereas high frequency (150-600 Hz) heart sounds signals to detect murmurs. Thus, the overlap between normal and abnormal heart sounds signals was ignored during the mode estimation process.

1.3.2.5 Time-frequency/Wavelet analysis-based methods

Other methods reported for the segmentation were based on time-frequency analysis of heart sounds signals. White *et al* [96] used the Pseudo-Wigner-Ville distribution to identify murmurs. Statistical data was missing for comparison and the study was conducted on heart sounds obtained from children. In another study, Kim and Tavel [97] performed a spectral analysis for the segmentation of murmurs using time-frequency representation. The reported study establishes a correlation with the Doppler ECG and derived a mean pressure gradient at different frequencies without proposing a generic algorithm. In the time-frequency domain based segmentation method proposed in [80], the authors mitigated the traditional problem of selecting parameters of the analysis window, by employing Rényi entropy and obtained the optimal window length. Sanei *et al* [81] also separated murmurs from heart sounds signals using adaptive singular spectral analysis in the time series. However, the only correlation was established, and the statistical parameters were not reported for proper comparison.

Wavelet transforms were frequently used in recent studies for the time-frequency analysis. Reed *et al* [15] demonstrated a wavelet decomposition based method for analyzing heart sounds signals. However, the classification accuracy of the proposed system was a function of the signal-to-noise ratio and the variance of the added noise. In another study, H. Nazeran [84] proposed a Morlet wavelet-based segmentation method using the estimated normalized average Shannon energy. The proposed method was effective in localizing heart sounds components compared to the Fourier transform. However, segmentation accuracy was not reported in the study for comparison. Further, Dokur and Ölmez [83] used an adaptive peak detector based on the third, fourth and fifth level decomposition coefficients. The reconstructed signal was squared and smoothed by the moving window integration so that the locations of S_1 and S_2 become apparent. Segmentation rate of the proposed algorithm was not reported. Later, Meziani *et al* [98] proposed an algorithm to locate the peaks of S_1 and S_2 and to extract different murmurs using the Daubechies wavelet. Wavelet transform was used to denoise signals and to construct Shannon energy envelope. The manual selection of the threshold was the major drawback of this algorithm and segmentation accuracy was not found in the article for the comparison.

Patidar and Pachori [82] segmentation method used a tunable-Q wavelet transform to remove murmur contents so that the fundamental heart sounds possessing maximum kurtosis are obtained in the reconstructed signals. This method was helpful in improving the localization of the fundamental heart sounds with respect to the overlapping murmurs by varying the Q-factor. Effective features were extracted in the study to reduce the computational complexity of the algorithm. The wavelet packet decomposition was used to extract diagnostic features and to detect and classify heart sounds, including murmurs [99]. Further, Wang *et al* [100] proposed an optimum multi-scale wavelet packet decomposition (OMS-WPD) and identified murmurs

and normal heart sounds using the optimum decomposition coefficients. The main drawback associated with this method was the requirement of additional computational power for extracting wavelet characteristics of the signals for the segmentation. Also, the proposed algorithm achieved an identification accuracy of 88.98%, which was not significant for the clinical applications. More recently, Fahad *et al* [101] also applied wavelet analysis to segment heart sounds. Shannon energy of the reconstructed wavelet signals was computed to detect S_1 and S_2 sound lobes. Peaks with larger time intervals were identified as diastole and those with smaller time intervals were identified as the systole intervals.

1.3.2.6 Learning-based methods

Learning-based approaches performed the segmentation by training algorithms with the extracted features. The time-delay neural network employed by Oskiper and Watrous [87] for identifying S_1 in the cardiac cycle, was comprised of a single hidden layer with time-delay links that connected the hidden units to the time-frequency energy coefficients of a Morlet wavelet decomposition. However, only S_1 sounds were identified in the article. Similarly, Tang *et al* [41] decomposed heart sounds signals into components characterized by different features in the frequency-time-frequency domains. A fuzzy detection method was used to separate signals of interest and then S_1 and S_2 sounds were identified. Nigam and Priemer [86] also proposed a simplicity-based fuzzy clustering approach for the identification of murmurs. However, the overall accuracy achieved was significantly reduced.

In recent approaches, researchers segmented cardiac signals directly into cardiac cycles and skipped the steps used to identify individual locations of S_1 and S_2 peaks [45], [102]. In such approaches, initial investigation of S_1 and S_2 was skipped if prior knowledge of cardiac cycles was available. For instance, Yuenyong *et al* [56] estimated cardiac cycles and classified them into normal and abnormal heart sounds without identifying fundamental heart sounds. Similarly, S. Deng *et al* [91] proposed the autocorrelation features using subband envelopes coefficients and diffusion maps based classification without performing the segmentation. These autocorrelation features were further employed to excerpt unique feature representations of heart sounds that could be classified using support vector machines. Similarly, authors in [103] classified heart sounds using feed-forward neural networks (FFNNs) without performing the segmentation. However, the assumptions regarding the systolic and diastolic intervals and adaptive threshold values, while identifying S_1 and S_2 are not valid in all conditions. Hence these methods have similar drawbacks to the envelope analysis.

1.3.3 Feature Extraction

Representation of the cardiac signals in different domains reveals various physiological and pathological characteristics and allow efficient feature extraction. To capture concurrent variations and structural components in both time and frequency, time-frequency representation of the transient signals has been reported as a preferred mean over the time-domain and frequency-domain representations. Qualitative and quantitative measurements of the signals are obtained using different transforms for heart sounds analysis; for instance, time-frequency representation using *S*-transform [39], Fourier transform, short-time Fourier transform (STFT) [104], Wigner-Ville distribution (WVD) [93], [105], Choi–Williams distribution (CWD) [106], [107], wavelet transform [108], [109], and short-time modified Hilbert transform (STMHT) [48]. Though, the short-time Fourier transform was found to be popular, obtaining a proper resolution for the feature extraction using STFT is challenging because of the fixed window available for the analysis [102], [110], [111]. Wavelet analysis emerged as an alternative by substituting the frequency shifting operation of the STFT by a time or frequency scaling operation [111]. Wavelet transform was widely reported in literature because of its suitability for representing signals where the length of the temporal window can be engineered for multi-resolution analysis with wide frequency range across the length [112]. Discrete wavelet transform (DWT) [56], continuous wavelet transform (CWT) [113]–[115] and mel-scaled wavelet transform [77], [116]–[118], have all been used for heart sounds analysis.

Feature extraction and selection play an important role in pattern recognition and classification. Features with high-order statistics, non-linear fractal complexity, entropy information and chaos theory helped in capturing relevant information from non-stationary PCG signals, required for the proper classification. Other features included the Shannon energy envelope of the frequency spectrum, wavelet coefficients, perceptual features such as mel-frequency cepstral coefficients (MFCCs), bispectrum, variance fractal dimension (VFD), and fractal features such as largest Lyapunov Exponents or Hurst Exponent. A brief overview of the different approaches reported for the significant feature extraction is discussed in the following.

Time-domain features are easy to extract and quantify. These features mainly included the timing characteristics such as locations of S_1 and S_2, systolic and diastolic intervals, and amplitude information (such as the mean absolute value of the S_1 and S_2 and other peaks in the cardiac cycle). Some physiological and pathological information that are missed in the time-domain analysis can be visualized in the frequency domain analysis. Frequency-domain based feature extraction methods used mainly included band-pass filter banks and zero-crossing analysis [64], [80], [107], [119]–[122]. Other discriminant features included the mean power

of distinct cardiac sounds segments (S_1, S_2, systole and diastolic) in different frequency bands and mel-frequency cepstral coefficients (MFCCs) [77], [116]–[118].

Though time-domain and frequency-domain based features are easy to extract, some physiological and pathological information is obscured. Most of the recent studies extracted pathological information by analyzing the signal in the time-frequency domain [80], [102], [123].

Iwata *et al* [64] firstly proposed an algorithm to detect the S_1 and S_2 by spectral tracking of low-frequency components. Later, Debiais *et al* [124] extracted the spectrogram, Choi-Williams and Bessel coefficients for the time-frequency representation of the heart sounds and applied these extracted parameters to the numerical simulations. Energy distribution coefficients in the time-frequency domain were also analysed for the feature extraction [120]. Further, Choi and Jiang [125] proposed maximum frequency (f_{max}) and maximum width (f_{width}) features for the normalized auto-regressive power spectral density (NAR-PSD) to classify heart sounds. The author in [46] proposed a short-time modified Hilbert transform based frequency feature matrix comprised of features mainly the frequency width of the envelope in the frequency domain and identified the fundamental heart sounds; which was later used to detect cardiac abnormalities. Power spectral density and energy of heart sounds signals were estimated using the auto-regressive model to estimate peaks' boundaries. Teo *et al* [126] proposed a power spectral analysis to extract features; mainly the ratios, mean and standard deviation of ratios between the sum of powers of the N highest peaks to the total power in the spectrum for individual segments, and later fed them to a neural network based classifier. In a subsequent study, Schmidt *et al* [35] extracted features from the different frequency bands using the sub-segments of the short duration windows in the systolic period. More recently, Whitaker *et at* [127] segmented heart sounds using Springer *et al's* [70] hidden semi-Markov model and transformed them with an N-point FFT and sparse coefficient matrix for the feature extraction.

Mel-frequency spectral coefficients (MFSCs) have also proven suitable for feature extraction. Temporal alignment techniques, such as the dynamic time wrapping (DTW) reported in [128], was applied to mel-frequency cepstral coefficients (MFCCs) based features for the classification, mainly to reduce the impact of subject-specific differences. More recently, authors in [129] extracted the mel-frequency spectral coefficients features that evaluated MFCCs without performing discrete cosine transform (DCT) and concluded that MFSCs analysis outperforms MFCCs.

Wavelet decomposition and reconstruction provided a significant information of the spectral components of heart sounds signals and was used in recent studies to extract features [56], [130], [131]. Methods for wavelet analysis included Debauchies transformation [110],

Morlet transformation [107] and Stockwell transformation. The type of wavelet used for the analysis was selected from the family of wavelets- Haar, Daubechies, Biorthogonal, Coiflets, Symlets, Morlet, Mexican hat and Meyer. Eugen *et al* [114] examined the common wavelets transforms on phonocardiography signals to determine the best wavelet to obtain time-frequency representations. The study found that the continuous wavelet transform using a Morlet wavelet was a reliable approach for the feature extraction [114].

In another study, Zheng *et al* [77] evaluated features such as energy fraction, sample entropy and maximum energy fraction, of S_1, S_2 and murmurs, in frequency sub-bands using wavelet packets decomposition. In one of the recent study, Patidar *et al* [132] proposed tunable-Q wavelet transform (TQWT) based decomposition method to extract the diagnostic features.

Earlier, Turkoglu *et al* applied wavelet packet decomposition (WPD) and wavelet packet entropy to extract features of the Doppler signals [133]. Later, Choi [134] suggested wavelet packet energy features using wavelet packet decomposition of heart sounds signals. Wavelet analysis was used by Dokur and Ölmez [83] to estimate powers of the decomposition level coefficients in all subbands for the feature extraction. In a further progress, Wang *et al* [100] proposed a feature extraction method based on the optimum multi-scale wavelet packet decomposition (OMS-WPD) and wavelet-time entropy to extract features of heart sounds signals. In a subsequent study, Zheng *et al* [135] extracted the multi-fractal spectrum parameters, frequencies corresponding to the maximum peak of the normalized power spectral density curve (f_{PSDmax}), and adaptive sub-band energy fraction. These features included the ratio of diastolic to systolic duration, the ratio of the amplitude of S_1 and S_2 (S_1/S_2), multi-fractal spectrum parameters, the frequency of the maximum peak of the normalized power spectral density curve (f_{PSDmax}) and adaptive sub-band energy fraction. Apart from these features, features such as perceptual and fractal features by Delgado-Trejos *et al* [58], 2D-PCA based features [136], were also reported to characterize different murmurs. More recently, Kay and Agarwal [69] extracted time-frequency features (mel-frequency cepstral coefficients, inter-beat and complexity features) using continuous wavelet transform, considering Morlet wavelet as the mother wavelet.

From the reported works, it was found that the time-frequency representation, mostly using wavelet analysis, proved suitable for the feature extraction. However, features were hard to extract especially those focused on the detection of a particular abnormality. Also, wavelets are non-symmetric and may produce artifacts at wavelet sub-bands borders [137]. Other than this, wavelet package decomposition level is sampling frequency dependent and hence, inaccurate estimation of the time-frequency behaviour of the signal may impact the performance of the algorithms [77]. Details for above-mentioned feature extraction methods and type of features extracted are summarized in Appendix D.

1.3.4 Classification of Cardiac Acoustic Signals

The final step of a heart sound analysis algorithm is to take the extracted features and fed them to an appropriate classifier to interpret them. Reported approaches for the classification include Support Vector Machine (SVM) [45], [46], [125], [127], [132], [134], [135], [138]–[142], [47], [143]–[147], [73], [77], [85], [91], [100], [102], [108], Hidden Markov Models (HMM) [101], [116], [117], [148], [149], *k*-Nearest Neighbour (*k*-NN) [58], [107], [136], [150], [151], Neural Networks [15], [39], [83], [84], [89], [103], [126], [129], [133], [152]–[154], [52], [155]–[164], [56], [165]–[169], [57], [59], [67], [69], [74], [75], rule-based classifier or decision trees [170], [171], BayesNet classifier [131], machine learning based approaches [54], [144], [155], Gaussian-Bayes model [172], Naïve Bayes [173], Gaussian Mixture Model (GMM) [174], random forest [173], and discriminant analysis [106]. Works reported using these approaches are presented in Appendix E.

1.3.4.1 Support Vector Machine

Support vector machines are non-probabilistic binary linear data-based machine learning models suitable for the classification of heart sounds using different kernel functions (such as a Gaussian, sigmoid etc.). Implementation of support vector network for non-separable training data was firstly reported by Cortes and Vapnik [175] which has been extended to handle supervised machine learning problems such as classification of heart sounds signals.

In general, most of the studies are focused on improving the classification performance either by modifying existing approaches using SVM classifiers or by adding new features to the classifier. Wavelet-based features are commonly used for the abnormality detection and classification. For instance, Choi [134] used wavelet packet decomposition and support vector machine to detect the aortic and mitral valvular heart disorders. These disorders were identified based on the fact that they differ from the normal heart sounds in the frequency spectrum. However, this may not be valid for all kind of disorders examined. Choi and Jiang [125] explored a normalized auto-regressive power spectral density (NAR-PSD) curve for the feature extraction and an SVM classifier to improve the classification accuracy. However, the accuracy of the algorithm was found to be dependent on the disease under investigation.

Other features are also investigated and used with an SVM classifier. For instance, Banerjee and Mondal [138] proposed features based on the inherent structural complexity of heart sounds computed in terms of sample entropy. These features were evaluated using an SVM classifier and tested using signals of different noise levels. The accuracy achieved was promising for the clean signals, however, there was a significant variation when the noise was added to the signals. In another study, Whitaker *et al* [127] extracted features using sparse coding and

obtained spectral characteristics that were fed to an SVM classifier. Classification results achieved were below 90% and some additional features were suggested.

Further attempts to reduce the classification errors included the modification in the existing SVM classifier and selecting an optimum kernel function and parameters. For instance, Avci [145] proposed a genetic-SVM (GSVM) classifier to diagnosis heart valve diseases and tested with eight different kernel functions (linear, polynomial, radial basis, exponential radial basis, sigmoid, spline, Fourier and bspline). It was found that there was a significant improvement in the classification performance with the genetic-SVM classifier using the exponential radial basis function. The proposed method also emerged better compared to the wavelet neural network classifier proposed by Turkoglu *et al* [165]. In a subsequent study, Ari *et al* [108] suggested a least square SVM (LS-SVM) classifier by modifying the Lagrange multiplier based on least mean square algorithm to classify heart sounds using wavelet-based feature vector. The performance of the LS-SVM with radial basis kernel function was tested in classifying normal and five different pathological sounds. Results achieved were highly dependent on the number of pathologies in abnormal cases. Later, Kwak and Kwon [73] proposed the murmur likelihood and HMM state likelihood based features to classify different types of murmurs and normal heart sounds using an SVM classifier [73]. It was concluded that by adding murmurs likelihood to an SVM classifier, there is a significant improvement in the classification performance compared to the multi-layer perceptron and radial-basis function networks.

In another study, Redlarski *et al* [85] combined an SVM classifier and modified Cuckoo search (MCS) algorithm to increase the diagnostic accuracy of the system. Also, linear predictive coding was used to solve the non-stationary problem of the heart sounds signals and feature extraction. Though the proposed algorithm was validated with a database comprised of normal and various abnormal heart sounds, the number of samples in the recording was too small.

More recently, Zheng *et al* [77] used an SVM classifier with four different kernel functions (Gaussian radial basis kernel function (GRKF), linear kernel function (LKF), polynomial kernel function (PKF) and sigmoid kernel function (SKF)) and evaluated the energy fraction and sample entropy features from the reconstructed selective frequency elements of heart sounds signals. It was found in the study that the Gaussian radial basis kernel function (GRKF) produced the best results in classifying normal, aortic stenosis, pulmonary stenosis, tricuspid insufficiency and mitral insufficiency using heart sounds signals. However, a limited number of heart sounds signals were recorded from 42 healthy and 38 patients for the analysis. Patidar *et al* [132] suggested a tunable-Q wavelet transform (TQWT) to extract features and classified heart sounds using a least square SVM with various kernel functions (radial basis function, Morlet wavelet kernel and Mexican hat wavelet kernel function). The proposed study proved

promising in identifying normal, valvular defects, septal defects, and other defects. The main findings were the segmentation followed by a tunable-Q wavelet transform and least-square support vector machine (LS-SVM) based classification with Morlet wavelet kernel function. Similarly, Zheng *et al* [135] also suggested a least square LS-SVM classifier. However, intense features were calculated based on multifractal detrended fluctuation analysis (MF-DFA), maximum entropy spectra estimation (MESE) and the empirical mode decomposition (EMD). The proposed diagnostic features achieved promising results in identifying cases of chronic heart failure. Also, the classification performance of LS-SVM was found to be better than that of the back-propagation artificial neural network (BP-ANN) and hidden Markov model (HMM). However, subjects with murmurs were not considered in the study that may lead to the misclassification.

In another study, Gharehbaghi *et al* [139] suggested a growing time support vector machine (GTSVM) to capture the characteristics of innocent murmurs and the results achieved were compared with a conventional SVM classifier. Significant improvements in the performance were achieved compared to the conventional SVM and multi-layer perceptron (MLP) in separating innocent and pathological murmurs, however, the overall classification accuracy achieved was still below 90%. Overall, the SVM classifier was used in most of the cases with novel features or feature extraction techniques.

1.3.4.2 Hidden Markov Models

Hidden Markov Models (HMMs) are one of the most promising probabilistic statistical classifiers and their classification performance is investigated in various reviewed articles and significant modifications are proposed to improve the classification accuracy.

In most of the studies, either modification to the existing hidden Markov model or adding new features to the existing classifier using different transforms was suggested. Like as Uğuz *et al* [149] used a hidden Markov model to classify heart sounds using wavelet transform and short-time Fourier transform to extract features. However, the Doppler ultrasounds were used for the analysis. Similarly, Wang *et al* [116] also extracted features using a time-domain transform, short-time Fourier transform and mel-frequency cepstral coefficients (MFCCs) and classified heart sounds signals using a modified-HMM. MFCCs based features using hidden Markov model yielded promising results compared to the time-domain and short-time Fourier transform based features. Later, Chauhan *et al* [117] proposed a mel-frequency cepstral coefficients (MFCCs) based selective features and a hidden Markov model to classify normal sounds, continuous murmurs, diastolic murmurs and systolic murmurs in a cardiac cycle. Promising results were achieved, however, a small dataset of recordings was considered for the

validation of the algorithm proposed to identify ejection clicks, opening snaps, split S_1, split S_2, S_3 and S_4.

In another attempt to improve the classification performance using a hidden Markov model, Schmidt *et al* [176] proposed a duration-dependent HMM in which the durations of the states were estimated using an autocorrelation analysis of a homomorphic envelogram. It was found in the study that the duration-dependent HMM performed better compared to the conventional HMM in classifying heart sounds. In a subsequent study, Saraçoğlu [148] classified frequency domain features obtained using a principal component analysis (PCA) dimension reduction technique using an HMM-based classifier, and achieved an average classification accuracy of 72.2% that was not suitable for the clinical applications. More recently, Fahad *et al* [101] suggested an adaptive-neuro fuzzy inference system and an HMM-based classifier to classify heart sounds signals. This work was focused on the classification of normal sounds and valve disorders (mitral stenosis, mitral regurgitation, and atrial regurgitation).

1.3.4.3 *k*-Nearest Neighbor (*k*-NN)

The feature distances (Euclidean, Manhattan, Minkowski, Mahalanobis, etc.) were estimated to compute the nearest neighbors when the most relevant patterns were close to each other in the feature space [170], [172]. The *k*-nearest neighbor algorithm was successfully applied to classify heart sounds recordings [58], [150], [177], [178]. Classification performance was found to be dependent on the *k* parameter used in the algorithm and accordingly, various features were suggested to improve the classification performance.

Bentley *et al* [107] applied the spectrum-analysis technique and extracted nine features to classify various heart valve sounds using an *k*-NN classifier. However, out of nine, two features were based on simultaneously recorded ECG signals. Delgado-Trejos *et al* [58] extracted the time-varying and time-frequency features, perceptual features and fractal features to classify murmurs using an *k*-NN classifier. The best results were achieved with the fractal features (97.17%) followed by the time-varying and time-frequency features (95.28%) and perceptual features (88.7%). In another study, Quiceno-Manrique *et al* [150] investigated spectral power distribution of abnormal heart sounds using various energy distributions (Wigner-Ville, Choi-Williams, smoothed pseudo-Wigner-Ville, exponential, and hyperbolic T-distribution). It was concluded that the addition of energy distributions to the analysis of heart sounds improved the classification performance compared to the short-time Fourier transforms and continuous wavelet transforms representation using an *k*-NN classifier. Amit *et al* [106] also investigated various signal transforms to extract features and presented a hierarchical clustering approach for heart sounds classification using an *k*-NN and discriminant analysis. However, results achieved were for the S_1 classification and no results were reported for other heart sounds.

Another time-frequency representation proposed was based on time-varying auto-regression [136]. Extracted features were fed to an *k*-NN classifier to detect murmurs. The proposed method yielded promising classification results, however, only 45 recordings were used for the validation. More recently, Hamidi *et al* [179] also proposed a method to classify heart sounds using a nearest neighbor classifier by evaluating the Euclidean distance. The proposed algorithm exhibited large variation in the classification accuracy with the different databases. In some of the studies, the classification task was achieved using an *k*-NN classifier along with another classifier to enhance the performance of the system. In general, an *k*-NN classifier offers advantages in terms of training time, simplicity and ease of implementation compared to others, however, it demands large memory space and offers slow estimation [180].

1.3.4.4 Neural Networks

Neural networks-based classifiers have proven potential as a computational tool for pattern classification of heart sounds with good classification performance. Artificial neural networks (ANNs) were most widely used along with the description logics [181] and linear/non-linear discriminant learning. Ölmez and Dokur [166] classified seven different heart sounds (aortic stenosis, mitral regurgitation, mitral stenosis, pulmonary stenosis, aortic regurgitation, summation gallop and normal sounds) using a grow and learn (GAL) network and linear vector quantization (LVQ) network. LVQ provided the benefit of quick learning time, however, it generated a large number of feature vectors because of its non-incremental network. Studies were also focused on the reduction of feature vectors to reduce the complexity of the neural networks. Like as, Andrisevic *et al* [163] suggested an artificial neural network to classify feature vectors extracted using a principal component analysis (PCA). The proposed approach was tested with a limited database and a low classification performance was achieved. Later, Uğuz [157] proposed a similar approach using a PCA-ANNs that achieved better results compared to [163]. The main contribution was to apply a discrete Fourier transform and a Burg auto-regressive spectrum analysis for the feature extraction; and a principal component analysis to reduce the dimension of the feature vector.

Some studies were focused towards the addition of new features or transformation to extract features that may help in the identification of cardiac abnormalities. Higuchi *et al* [162] proposed an automated approach to diagnosis cardiac diseases by analyzing heart sounds signals using a three-layered artificial neural network. However, the algorithm failed to classify ventricular septal defects and yielded an overall classification accuracy below 70%. Similarly, H. Nazeran [84] suggested wavelet features extracted from heart sounds signals and fed them to a neural network based classifier. This study also achieved a limited classification accuracy (70%) that was not suitable for the clinical applications. In a subsequent study, Sepehri *et al*

[160] proposed a neural network to classify energy features extracted from the Arash-Band with lowest clustering errors. Promising results were achieved using the algorithm in identifying congenital heart diseases in children.

Other significant modifications in the existing neural networks-based approaches were also found. In a first, Akay *et al* [169] applied a fuzzy-neural network to detect coronary artery diseases using wavelet features and achieved a limited classification accuracy. The results obtained also suggested that further modification in the terms of feature extraction or the classification method was required. Later, Folland *et al* [167] proposed a multi-layer perceptron (MLP) and radial basis function neural network (RBF-NN) that were trained with features extracted using FFT and the Levinson-Durbin auto-regression algorithms. These proposed algorithms also demonstrated a limited accuracy in classifying normal and abnormal heart sounds. In a subsequent study, Turkoglu *et al* [133] proposed a wavelet packet neural network based on the wavelet transform theory and added a multi-layer perceptron (MLP) layer to the classifier. Though the algorithm proved successful in classifying the normal and abnormal heart sounds, murmurs were not considered for the evaluation. Also, the Doppler heart sounds were used in the study. Later, Javed *et al* [52] extracted features from heart sounds signals segmented into fundamental heart sounds and murmurs and fed to a multi-layer perceptron (MLP) neural network. Further, a smoothed Pseudo Wigner-Ville distribution (SPWVD) was used to classify murmurs. However, the correlation between the two segments (FHSs and murmurs) and overlap were ignored in the analysis.

In another study, Babaei and Geranmayeh [159] proposed a versatile multi-resolution wavelet-based algorithm using a multi-layer perceptron structure with back-propagation (MLP–BP) algorithm based artificial neural network. Though the proposed algorithm achieved promising results in classifying aortic insufficiency, aortic stenosis, pulmonary stenosis and normal heart sounds, other abnormalities need to be investigated. In a further improvement, Jabbari and Ghassemian [57] proposed a multivariate matching pursuit (MMP) based on a three-layer feed-forward multilayer perceptron network. Though results were promising, the multilayer perceptron was based on a back-propagation algorithm and hence required a longer training time [166]. Also, assumptions made regarding the time interval of S_1-S_2 and S_2-S_1, are not always valid. Recently, Gharehbaghi *et al* [156] proposed a time growing neural network (TGNN) to classify features extracted from the spectral analysis of the frequency bands with a growing length window. The performance of the proposed neural network was found to be better compared to a time-delay neural network and the multi-layer perceptron mainly because of the reduced environmental noise using the suggested growing length window. The algorithm needs to be validated using heart sounds database from adults.

Dokur and Ölmez [83] proposed an incremental neural network for the classification based on the Kohonen's self-organizing map (SOM) network and an incremental self-organizing map (ISOM). A significant performance improvement was found in the study while classifying fourteen different types of heart sounds, mainly because of the incremental network that automatically determined the optimum number of nodes required for the segmentation. Although the proposed algorithm achieved a good classification accuracy (95%), the algorithm needs to be validated with a large database. More recently, Abdollahpur *et al* [67] employed a feed-forward neural network to classify features extracted from the segments that had successfully passed the cycle quality assessment. It was concluded that the cycle quality assessment algorithm proposed in the study boosted the classification performance.

An effort to classify heart sounds using a combination of different classifiers was also reported. For instance, authors in [153] proposed a classification approach using a 2-means clustering and an ANN classifier. The unsure data were identified by employing a *k*-mean clustering approach and an ANN model was used for the classification. This method yielded a mean accuracy of 86.5%. Conroy *et al* [182] proposed a dynamic ensemble approach to generate a complete dataset from the missing information of data using a two-stage machine learning process. This was based on the AdaBoost, a powerful machine learning technique, and the goal was to construct a high-performance classifier from an ensemble of "weak learners". This work was further extended by Potes *et al* [154] to classify heart sounds using a variant of the AdaBoost classifier and a deep learning convolutional-neural network classifier. It was found that the ensemble of the AdaBoost and CNN classifiers significantly boosted the sensitivity, but the overall performance was still limited (86.02%).

Kang *et al* [89] presented an automated identification of innocent murmurs in children using an artificial neural network and a support vector machine. The main focus of this work was to differentiate between the normal and pathological murmurs while classifying heart sounds in children. More recently, Beritelli *et al* [152] proposed a method to diagnose heart activity based on the Gram polynomials and probabilistic neural network. Though the proposed method achieved an overall accuracy of 94%, information regarding the types of abnormality under investigation and segmentation process was not found in the study.

Attempts to classify heart sounds using a modified classifier or rule-based classifier were also investigated. Decision trees classifiers based on certain rules and decision nodes were considered as a rule-based algorithm [170]. Pavlopoulos *et al* [171] proposed a decision tree algorithm that successfully classified fundamental heart sounds, the aortic stenosis and mitral regurgitation. Although the proposed algorithm achieved promising results, other heart sounds (such as artifacts) were not considered in the study. Later, Chen *et al* [183] suggested continuous wavelet transform based features and a regression tree (CART) based classification

algorithm. Although the work was intended to classify murmurs, specific types of murmurs were not investigated in the study. Safara *et al* [131] found that the entropy of wavelet packets as one of the feasible characteristics to classify four common murmurs (aortic regurgitation, mitral regurgitation, aortic stenosis and mitral stenosis). The study investigated the performance of the five different classifiers (SVM, decision trees, *k*-nearest neighbor, BayesNet and multi-layer perceptron networks) and the best results were achieved by a BayesNet classifier (96.94%). Other than these attempts, machine learning based approaches [54], [144], [155], Gaussian-Bayes model [172], Naïve Bayes [173], Gaussian Mixture Model (GMM) [174], random forest [173], and discriminant analysis [106], were also found in the literature. Appendix E summarizes the above-discussed heart sounds classification approaches and corresponding performance matrices are reported.

1.4 State-of-the-Art Cardiac System Monitors

A review of various cardiovascular monitoring systems was conducted to understand the evolution of cardiac monitors over the years and the various techniques involved in their development. Automated continuous real-time physiological status monitoring of the cardiovascular system helps in early-stage detection of diseases. With considerable advances in the technology, many invasive and non-invasive cardiac monitors have been developed. The main objective of these devices is to interpret different types of signals obtained from the cardiovascular system. The type of signal (electrical, acoustic, optical and seismic signals) under investigation depends on the sensor deployed and the technique used for the analysis. In general, electrocardiography (ECG), phonocardiography (PCG), photoplethysmography (PPG), and seismocardiography (SCG) based systems were found and briefly discussed in this section.

Over the decades, ECG has been a very popular non-invasive method that measures the electrical activity of the heart over a period of time with the help of electrodes and produces graphical results in the form of electrocardiograms (EKGs) for interpretation. Careful examination of EKGs reveals many important characteristics of electrical activities of the heart that are helpful in the identification of cardiac conditions. ECG signals are pre-processed and then PQRST waveform is extracted for the analysis. Despite recent developments, methods for automated classification of the ECG signals are not well established. ECG machines are also cumbersome, and, in some situations, it is tough to establish proper contact between the electrodes and skin surface.

ECG devices can be classified as a function of their number of leads: 12-lead ECG (Alirus 12-Lead ECG, Alirus, USA or Nasiff CardioCard Mobile Resting ECG, Nasiff Associates Inc,

USA), devices with less than 12 leads (Alirus 5-Lead ECG, Alirus, USA, or PocketECG, Medicalgorithmics, Poland or QardioCore, Qardio Inc., USA or eMotion ECG Mobile, Mega Electronics Ltd., Finland or LifeWatch MCT, LifeWatch AG, Switzerland or MobilECG 1–7 lead recorder, MobileECG, Hungary), and single-lead ECG (iHealth single lead wearable ECG, iHealth Labs Inc., USA or AliveCor, USA known as Kardia or ECG Check, CardiacDesigns, USA or HeartCheck Handheld™ ECG, CardioComm Solutions Inc., USA or Zenicor ECG, Zenicor Medical Systems AB, Sweden or InstantCheck, DailyCare BioMedical Inc., Taiwan). ECG based monitoring devices are also available as biopatches (Zio Patch, iRhythm Technologies, Inc., USA or BioPatch by Zephyr Technology Corp., Medtronic, USA, or BodyGuardian Heart, Preventice Solutions, USA or Metria, Avery Dennison Medical Solutions, USA or AVIVO Mobile Patient Monitoring system by Corventis), chest-strap and wrist-only models (chest-belt wireless ECG monitoring in real time, Suunto Quest, Polar RS300X, MyZone MZ-3, Polar H7) and other smart clothing (Hexokin, Carré Technologies Inc., Montréal or OMbra by OMsignal). Currently, flexible thin film ECG sensors (ultrathin flexible stretchable silicone membrane stick-on ECG device) are under examination to address shortcomings of the currently available commercial systems. Among these devices, most of the chest strap and wristbands such as Polar H7 Heart Rate Sensor, Polar, Finland and HRM-Run, Garmin, USA, are designed to obtain instantaneous heart rate (HR) or heart rate variability (HRV).

Traditional Holter devices are one of the most popular ambulatory ECG monitors, and have an average dimension of 110×70×30 mm. These can keep a log of the heart's electrical activity using a series of electrodes for a period of 24-48 hours. However, Holter monitors are not suitable for long term monitoring and can be a burden for patients due to its design. Furthermore, the real-time interpretation of the recorded signal is not possible [184]. Recently, researchers have integrated the Holter device with data transmission platforms such as Human++ [185], MEDiSN [186] and PlaIMoS platform [184], that can help in real-time data monitoring, transmission, and analysis of the data. A heart patch called Human++, flexible stick-on device, developed by the researchers at the Interuniversity Micro-Electronic Centre (IMEC) is a variation of the Holter monitor [187]–[189]. This portable EKG device consists of a patch like flexible organic electronics and thin-film silicon electronics-based circuit board with a circuit arrangement to detect and transmit the EKG signals up to 10 meters.

Other developments towards a wearable cardiac monitoring include VitalPatch, a wireless patch sensor packed with tiny ECG electrodes, a 3-axis MEMS accelerometer, and a temperature-sensing thermistor with a battery life of 3-4 days [190], [191]. Another device is Zio Patch by iRhythm Technologies, a single-lead ECG monitor that adheres to the left pectoral region of the chest, can provide continuous observations up to 14 days with 57% more

significant clinical events than the conventional Holter [192], [193]. Similarly, NUVANT Mobile Cardiac Telemetry (MCT) system by Corventis Inc, USA, rebranded as SEEQ by Medtronic Inc, USA, is also available for the ECG signal analysis [194]. AliveCor, USA, known as Kardia, is a compact size (8.2 cm × 3.2 cm × 0.35 cm), lightweight single-lead real-time ECG device for monitoring heartbeats, diseases like arrhythmia, shortness of breath, dietary habits, and heart activities [195]. AVIVO mobile patient monitoring system, Corventis, is a wireless ECG sensor, with an accelerometer and an impedance monitor for automated insight into the health status of the ambulatory patients. This non-invasive device adheres to a patient's chest and provides an understanding of heart and respiration rates, levels of patient activity, and fluid assessment. Similarly, the EPI Mini, a smartphone-enabled ECG device [196], collects ECG readings from the user's fingertips. ECG Check by Cardiac Designs [197] and ViSi Mobile system by Sotera Wireless are other ECG-based wearable cardiac monitors. Furthermore, smart clothing such as Hexoskin by Carré Technologies Inc., Montréal, is a body-worn sensor that can assist in ECG cardiac system and activity monitoring [198]. Other ECG smart wearables include OMbra by OMsignal [199] that keeps the record of heart rate and other sophisticated parameters related to the cardiac system. CoVa monitoring system by toSense is also a body-worn sensor for monitoring congestive heart failure, chronic obstructive pulmonary diseases, and hypertension, by measuring thoracic bioimpedance and electrocardiogram waveforms [200].

Companies such as BioMedical, Cardionet, Lifewatch, Medicomp, MedNet, ScottCare and Medicalgorithmics provide mobile cardiac telemetry based on the data available from the aforementioned ECG devices. Comparative studies of such monitors have been already reported in [201], [202]. ECG devices are efficient in monitoring cardiac system as explained above. However, some of the structural abnormalities such as defects in valves, are characterized by abnormal heart sounds and murmurs, which are difficult to trace with electrocardiographic methods.

Recently, non-invasive techniques such as Ballistocardiography (BCG), Seismocardiography (SCG) and Photoplethysmography (PPG) have been reported. BCG and SCG are mainly being explored to measure the mechanical activity of the heart using accelerometers to capture body micro-motions and to measure the recoil forces of the body and precordial vibrations caused by the ventricular contractions, respectively. These techniques don't utilize cumbersome wires and cables for the analysis. BCG measures whole-body recoil forces in response to the blood ejection into the vascular tree, while SCG measures the positional vibrations of the chest wall in reaction to the myocardial motions and respiration. However, so far, the applications of these techniques are limited to heart rate and heart rate variability measurements.

Photoplethysmography (PPG) is a non-invasive technique that can track the volumetric changes in the blood with an optical pulse wave monitors (pulse oximetry). Recorded PPG signals are used to determine heart rate, respiration, and cardiac cycles. The examples of these devices are such as Apple watch and Fitbit Charge. Apart from these devices, Philips Vital Signs Camera is a remote PPG device with standard/IR camera that senses tiny changes in the skin color for heart rate measurement and motion of chest and abdomen to determine the respiration rate [203]. The accuracy of PPG based devices depends on the type of the parameter under investigation and skin photosensitivity. Also, the pulse detection rate is largely depended on the light wavelength which can be influenced by the internal parameters such as skin pigmentation and melanin concentration.

Apart from these techniques, echocardiography (sonograms of the heart) also finds its application in high-resolution pocket mobile echocardiography (PME) [204]. However, echocardiography is not suitable for the overweight and lung patients. Additionally, non-invasive cardiac implantable electronic devices (CIEDs) such as implantable pacemakers, implantable loop recorders (Medtronic RevealXT, Reveal LINQ by Medtronic Inc, Dublin, Ireland [205], and Chronicle by Medtronic, Minneapolis, MN), are small USB sized long term rhythm monitors for a duration up to 3 years [206]. Though these devices are able to assist in continuous monitoring, invasiveness and stability in terms of the pull-out forces and the rotational stability of the device are the main concerns [207].

Recent studies suggest that automated cardiac sounds analysis can be used as a powerful tool to diagnose heart diseases. In general, auscultation using a stethoscope stands stalwart diagnostic method to identify cardiovascular diseases. However, this requires time, and experience to draw a conclusion from the phonocardiograms. With the advent of technology to capture heart sounds, self-care monitoring can be performed for primary screening of heart diseases. However, to achieve a robust, high sensitivity, and high yield system to classify heart sounds signals, intelligent algorithms are required.

1.5 Discussion

This chapter provided an overview of the current state-of-the-art in algorithms developed for the computerized heart sounds analysis and classification. Algorithms reviewed in this work investigated advanced signal processing tools and learning based approaches to automate the process. These algorithms were carefully evaluated to understand current challenges. Segmentation and classification of heart sounds were found to be still challenging, mainly because of the noise associated with the acquired signals that affected the quality of analysis.

Also, the complexity and non-uniformity associated with heart sounds signals were difficult to model.

Most of the segmentation approaches reviewed utilized adaptive threshold values of peak amplitudes, assumptions related to the systolic and diastolic intervals, and cardiac cycle period, to localize the peaks in the heart sounds signals. These assumptions are not valid for all kinds of subjects. Also, most of the segmentation algorithms fail in case the systole and diastole periods are of nearly equal duration. Thus, the error at the segmentation level may propagate to the next level of analysis. Only a few reviewed articles studied the noise tolerance while segmenting heart sounds signals [19], [41], [60], [78], [86], [95]. Recently suggested probabilistic models by Springer *et al* [70] and Schmidt *et al* [34] achieved good segmentation accuracy even for noisy signals.

It was found that the identification of fundamental heart sounds S_1 and S_2 achieved higher accuracy compared to the pathological sounds' identification during the segmentation process: (94.54±5.15%) and (93.96±5.01%), respectively, versus (88.50±5.93%) [208]. While most of the articles identified S_1 and S_2 heart sounds at the event level, the identification of S_1 sounds achieved higher accuracy compared to the S_2 sounds.

Among the articles reviewed in this thesis, only a few articles aimed to identify pathological heart sounds at the event level. These articles include detection of S_3 ([72], [95]), S_4 ([72], [95]) and murmurs ([19], [74], [86], [95]). The Hilbert-Huang Transform (HHT) was suggested for the identification of S_3 and S_4 [72]. However, the selection of intrinsic mode functions (IMFs) required in the model was challenging because of the varying temporal-spectral characteristics of heart sounds.

More recently, the Empirical wavelet transform was also suggested as a decomposition approach to segment heart sounds and to detect S_1, S_2, S_3 and murmurs [95]. Though these results appear to indicate that pathological sounds can be identified at the event level during the segmentation, most of the studies performed classification to diagnose pathologies. A few studies also suggested to segment cardiac signals directly into cardiac cycles rather than identifying the peak locations [45], [56], [91], [102], [103]. In these approaches, initial localization of S_1 and S_2 was skipped if prior knowledge of cardiac cycles was available.

From the data summarized in Appendix D, it can be found that the different characterization measurements/features are evaluated based on the classification problems. Statistical features are mostly extracted to identify the fundamental heart sounds and systole and diastole intervals in a cardiac cycle. Additionally, morphological, spectral, perceptual, fractal features, wavelet features, higher-order statistics and other time-varying and time-frequency domain discriminative features are recommended to distinguish pathological sounds. These features

take into account the dynamics of heart sounds under pathological conditions. Most of the features yielded promising results for classification between normal and abnormal heart sounds. Only a few reviewed articles reported features to identify particular cardiac pathologies which are discussed in the following paragraphs.

Reported characterization measurements are extracted using various heart sounds signal transformations and decompositions suggested. Among them, wavelet-based decomposition and reconstruction methods to obtain signal characteristics in both, time and frequency domains, for feature extraction were suggested in most of the recent articles [56], [130], [131]. The coefficients of mother wavelet transform are also evaluated as promising features. Some articles presented a comparative study of mother wavelets and suggested continuous wavelet transform using a Morlet wavelet as a potential transformation to extract features for detection of cardiac abnormalities - such as S_3, S_4, aortic stenosis, mitral regurgitation, midsystolic click, ventricular septal defect, atrial septal defect, mitral stenosis and aortic regurgitation [69], [114]. While others suggested the Daubechies wavelet for heart sounds analysis [45], [74], [149], [209].

Mel-Frequency Cepstral Coefficients (MFCCs) and Mel-Frequency Spectral Coefficients (MFSCs) have yielded promising results, compared to time-domain and short-time Fourier transform based features [116], [117], [129], in classification of fundamental heart sounds, S_3, S_4, ejection click, opening snap and diastolic and systolic murmurs [117]. However, MFCCs are not efficient in murmur classification with large energy lobes [101]. Hence, in addition to time-frequency domain features (such as STFT, wavelet transform, *etc*.), perceptual features (such as MFCCs), non-linear and chaos based features (such as recurrence quantification analysis and higher order statistics) and fractal features (such as correlation dimension, Largest Lyapunov Exponent and Hurst exponent) are recommended for the identification of valve disorders [58], [67].

Other features included multi-fractal spectrum [210], that achieved 96.91% accuracy in identifying prolapsed mitral valve; and multi-level basis selection [143] which yielded 97.56% accuracy in identification of aortic stenosis, mitral insufficiency, and atrial insufficiency. Similarly, along with time-domain based features, the center of gravity and the width of the frequency distribution extracted using a moving windowed Hilbert transform, reported up to 98.40% accuracy for identification of ventricular septal defects. Identification of systolic ejection click using spectral analysis with a time growing window also reported promising results (97.00% accuracy). Other than this, instantaneous frequency and amplitude of decomposed signal were found to be useful for the identification of splitting of fundamental heart sounds [16].

Reviewed articles also suggested other features extraction methods - partial least squares regression method [45], matching pursuit based methods [61], sparse coefficient matrix [127] and multivariate matching pursuit [57] - for which extracted features achieved promising results in classifying normal and abnormal heart sounds. In general, temporal, statistical, wavelet coefficients, spectral and instantaneous amplitude, and frequency-based features were extracted for abnormality detection.

Most of the recent studies classified pathological heart sounds using learning-based approaches (Artificial Neural Network (ANN) or Support Vector Machine (SVM) or ensemble learning). Articles also suggested modified support vector machines (such as Genetic SVM (G-SVM) [145], Least-Square Support Vector Machine (LS-SVM) [108], [211], Growing Time Windows based Support Vector Machine (GTSVM) [139], Support Vector Machine and Modified Cuckoo search (SVM-MCS) [85]), and validated the classification performance of SVM using different kernel functions for the identification of normal and pathological sounds. It was found that the Gaussian Radial Basis Kernel Function (GRKF) produced the best results in classifying normal, aortic stenosis, pulmonary stenosis tricuspid insufficiency and mitral insufficiency heart sounds compared to Linear Kernel Function (LKF), Polynomial Kernel Function (PKF) and Sigmoid Kernel Function (SKF)) [77]. Also, the least-square support vector machine (LS-SVM) classifiers were found promising in identifying normal, valvular defects, septal defects and other defects [132], with Morlet wavelet kernel function. Least square SVM was also suggested for identifying cases of chronic heart failure [135]. This achieved similar results to the back-propagation artificial neural network (BP-ANN) and hidden Markov models (HMM) and required less training time compared to its counterpart [147]. Other classification approaches such as decision trees [171], were also reported for the classification of fundamental heart sounds, aortic stenosis and mitral regurgitation. However, these methods are not suitable for complex feature classification [102].

Most of the studies reported methods to identify murmurs as systolic or diastolic murmurs. However, the classification of these murmurs into various sub-classes was not found in general. In a recent study, it was found that wavelet-based features and coefficients such as entropy, achieved promising results using a decision-based classification algorithm in classifying murmurs into systolic murmurs (early, mid, late, pan) and diastolic murmurs (early, mid, late, pan) and continuous murmurs [95]. Murmurs of valvular defects, mainly because of the stenosis, regurgitation, and insufficiency, were mostly investigated. Wavelet transformation and wavelet coefficients such as entropy, were found useful to classify normal, aortic stenosis, aortic regurgitation, mitral stenosis and mitral regurgitation [95], [134], [143]. Another method based on murmur likelihood computation and SVM classifier was found useful in classifying normal, aortic stenosis, mitral regurgitation, ventricular septal defect, aortic regurgitation,

mitral stenosis, and mitral valve prolapse [73]. SVM classifier based approaches were also extended to diagnose ventricular septal defects and atrial fibrillation and achieved satisfactory results in abnormalities classification [47], [125]. Other diagnostic heart sounds such as gallop rhythm were also classified using an SVM based classifier after pre-processing signals using the optimum multi-scale wavelet packet decomposition [100].

Other than these sounds, splitting of fundamental heart sounds has also been identified as a pathological event. While a limited number of articles investigated the splitting of second heart sounds (S_2) at the event level, no quantitative measurement of splitting of the first heart sounds (S_1) was found. The split identification was found to be obscured mainly because of the overlap of the components (M_1 and T_1 of S_1 and A_2-P_2 of S_2). In the articles reviewed in this work, it was not possible to ascertain the accuracy level in detecting the splitting of fundamental heart sounds due to the lack of articles available and the lack of quantitative analysis. Apart from these diagnostic sounds, a large amplitude of S_3 or S_4 and the presence of extra peaks in the cardiac cycle may reflect valvular malfunctioning or abnormalities, but these have not been investigated.

Although existing approaches reported promising results, algorithms were specifically developed for identification and classification of certain types of pathological sounds. In some cases, the accuracy of the algorithm was greatly dependent on the disease being investigated. Thus, these results cannot be interpolated to analyze other heart sounds that may be present in a cardiac cycle. Other than this, in some of the studies, the class of murmurs was not specified.

Data acquisition systems and databases used by the reviewed studies were also examined, coming to the conclusion that databases available for the validation of the algorithms are limited. In addition, demographics of the subjects and protocols followed when performing signal acquisition were not always fully specified. Sensors locations were also generally missing. Only a few studies validated the proposed algorithm with a database containing normal and abnormal heart sounds [85]. As most of the algorithms for the heart sounds analysis were validated with limited duration of recordings, the performance of these algorithms is not statistically significant. Thus, the robustness of algorithms still needs to be validated using large databases and with signals obtained from different subjects populations, including wider age ranges, and in real use scenarios. This is even more important considering that heart sounds are very sensitive to noise and interference, and different databases show different levels of data corruption. Furthermore, libraries of auscultatory recordings containing sounds signals from all possible auscultation sites from different subjects have not been reported. The analysis should be extended to test the robustness of the algorithms against the placement of the sensor (auscultation positions) while performing the signal acquisition. In relation to this, acquisition systems and noise reduction techniques should be developed in parallel, since different

acquisition systems respond differently to artifacts, which consequently might affect the performance of specific noise reduction algorithms.

Overall, existing algorithms show satisfactory results in classifying heart sounds in controlled conditions. However, it is not possible to extrapolate from this how they would operate in long term continuous monitoring of signals in real-life environments, mostly when subject-specific training is not an option.

2 Materials and Methods

The following chapter provides a detailed description of the datasets and methods used to facilitate and evaluate the performance of the approaches presented in Chapters 3, 4, 5 and 6. The four different studies present machine-learning based approaches to try to address the challenges in the automatic classification of cardiac recordings from different conditions.

2.1 Heart Sounds Datasets

As part of this work, signals acquired during a clinical study, from subjects with and without cardiac disease, in real-world, non-controlled environments were used. The study was originally approved by the Ethics Committee of All India Institute of Medical Sciences (AIIMS), Patna, India (approval number: AIIMS/Pat/IEC/2019/375). Written consent was obtained from all the subjects before physiological signals were recorded by trained medical personnel in the hospital. A total of 11 healthy subjects and 10 cardiac patients (14 males and 7 females), with average age of 38 ± 19 years old and weight of 54 ± 9.84 kg, participated in the experiments. Among the 10 cardiac patients, two patients were suffering from atrial fibrillation (AF), one from coronary artery disease (CAD), one from rheumatic heart disease (RHD), two from mitral regurgitation, one from bicuspid valve murmur, one from aortic stenosis, one form heart palpitations, and one from valvular heart disease. The protocols used for data acquisition were designed to address two major issues in the development of the algorithm. Firstly, for each auscultation experiment cardiac signals were recorded for a continuous duration of 5-minutes. This is in line with the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology recommendation [212] for evaluating clinical parameters such as heart rate and heart rate variability (HRV). Thus, the validation results obtained with this database could be extrapolated to evaluate the performance of algorithms developed for long term continuous monitoring of cardiac conditions. Secondly, to account for the variations in the signal characteristics, recordings were obtained using an electronic stethoscope [213] from each of the four possible auscultation sites. These included the aortic valve (between the 2nd and 3rd intercostal space (ICS) at the right upper sternal border), the mitral valve (between the 5th and 6th ICS near the apex of the heart), the pulmonary valve (between the 2nd and 3rd ICS at the left upper sternal

border) and the tricuspid valve (between 4th and 5th ICS at the left sternal border)). These auscultation sites are also shown in Figure 2.1.

All the recordings were carried out in uncontrolled environments, in consecutive patients that were being monitored with ECG as part of the conventional diagnostic pathway in the hospital. Patients were not asked to follow any instruction whilst being monitored. The recordings were carried out using an electronic stethoscope from ThinkLabs (shown in Figure 2.2) [213]. As part of the normal patient pathway, they were also being monitored with a 12-lead ECG system from Contec Medical Systems Co. Ltd. (shown in Figure 2.3) [214]. The locations of *R*-peaks and *T*-end peaks in the simultaneously recorded ECG signals were used as a reference to label S_1 and S_2 sounds, respectively. Overall, a 20-minute recording (5-minute from each of the 4 sites) was obtained from each subject. In this work, this database is referred as RHSS database. The study designed to record acoustic signals is described in Appendix G.

In the study presented in Chapter 3, heart sound signals recorded (RHSS database) from 21 subjects with and without cardiac disease, were used to form a database of 16825 S_1 and 16808 S_2 signals. The locations of *R*-peaks and *T*-end peaks in the simultaneously recorded ECG signals were used as a reference to label S_1 and S_2 sounds, respectively. As the signals were acquired from four different auscultation sites, the created database takes into account the characteristic variations in S_1 and S_2 with respect to the auscultation sites. In addition to the created database of S_1 and S_2 sounds, the datasets A and B of the PASCAL database available from the Classifying Heart Sounds Challenge (CHSC) [23] was also used in this study. This

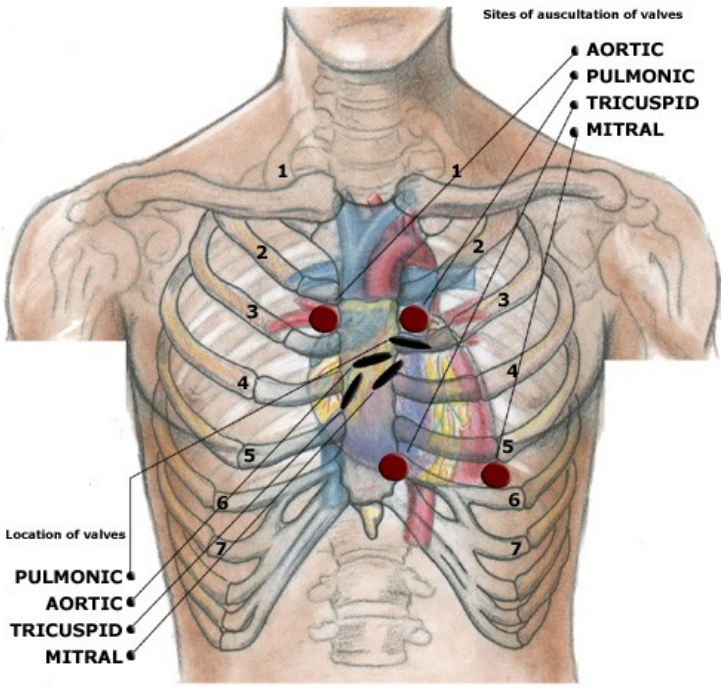


Figure 2.1. Location of heart valves and sites of auscultation [285].



Figure 2.2. An electronic stethoscope from the ThinkLabs used in this study to record cardiac acoustic signals.



Figure 2.3. A 12-lead ECG system from Contec Medical Systems Co. Ltd., used in this study to record ECG signals.

contains 21 and 90 recordings, respectively, from healthy subjects, with marked locations of S_1 and S_2, which were used to extract 753 S_1 and 736 S_2 segments. Overall, the combined database used to validate algorithm presented in Chapter 3 contained 17578 S_1 and 17544 S_2 sounds.

In the study presented in Chapter 4, the results obtained were evaluated using recordings obtained from the PASCAL database available from the Classifying Heart Sound Challenge (CHSC) database [23]. 231 recordings (31 from the Dataset A and 200 from Dataset B, of the PASCAL database) with normal fundamental heart sounds - S_1 and S_2; and 19 recordings labelled as "extra heart sounds" in the Dataset A, with additional sounds (i.e. a "lub-lub dub" or a "lub dub-dub"), either in the systolic or diastolic interval in the cardiac cycle, at regular intervals, were used. Additionally, 40 recordings contained artifacts, that were used to validate the performance of the developed algorithm in closer to real life applications. Information related to the category of recordings available from the database was used to create labels. Recordings in the Dataset-A had been originally collected using the iStethoscope Pro iPhone app, while recordings in the Dataset-B were collected using a digital stethoscope DigiScope

[23]. It was noted that the recordings in the Dataset B were superimposed with substantial amount of background noise or distortion as they were collected from a clinical trial in hospitals [23]. Among the 231 recordings with normal fundamental heart sounds, 111 recordings marked with the exact locations of S_1 and S_2 sounds were used to study the characteristics of fundamental heart sounds. This helped in identifying potential features to assist the classification of normal and abnormal recordings.

Recorded Heart sounds signals (RHSS database) from 21 subjects, labelled as normal and abnormal, were used in the study presented in Chapter 5. Due to the limitation of recording devices, there was a variation of ±30 seconds in the lengths of the recordings. Also, in some instances, no signal was captured during the recordings, resulting on these being removed from the analysis. Overall, 684 segments created from the RHSS database, each of 30 seconds window, were used to validate the algorithm. It is to be noted that the electrocardiogram (ECG) signals recorded in parallel to cardiac sounds using a 12-lead ECG system from Contec Medical Systems Co. Ltd. [214] were used as a reference to label S_1 and S_2 sounds using locations of *R*-peaks and *T*-waves, respectively, as reported in [215]. Additionally, recordings from three well-known public databases were also obtained to evaluate the performance of the presented approach by comparing results achieved in previous works using them. These included the PhysioNet Computing in Cardiology (CinC) Challenge 2016 database [22] (referred as Database1 in this study), the PASCAL database available from the Classifying Heart Sound Challenge 2011 (CHSC) [23] (referred as Database2 in this study), and a public database with recordings collected from different sources available from [216] (referred as Database3 in this study).

The PhysioNet database (Database1) contains recordings, labelled as normal and abnormal, from subjects with both normal and pathological conditions (such as murmurs related to mitral valve prolapse (MVP), aortic diseases, miscellaneous pathological conditions (MPC), coronary artery disease (CAD), mitral regurgitation (MR), and aortic stenosis (AS)). In case of the PASCAL database (Database2) and Database3, recordings with different pathological conditions were grouped together and labelled as the abnormal class. For instance, recordings categorized as murmur, extra heart sounds, and extrasystole in the PASCAL database were labelled as abnormal class. Similarly, mitral stenosis (MS), mitral regurgitation (MR), mitral valve prolapse (MVP), and aortic stenosis (AS) recordings from the Database3 [216] were labelled as abnormal class. Overall, 3153 recordings from the PhysioNet database (Database1), 585 (391 normal and 194 abnormal) from the PASCAL database (Database2), and 1000 (200 normal and 800 abnormal) from the Database3, were used in the study presented in Chapter 5. The length of recordings from the PhysioNet CinC database varied from 5 seconds to 122

seconds [8], whereas the recordings from the PASCAL database varied from 1 to 30 seconds. The length of recordings from the Database3 was limited to 3 cardiac cycles [216].

For the work presented in Chapter 6, heart sounds recordings were obtained from two public databases:

1) A public database reported in [216]. This database contains 1000 recordings from normal (200) and abnormal (800) subjects. The abnormal recordings include subjects suffering from aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP). 200 recordings were collected for each class and stored in wav format. These recorded signals were sampled to 8000 Hz. Each recording contains three period heart sounds signals or three cardiac cycles [216]. In this study, this dataset has been represented as DB1.

2) The PASCAL database available from the Classifying Heart Sound Challenge (CHSC) [23]. 65 recordings (31 normal and 34 murmurs) from the Dataset-A, and 415 recordings (320 normal and 95 murmurs) from the Dataset-B of the PASCAL database were used in this study. The recorded signals in the Dataset-B were superimposed with substantial amount of background noise or distortion as they had been collected from a clinical trial in a hospital [23]. Out of 415 recordings in Dataset-B, 149 recordings (120 normal and 29 murmurs) are marked as noisy recordings in the PASCAL database. Because of this, for this study the original datasets were re-labelled, as two different datasets, DB2 and DB3. First, clean normal and abnormal recordings (represented as DB2); and second, clean and noisy, normal and abnormal recordings (represented as DB3). Overall, DB2 contained 331 recordings (97 murmurs and 234 normal recordings without noisy signals), and DB3 contained 480 recordings (129 murmurs and 351 normal recordings with noisy signals) from the PASCAL database (both Dataset-A and Dataset-B). Information related to the category of recordings was used to label recordings.

It is worth nothing that these datasets are the ones that have been conventionally used in literature to test algorithms. However, most studies have used one or the other, not both [216] or [23]. The method presented in this work have been tested in both, since taking into account feature variations or changes in the characteristics of the recordings, especially under varying environmental conditions and recording devices used to collect data. This makes the results more generalizable when taking into account real life conditions. Thus, the validation results obtained were significant, addressing key issues in the development of algorithms for automatic detection and classification of recordings in real-world conditions.

2.2 Feature Extraction & Selection

2.2.1 Statistical time and frequency domain features

The characteristics of fundamental heart sounds in normal recordings were used to identify potential features for the classification. It was found that any physiological alteration in the cardiovascular events associated with the contraction and relaxation of the atria and ventricles influences the occurrence of S_1 and S_2 sounds, since these are produced due to the closure of atrioventricular and semilunar valves, respectively. Subsequently, the duration of the systolic interval (starting at the beginning of S_1 and lasting until the beginning of S_2), as well as the diastolic interval (starting at the beginning of S_2 and lasting until the beginning of S_1) may vary significantly as a function of different cardiovascular conditions and so will the sound generated in the process [217]. For instance, authors in [218] demonstrated the variation in the duration of S_1 and S_2 sounds for different pathological conditions. This is the reason why statistical time and frequency domain features were considered in this work. From now onwards P will refer to cardiac peaks in the sounds, the subscript *1* to characteristics associated to S_1 and the subscript *2* to characteristics associated to S_2. The features extracted were: the mean of the absolute value of P_1 sound intervals (P_{1T}), the mean of the absolute value of P_2 sound intervals (P_{2T}), the ratio of subsequent P_1 and P_2 intervals (P_{1T}/P_{2T}), the interval between the beginning of subsequent P_1 and P_2 sounds ($T_{P1\text{-}P2}$), the interval between the beginning of P_2 and the next P_1 sounds ($T_{P2\text{-}P1}$), the interval between the two adjacent P_1 sounds ($T_{P1\text{-}P1}$), the interval between the two adjacent P_2 sounds ($T_{P2\text{-}P2}$), the ratio of $T_{P1\text{-}P2}$ to $T_{P2\text{-}P1}$ intervals ($T_{P1\text{-}P2}/T_{P2\text{-}P1}$), the ratio of $T_{P1\text{-}P2}$ interval to P_1 interval ($T_{P1\text{-}P2}/P_{1T}$), the ratio of $T_{P2\text{-}P1}$ interval to P_2 interval ($T_{P2\text{-}P1}/P_{2T}$). These statistical features are also shown in Figure 2.4. Here, P_1 and P_2 represent the first and second sound peaks of interest localized in a recorded signal, occurring in an alternate manner, respectively. As the algorithm might have detected sound peaks that were neither S_1 nor S_2 sounds, P_1 and P_2 were used to indicate the peaks so that the features extracted were independent of the accurate identification of S_1 and S_2 sounds. Thus, a total of 10 statistical features were extracted for each analysis window using the localization approach demonstrated in Chapter 3.

As shown in Figure 2.4, the zero crossover points of sound peaks in the Shannon energy envelope were used to compute the start and end locations of sound peaks, P_1 and P_2. These locations were stored in two arrays (P_{12S} and P_{12E}) of length $2n$ (n is the number of all sound peaks localized). In array P_{12S}, odd locations represented the start of P_1 (denoted as P_{1S}); whereas even represented the start of P_2 (denoted as P_{2S}). Similarly, in array P_{12E}, odd locations represented the end of P_1 (denoted as P_{1E}); whereas even represented the end of P_2 (denoted as P_{2E}). Thus, the features were computed for the i^{th} iteration as follows:

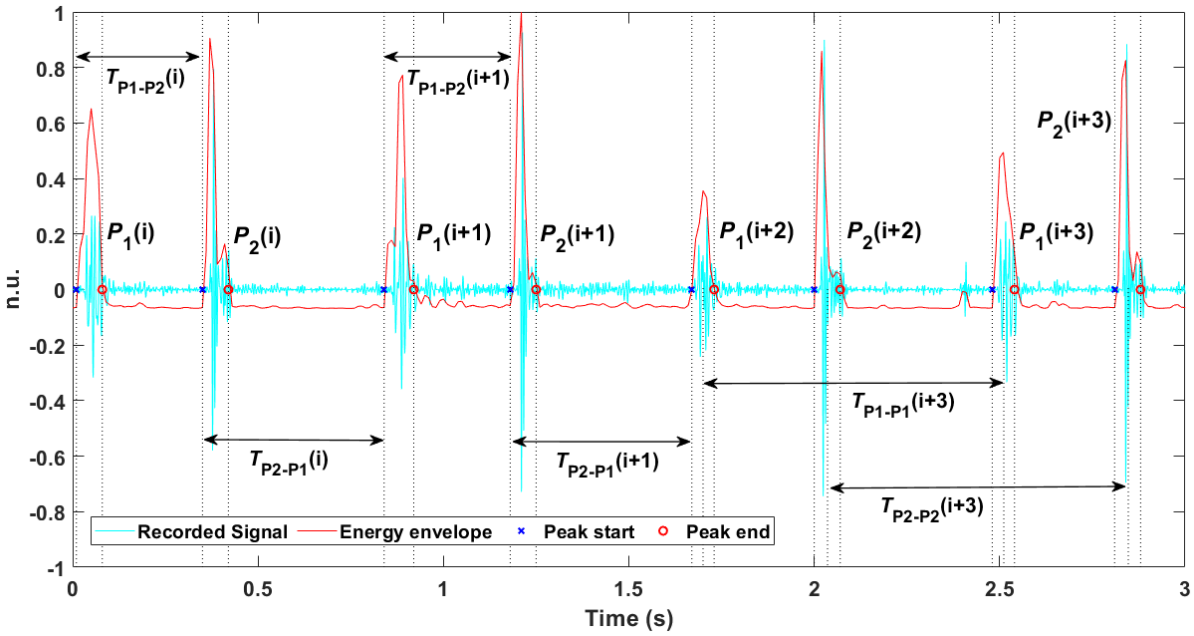


Figure 2.4. Various statistical features extracted from the localized sound peaks of interest in a recorded heart sounds signal from normal subject, from the PASCAL database.

$$P_{1T}(i) = P_{1E}\,(2i\ -\ 1)\ - P_{1S}\,(2i\ -\ 1) \tag{2.1}$$

$$P_{2T}(i) = P_{2E}\,(2i) - P_{2S}\,(2i) \tag{2.2}$$

$$T_{P1-P2}(i) = P_{2S}\,(2i) - P_{1S}\,(2i\ -\ 1) \tag{2.3}$$

$$T_{P2-P1}(i) = P_{1S}\,(2i\ +\ 1) - P_{2S}\,(2i) \tag{2.4}$$

Further, the midpoint of P_1 (denoted as P_{1M}) and P_2 (denoted as P_{2M}) were used to evaluate T_{P1-P1} and T_{P2-P2} for the i^{th} iteration as follows:

$$P_{1M}(i) = \frac{P_{1S}\,(2i\ -\ 1) + P_{1E}\,(2i\ -\ 1)}{2} \tag{2.5}$$

$$P_{2M}(i) = \frac{P_{2S}\,(2i) + P_{2E}\,(2i)}{2} \tag{2.6}$$

$$T_{P1-P1}(i) = P_{1M}\,(i+1) - P_{1M}\,(i) \tag{2.7}$$

$$T_{P2-P2}(i) = P_{2M}\,(i+1) - P_{2M}\,(i) \tag{2.8}$$

The statistical time-domain features that can potentially capture the amplitude variations, outliers and timing information, were extracted by representing the signal as a series of coefficients. Thus, in addition to these features, other statistical time-domain features evaluated for each analysis window of heart sounds signal ($h[n]$) were the signal's mean $\left(\overline{h[n]}\right)$ (F_{11}), the median (F_{12}), standard deviation (F_{13}), mean absolute deviation (F_{14}), quantile (F_{15} and F_{16}),

interquartile range (F_{17}), skewness (F_{18}), kurtosis (F_{19}), spectral entropy (SE) (F_{20}) [219], dominant frequency (f_D) (frequency corresponding to the maxima of power spectrum denoted as F_{21}), maximum frequency for each sound lobes (f_{max}) (F_{22}), and the ratio of dominant frequencies (F_{23}) and zero-crossings (F_{24}). These statistical features are described in Appendix F, where $h[n]$ and $\overline{h[n]}$ represent the signal and it's mean. N is the length of the signal. $H(i)$ is the discrete Fourier transform of $h[n]$. These features have been used for detection of atrial fibrillation [220].

2.2.2 Deterministic and non-deterministic energies

A cardiac cycle consists of a series of deterministic and non-deterministic events. Among them, fundamental heart sounds (S_1 and S_2) are considered as deterministic (due to their cyclic nature [42]), while abnormal sounds, such as murmurs, that may be present due to the different cardiac abnormalities are considered as non-deterministic events (or stochastic events) [221]. For instance, abnormalities such as structural defects or malfunctioning of heart valves result in non-deterministic events in a phonocardiogram of a recorded signal [221]. Depending on the pathological condition of a subject, non-deterministic (stochastic) events may be present in any segment of cardiac cycle (i.e. systole or diastole or both) and potentially correspond to the cavitation, valvular defects, regurgitation and stenosis [221]. This is because the energies of deterministic and non-deterministic events were estimated to distinguish normal and abnormal recordings. As shown in Johansen *et al* [221], [222], the difference between the total energy ($Energy_{\text{total}}$) and the energy of the deterministic events ($Energy_{\text{det}}$) was used to compute the energy of non-deterministic events ($Energy_{\text{non-det}}$) for each analysis window (as shown in Figure 2.5).

$$Energy_{non-det} = Energy_{total} - Energy_{det} \quad 2.9$$

Two features F_{25} and F_{26} were extracted, where F_{25} was the deterministic energy of the fundamental heart sounds, while F_{26} was the total energy of the systole and diastole intervals (i.e. non-deterministic energy or stochastic energy) in a recorded signal. The locations of peaks were used to determine the intervals between the peaks. For the systole interval, this was evaluated as the energy of segment between the end point of P_1 and the start of P_2, expressed as: $P_{2S}(2i) - P_{1E}(2i - 1)$). Similarly, for the diastole interval, this was evaluated as the energy of the segment between the end of P_2 and start of P_1, expressed as: $(P_{1S}(2i + 1) - P_{2E}(2i))$.

2.2.3 Multiresolution analysis using a non-decimated transform

Although wavelet-based features have been extensively used in the past few years, there is

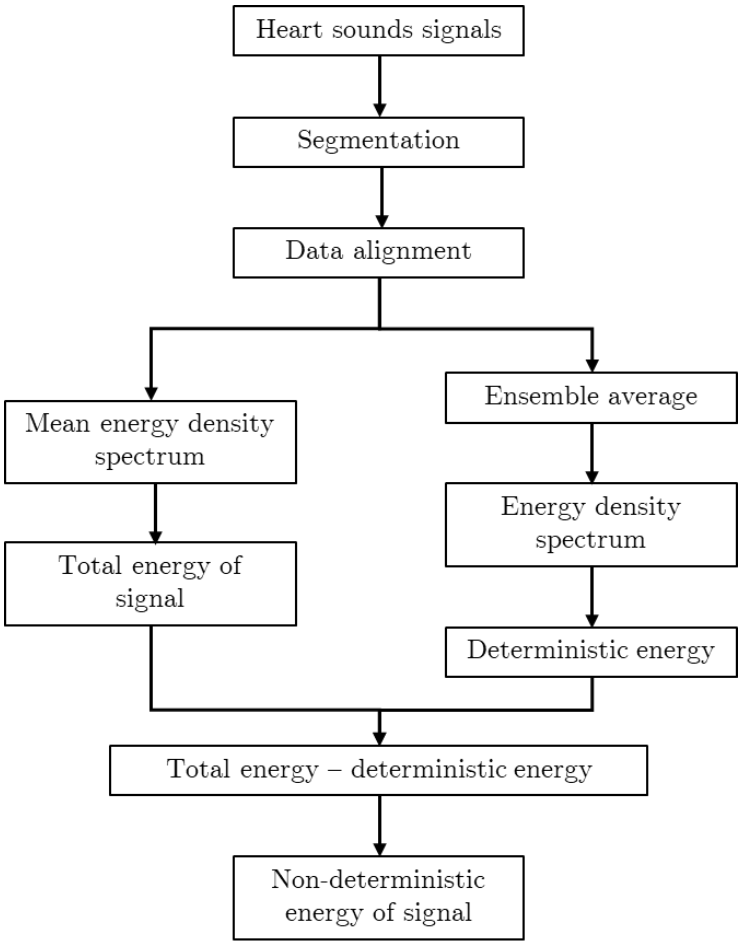


Figure 2.5. Flowchart for the non-deterministic energy calculation.

limited agreement on the potential wavelet features suitable for S_1 and S_2 identification and classification of cardiac acoustic signals. In this study, a variant of non-decimated transform i.e. Maximum Overlap Discrete Wavelet Transform (MODWT), was used for feature extraction as MODWT is well-defined for all sample sizes [223].

For a normalized signal $h[N]$ with N number of samples, MODWT gives $1 \times N$ scaling coefficients (C_S) for the final decomposition level and $n \times N$ detail coefficients (C_D), where n is the number of wavelet decomposition levels, given as $n = floor(\log_2 N)$. 11 features (denoted as F_{27}, …, F_{37}) were extracted, representing the energy across the scales and scaling coefficients. The energy of the signal can be computed for the signal $h[N]$ as:

$$\|h[N]\|^2 = \sum_{n=1}^{n_0} \left\|C_{D_n}\right\|^2 + \left\|C_{S_{n_0}}\right\|^2 \qquad 2.10$$

where, $n_0 = 9$, is the final level of decomposition. F_{37} represents the total energy of all the scales.

Furthermore, multiresolution analysis (MRA) of the MODWT matrix was used to analyse heart sounds at different frequencies with a different resolution by projecting the signal onto

various wavelet subspaces and a scaling subspace. The multiresolution analysis of the MODWT matrix at each level is time invariant. Thus, it facilitated the zero-phase filtering and time-aligned feature extraction [224]. This can also be observed from Figure 2.6, which shows the multiresolution analysis of the MODWT matrix obtained for 10-levels of decomposition of a recording.

Additionally, at a certain analysis level of multiresolution analysis, extracted coefficients can be used to distinguish S_1 and S_2. This can also be verified from the coefficients plotted in Figure 2.6(IV) and Figure 2.6(V), demonstrating the difference in coefficients of S_1 and S_2. Thus, 11 features (denoted as F_{38}, …, F_{48}) were extracted representing the energy across the scales for the multiresolution analysis of the MODWT matrix. F_{48} represents the total energy of all the levels. As per the knowledge gathered from the literature review, these features have not been explored before for fundamental heart sounds identification.

2.2.4 Homomorphic envelogram

The homomorphic envelogram, capturing various physiological attributes of acoustic signals, was extracted using a homomorphic filter as demonstrated in [34], [70]. Cardiac acoustic signals were expressed as a product of two components (the amplitude and oscillation). The log-transformed signal was passed through a low pass filter to separate the fast-varying components (or high-frequency components) representing the oscillations. In this work, a first-order Butterworth low-pass filter with cut off frequency 10 Hz was used to extract 3 features (F_{49}, F_{50} and F_{51}) representing peak amplitude, maximum frequency, and its corresponding index value, respectively, for each window. These features were extracted from the smooth homomorphic envelogram of recordings as shown in Figure 2.7.

2.2.5 Hilbert-Huang transform-based features

As heart sound signals are non-stationary and nonlinear with large variation in amplitude and frequency, Hilbert-Huang transform (HHT) is a powerful method for their analysis [72], [225] to capture their characteristics. The feature extraction using the HHT involves two steps: Empirical mode decomposition (EMD) and Hilbert transform. The EMD was used to decompose recordings into a finite number of intrinsic mode functions (IMFs) as shown in Figure 2.8(a). These extracted IMFs were used to compute the analytical signal using the Hilbert transform which provided the instantaneous frequency of the signal [51].

To obtain the first IMF ($\widetilde{h_1[n]}$), the average of two envelopes, constructed using the local maxima and local minima of the signal, was subtracted from the original signal ($h[n]$) [72]. The first extracted IMF was then subtracted from the original signal to obtain a new residual

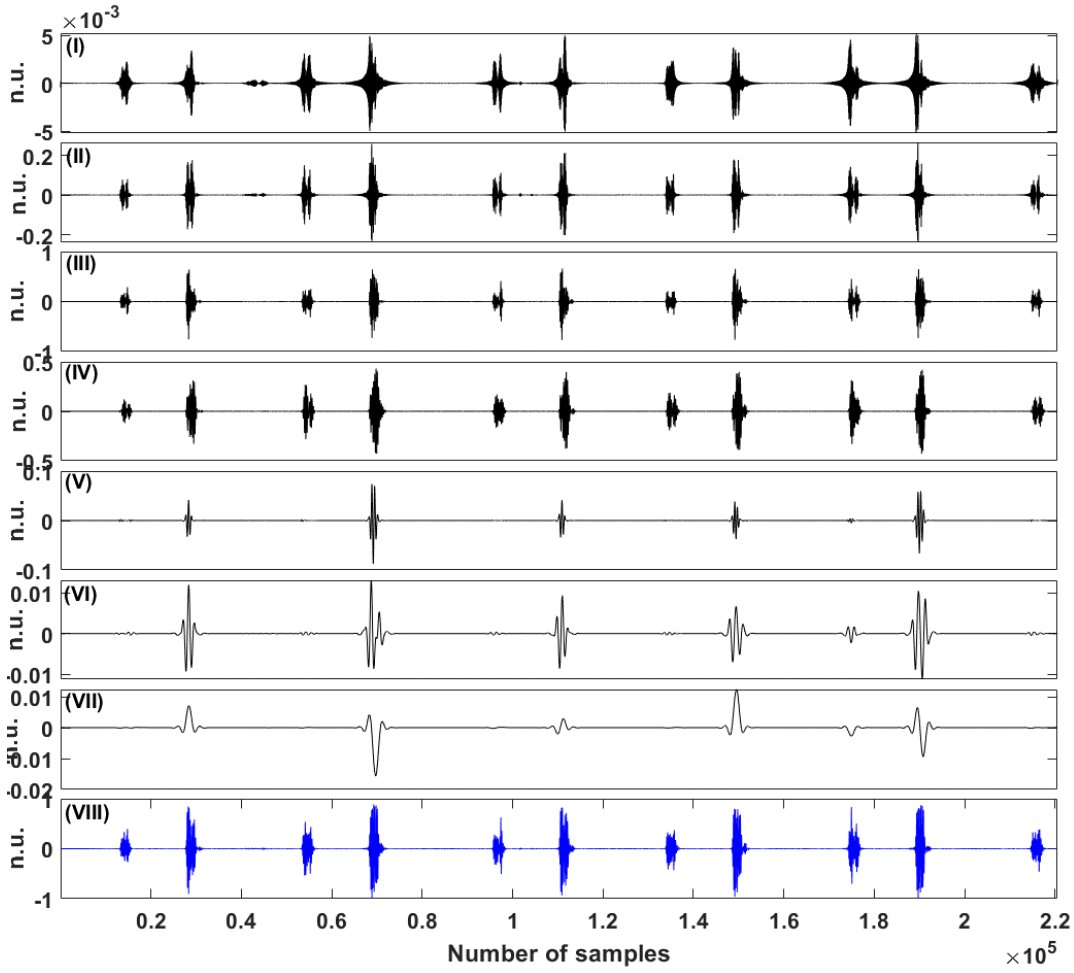


Figure 2.6. Multiresolution analysis of a recorded acoustic signal using the maximum overlap discrete wavelet transform (MODWT). Only meaningful multiresolution analysis (level 5 to 11) of the MODWT matrix coefficients are shown in (I)-(VII) for a recorded signal shown in (VIII).

signal $r_i[n]$, which was used as the original signal to obtain the next IMF. This step was repeated for *N* iterations, so that the original signal ($h[n]$) can be expressed as:

$$h[n] = \sum_{i=1}^{N} \widetilde{h_i[n]} + r_N[n] \qquad 2.11$$

The stopping criteria (S_{IMF}) used was:

$$S_{IMF} = \sum_{i=1}^{N} \frac{\left|\widetilde{h_{i-1}[n]} - \widetilde{h_i[n]}\right|^2}{\widetilde{h_i[n]}^2} \qquad 2.12$$

where, $\widetilde{h_i[n]}$ represents one of the IMFs changing from $\widetilde{h_1[n]}$ to $\widetilde{h_N[n]}$ [225]. In the next step, extracted IMFs (shown in Figure 2.8(a)) were used to compute the analytical signal ($h^*[n]$)

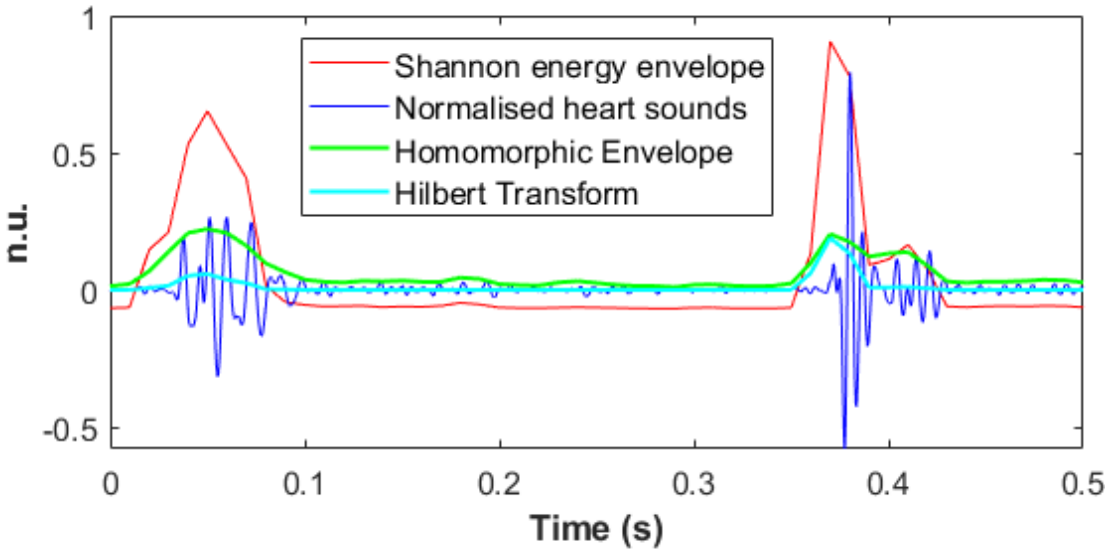


Figure 2.7. Different envelopes extracted for feature extraction of S_1 and S_2 sounds.

using the Hilbert transform, as follows [51]:

$$h^*[n] = h[n] + jH\{h[n]\} = h[n] + j\left[\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{h[\tau]}{\tau - 1}d\tau\right] \quad 2.13$$

$$h^*[n] = h[n] + j\left[h[n] \times \frac{1}{\pi n}\right] = HE[n]e^{j\theta[n]} \quad 2.14$$

where, $H\{h[n]\}$ represents the Hilbert transform as:

$$H\{h[n]\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{h[\tau]}{\tau - 1}d\tau = h[n] \times \frac{1}{\pi n} \quad 2.15$$

and $HE[n] = \sqrt{h^2[n] + H^2\{h[n]\}}$ is the envelope of $h[n]$ and the derivative of the instantaneous phase $\left(\theta[n] = \tan^{-1}\frac{H\{h[n]\}}{h[n]}\right)$ gives the instantaneous frequency $\left(d\omega[n] = \frac{d\theta[n]}{dn}\right)$ which represents the frequency at a particular time and is well suited for non-stationary signals analysis.

3 features (F_{52}, F_{53} and F_{54}) representing the peak amplitude, maximum frequency, and its corresponding index value in the spectrum, from the constructed Hilbert envelope for each window were extracted. 6 features (F_{55}-F_{60}) were the energy computed for the first 5 IMFs and the residual part; and 8 features (F_{61}-F_{68}) (2 features from each of the 4 selected IMFs (IMF2, IMF3, IMF4, and IMF5)) represented the maximum frequency and its corresponding index value in the spectrum. Further, a *k*-means cluster algorithm was used to cluster the instantaneous frequency and the instantaneous energy of different modes obtained from the Hilbert-Huang transform. This produced 8 features (F_{69}-F_{76}) from 4 different modes (IMF2, IMF3, IMF4, and

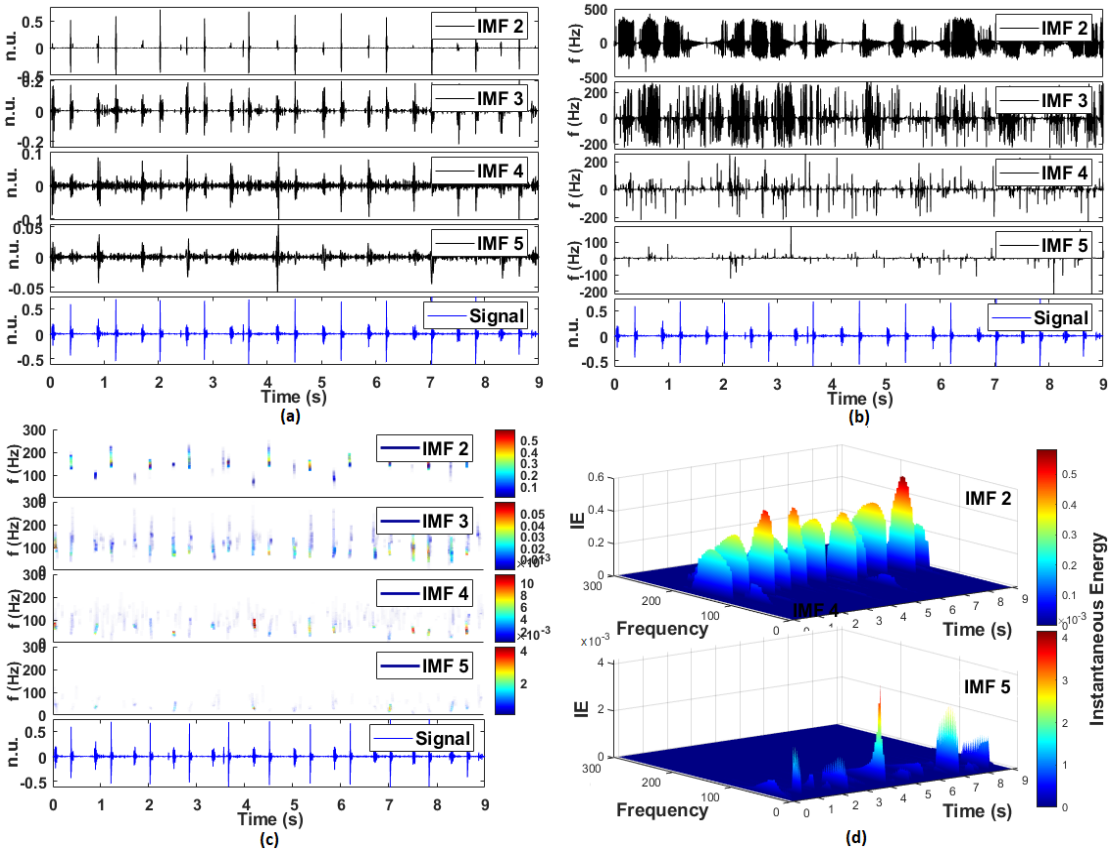


Figure 2.8. (a) Extracted 4 out of 10 intrinsic mode functions (IMF) of a heart sounds signal using EMD. (b) Intrinsic frequencies of the different IMFs. (c) Instantaneous energy (IE) of the different IMFs. (d) IE of IMF 2 and 5 to demonstrate the variation of the energy content with respect to the IMFs.

IMF5), by keeping the number of modes equal to the number of clusters in the algorithm. Here, IMF1 was eliminated as it corresponded to high-frequency components; and only IMF2, IMF3, IMF4, and IMF5, containing meaningful information, were selected for feature extraction. This is also because most of the physiological information corresponding to S_1 and S_2 was found concentrated in the frequency range of 20-300 Hz, present in these IMFs. This can be observed in Figure 2.8(b). Similarly, the instantaneous energy was computed for only 4 modes (IMF2, IMF3, IMF4, and IMF5) containing most of the energy that produced 8 features (F_{77}-F_{84}). Figure 2.8(c) and Figure 2.8(d) show the instantaneous energy content present in the different modes.

2.2.6 Perceptual features

Mel-frequency cepstral coefficients (MFCCs), that map the acoustic signal onto a mel-scale space, exhibiting logarithmic characteristics for high frequencies, were also extracted as

potential features [216], [226]. Segments were represented in parametric form through MFCCs computed by applying a discrete cosine transform on the signal obtained from a log-spaced filter bank. To extract features, the acoustic signal was pre-emphasized using a pre-emphasis coefficient (ALPHA) of 0.97. An analysis window of 0.02s and an overlapping window of 0.01s were used to extract 15 cepstral coefficients (represented as F_{85}-F_{99}) using 45 mel-scaled filters with a frequency passband of 5 Hz to 500 Hz. Each channel produced a set of MFCCs which were summed to extract one feature from each channel. The steps used to extract these features are shown in Figure 2.9.

2.2.7 Feature selection and reduction

Feature weights were computed using a diagonal adaptation of neighborhood component analysis (NCA), to derive a set of most significant features from all the features extracted, that could potentially best represent the signal. Neighborhood component analysis allocated weights to each feature by maximizing the expected leave-one-out classification accuracy [219]. NCA is a nearest neighbor-based feature selection method suitable for high-dimensional data analysis. Furthermore, authors in [227] demonstrated that NCA based feature extraction is almost insensitive to the increase in the number of non-essential features and results achieved were found promising compared to several nearest neighbor-based feature weighting approaches. For these reasons, in this work, NCA was used to select features from the set of features described in section 2.2. Features that corresponded to large feature weights (or above a specified threshold (δ)) were selected as optimum features.

2.3 Classifier Training and Hyperparameter Optimization

The approach is to develop a classifier using a training set that decodes the state of unseen test samples given their input features. The main problem with the supervised learning is that the

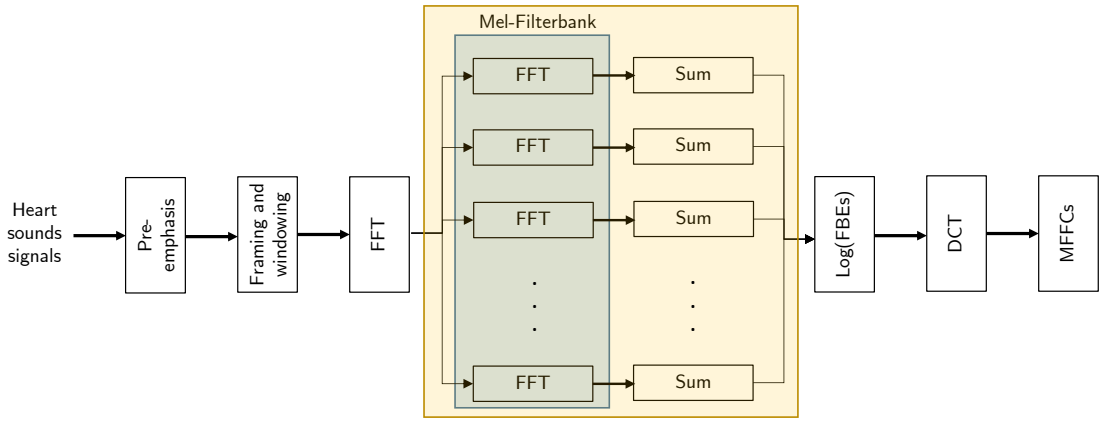


Figure 2.9. MFCCs extraction from heart sounds signals using a mel-filter bank.

dataset may not contain physiological measurement from varied population. This is really important as the characteristics of the signals vary widely from subject to subject. To address this, different databases were used in the presented studies, which included signals recorded in this study as well as signals available from different open sources.

2.3.1 Classifier training

Training a classifier using the statistics obtained from the true known states of the dataset is one of the ways to extract global parameters of the classifier that are robust in nature. In this study, a feature vector from $x^{(i)} \in R^n$ was used to train an ensemble classifier to interpret the class of $y^{(i)} \in R$; enabling the trained classifier to potentially decode the state of unseen test samples given their input features.

In this work, classifier mainly used for the heart sounds analysis, found during the literature review, were studied. It was found that the ensemble of classifiers are used to achieve improved classification accuracy by combining the prediction of several learners to derive a powerful model that performs better than any single classifier [68], [228], [229]. The ensemble learning approaches have been used in recent studies for successful classification of cardiac auscultatory recordings [68], [230], [231]. Furthermore, recent works on identification of S_1 and S_2 sounds using individual machine learning classifiers have demonstrated promising results [75], [232], [233]. This approach constructs a set of classifiers with different weights assigned to the weak learners using an adaptive ensemble algorithm and then classifies a new dataset by taking a weighted vote of their predictions. Thus, the weighted combination of multiple algorithms reduces the variance of predictions to achieve good performance and robustness. For these reasons, the ensemble machine learning approach was used to derive a powerful classifier model with the objective to achieve improved classification performance and robustness.

2.3.2 Hyperparameter optimization

To obtain optimized values of the key hyperparameters (learning rate, number of learning cycles, number of weak learners/trees and the maximum number of splits) considered for the learning algorithm, the misclassification rate (or loss function) was computed iteratively for different training sessions using 10-fold cross-validation, during the training stage. The cross-validated misclassification rates with respect to the number of trees, learning rate, and for a different number of splits were obtained iteratively for different training sessions. The values obtained from different training sessions have been reported in the following chapters.

3 Automatic Classification of S_1 and S_2 Sounds from Cardiac Auscultatory Recordings

The research presented within this chapter is an edited version of research article:

Amit Krishna Dwivedi, Syed Anas Imtiaz, Stuart Bowyer, and Esther Rodriguez-Villegas, "Automatic Classification of S_1 and S_2 Sounds from Cardiac Auscultatory Recordings", IEEE Access (2020) (under review).

3.1 Introduction

Segmentation is an important initial step in automated analysis of heart sounds. Segmentation involves automatic localization and identification of the fundamental heart sounds (first heart sound (S_1) and second heart sound (S_2) produced due to the mechanical activity of heart valves as shown in Figure 3.1), leading to the subsequent classification of the recorded acoustic signals.

A detailed discussion on recent works on segmentation, and subsequent identification and classification of S_1 and S_2 is presented in Chapter 1. Although some existing algorithms showed promising results in segmenting S_1 and S_2, these were validated using public databases as well as small datasets of recordings collected by researchers in a controlled environment. Hence, there are still important challenges when dealing with signals that are recorded in more representative, real conditions. One of these is how to separate the sounds of interest (S_1 and S_2) in raw cardiac auscultatory recordings that may also contain abnormal sounds- such as third heart sounds (S_3), fourth heart sounds (S_4) and murmurs associated with the malfunctioning of heart valves [50]. Another challenge is that they still need to target performance in real-world conditions, taking into account signals obtained in non-controlled environments, in which signals will also be more contaminated by artifacts. An additional issue in existing studies is the fact that in most cases the auscultation sites have not been reported [234]. This is important considering there can be significant variations in features extracted in signals recorded from different auscultation sites, that may lead to localization errors in learning-based approaches. For instance, S_1 possesses higher energy than S_2 in acoustic signals recorded from the apex or

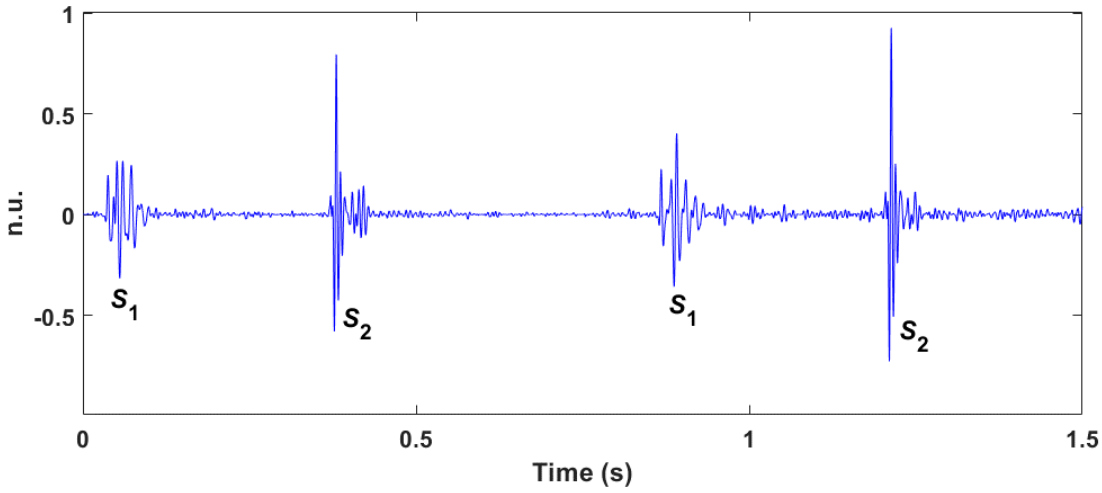


Figure 3.1. Fundamental heart sounds (S_1 and S_2) in a recording from healthy subject.

left lateral sternal border, while this is reversed for the signals recorded from the aortic or pulmonary auscultation site [235].

There were two main objectives of the work described in this chapter. The first was to find potential features that can help in the identification of S_1 and S_2 sounds from both normal and abnormal signals acquired from different auscultation sites. The second objective was to use these heart sounds features to develop an intelligent algorithm for the classification of S_1, S_2 and artifacts segments, in recorded signals. Both of these objectives ultimately aid the development of an algorithm that could be used as part of a wearable monitoring system.

3.2 Methods

This chapter presents a machine learning-based approach for automatic classification of fundamental heart sounds as shown in Figure 3.2. Recorded acoustic signals were pre-processed and sound peaks of interest were localized using the Shannon energy envelope. Later, segments of S_1 and S_2 were extracted among all the localized sounds and labelled using reference ECG signals. Overall, the database used in this study contained 17578 S_1 and 17544 S_2 sounds. The details of the individual databases, and corresponding number of S_1 and S_2 sounds extracted, are given in Table 3.1. An informative feature set was extracted from both labelled S_1 and S_2 sounds. Finally, selected features were used to train a classifier to establish a classification boundary. All the steps involved are discussed in detail in the sections below.

3.2.1 Pre-processing of recorded signals

Phonocardiograms of heart sounds signals recorded in real-world conditions are often

contaminated with noise and artifacts, hence, it is challenging to interpret recordings. Low-frequency noise signals in the spectra overlap with the fundamental heart sounds make the analysis complex. The presence of these noise components in the acquired signals are usually high, hence, a reliable method needs to be investigated to eliminate the internal and external noise content and artifacts.

A recorded heart sounds signal (HSS(n)) can be expressed as [217]:

$$HSS[n] = FHS[n] + EHS[n] + N[n] \quad 3.1$$

where, $FHS[n]$ represents fundamental heart sounds (S_1 and S_2), $EHS[n]$ the extra heart sounds (i.e. pathological sounds such as murmurs, S_3 and S_4), and $N[n]$, the noise/interference present in the recording. As most of the pathological and physiological information related to S_1 and S_2 is present in the frequency band below 500 Hz [99], a low pass 4th order Butterworth filter was used to eliminate high-frequency components in recorded signals. Recordings were then normalized to a fixed scale of [-1, 1] as follows:

$$h[n] = \frac{h'[n]}{\max_n(|h'[n]|)} \quad 3.2$$

where, h[n] is the normalized signal and $\max_n(|h'[n]|)$ represents the absolute maxima of the filtered heart sounds signal ($h'[n]$). The pre-processed and filtered recorded heart sounds signal is shown in Figure 3.3.

Further, to isolate components of $FHS[n]$, a wavelet-based adaptive sub-level tracking approach was used [236]. This is because traditional filtering approaches mainly involve decomposition of signals into successive time series or at different frequency bands using the Fourier transform and, hence, lack the ability to localize events occurring over multiple time

Table 3.1: Details of heart sounds signals used in this work.

Source		Subjects	No. of Recordings	S_1 and S_2 sounds extracted
The PASCAL database [23]	Dataset-A	Normal	21	178 S_1 and 176 S_2 from Dataset A
	Dataset B	Normal	90	575 S_1 and 560 S_2 from Dataset B
Recorded heart sounds signals (RHSS database)		11 normal and 10 abnormal (14 M and 7 F)	21	4217 S_1 and 4209 S_2 from AV
				4289 S_1 and 4285 S_2 from TV
				4171 S_1 and 4163 S_2 from MV
				4148 S_1 and 4151 S_2 from PV

*AV, TV, MV, and PV represent auscultation sites for the aortic valve, tricuspid valve, mitral valve, and pulmonary valve, respectively. M: male and F: female.

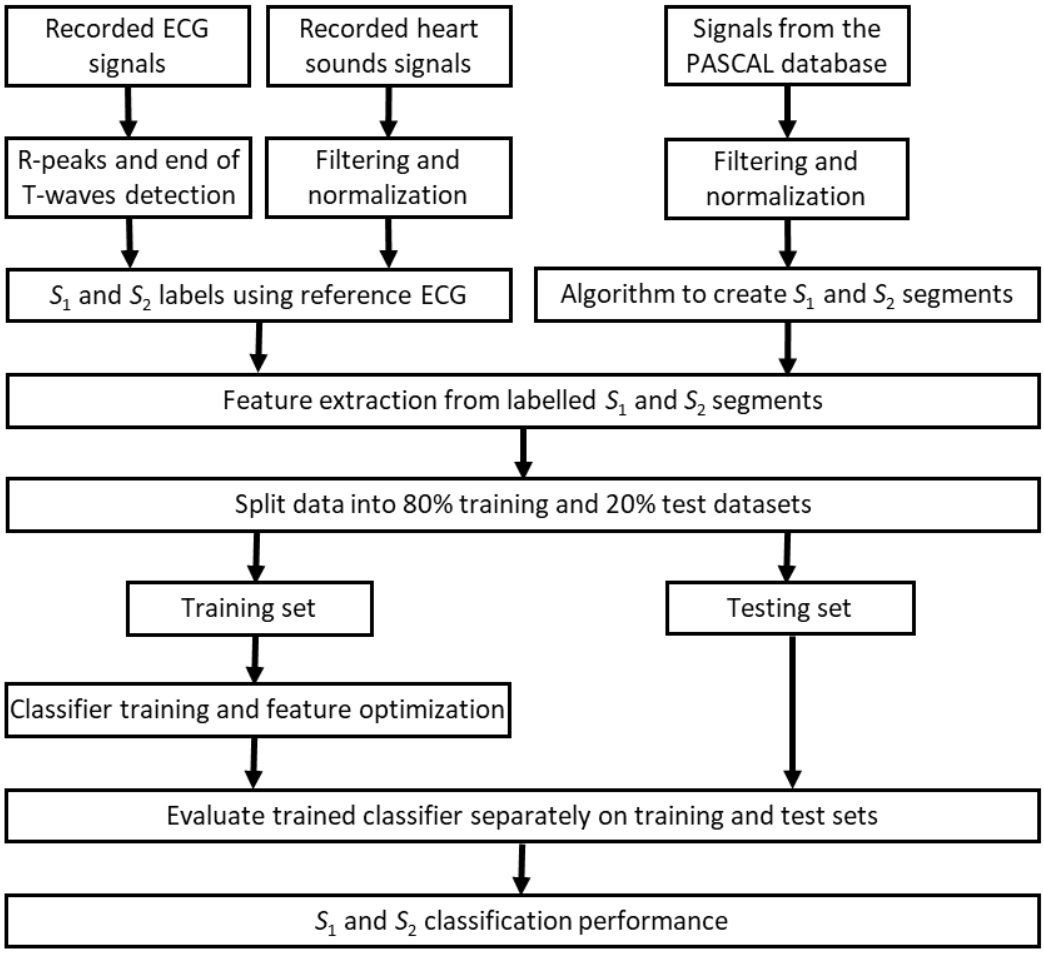


Figure 3.2. Flow-chart of the overall system to classify S_1 and S_2 sounds using features extracted from labelled segments of S_1 and S_2.

scales. As the nonstationary part of an acoustic signal in the time domain produces large wavelet coefficients over many wavelet scales, it is possible to characterize the coefficients with respect to the amplitude [236], [237]. The scale (dilation) and shift (translation) parameters, a ($a > 0$) and b ($b \in \Re$), respectively, are used to produce families of functions $\psi_{a,b}(t)$ from the mother wavelet $\psi(t)$ as [236]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3.3}$$

The projection of the signal $f(t) \in L^2(\Re)$, where $L^2(\Re)$ is the vector space of $f(t)$, onto the $\psi_{a,b}(t)$ is given as:

$$W(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{3.4}$$

where, '$*$' indicates the complex conjugate. Further, the inverse wavelet transform (*IWT*) is used to reconstruct $f(t)$ as:

$$f(t) = C_{\psi}^{-1} \int_{0}^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} W(a,b)\psi_{a,b}(t)db \qquad 3.5$$

where,

$$C_{\psi} = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty \qquad 3.6$$

Here, $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. This basis of noise reduction using the decomposition and reconstruction of wavelet coefficients in combination with the selectively modifying these coefficients prior to reconstruction, has been used in the adaptive sub-level tracking approach shown in Figure 3.4. Recordings were decomposed using the db5 wavelet by iterating the decomposition process *n* times (or levels) to obtain approximation ($C_A^{(n)}$) and detail coefficients ($C_D^{(n)}$) (where n=1, …, 5). Further, to find the optimum level of decomposition among the various possible levels representing the informative set of coefficients using the classical entropy-based criteria, Shannon entropy, norm entropy (p=1.5) and log energy were

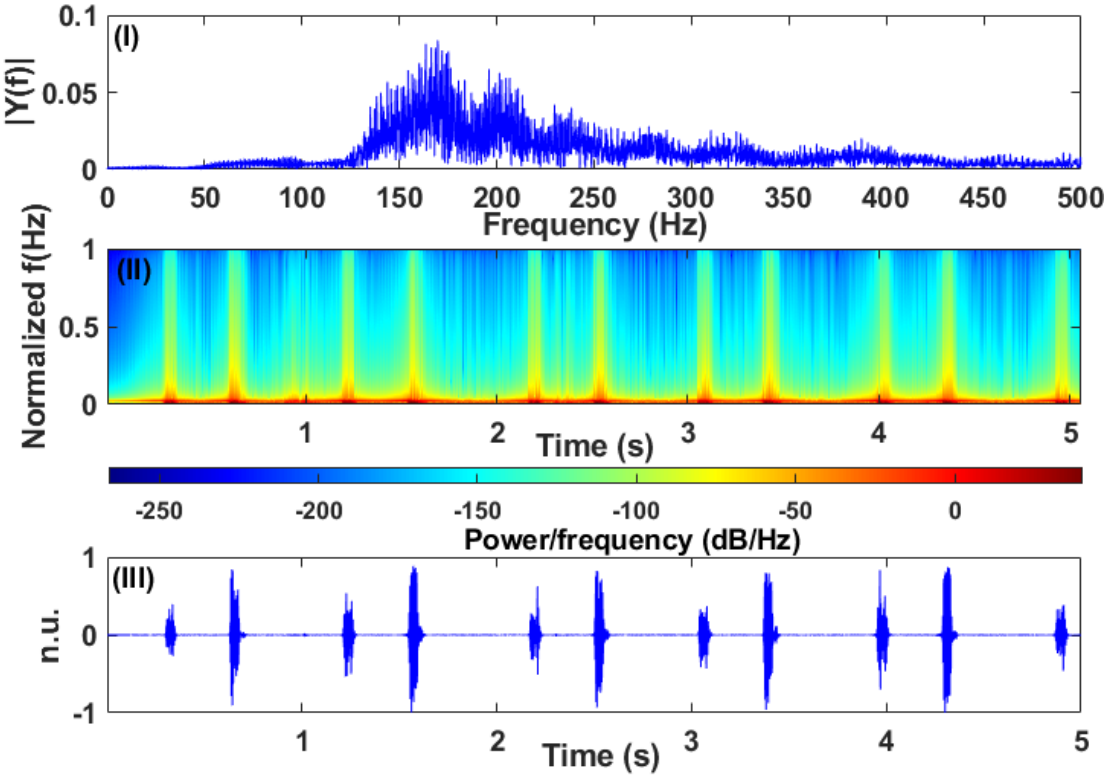


Figure 3.3. (I) Frequency response of a pre-processed recorded acoustic signal. (II) Time-frequency analysis of a recorded acoustic signal obtained using STFT. (III) Pre-processed and filtered recorded heart sounds.

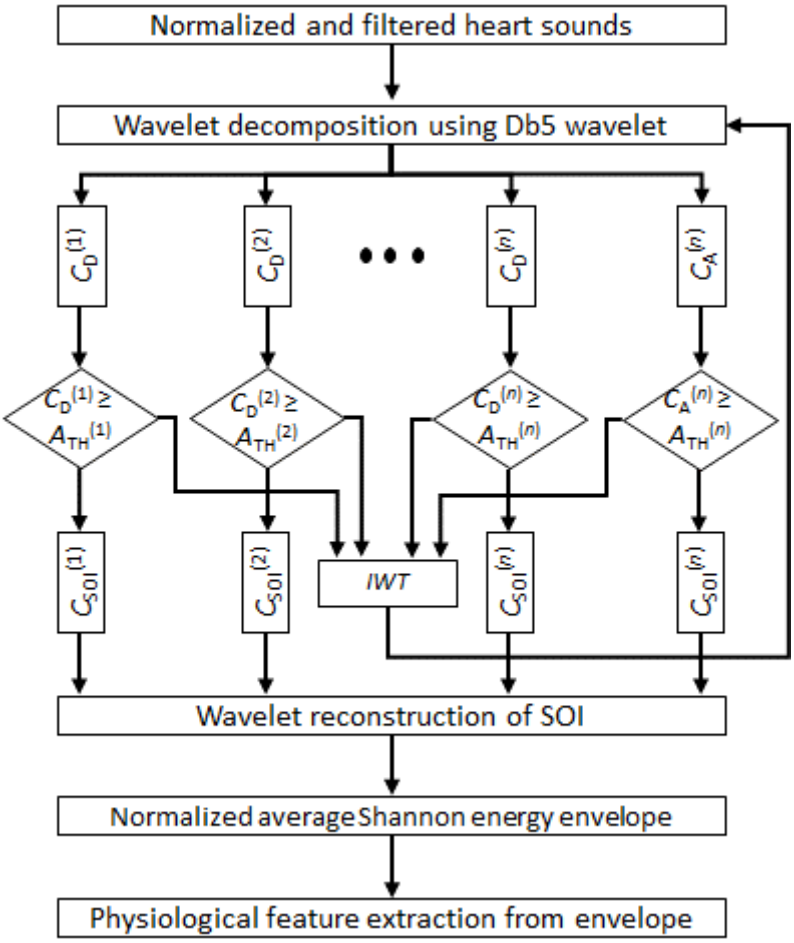


Figure 3.4. The different steps used in the adaptive wavelet-based filtering and the localization of sound peaks of interest in recordings. (In this work n=5).

estimated as shown in Table 3.2, and 5th-level with minimum entropy was selected for the analysis [238]. This was also verified using the subjective approach suggested in [239] to extract the optimum level for the analysis.

For each decomposition level, the most significant large coefficients ($C_D^{(n,i)} \geq A_{TH(n,i)}$) at each scale were extracted as coefficients of sounds of interests $\left(C_{SOI}^{(n)}\right)$. And lower coefficients

Table 3.2: Various entropy estimation for the signal shown in Figure 3.5.

Decomposition level	Shannon entropy (×10^{-3})	Norm (p=1.5) (×10^{-3})	Log energy (×10^{-7})
Signal	5.0290	3.8715	-3.4642
Level 5	2.1076	1.5797	-1.4132
Level 6	2.2507	1.7366	-1.3668
Level 7	2.2881	1.7858	-1.4142

$\left(A_{TH(n,i)} \geq C_D^{(n,i)}\right)$ were used as input for the next iteration (represented as $\left(C_{EHS}^{(n)}\right)$). $C_D^{(n,i)}$ and $A_{TH(n,i)}$ represent the decomposition coefficients and adaptive threshold (A_{TH}), respectively, for the n^{th}-level decomposition and i^{th}-iteration. These can be expressed as [235], [240]:

$$A_{TH(n,i)} = \left|mean(C_D^{(n,i)})\right| + A_{F(n,i)} \times \left|Std(C_D^{(n,i)})\right| \quad 3.7$$

where, $mean(C_D^{(n,i)})$ and $Std(C_D^{(n,i)})$ represent the mean and standard deviation of coefficients, respectively. $A_{F(n,i)}$ is an adjustment factor ranging between 2 and 3 [240]. This step was repeated for each decomposition level until the residual part became comparable to the stopping criterion (S_C), determined using the maximum likelihood method:

$$S_C = \left|\frac{E\left(C_{EHS}^{(n,i)^2}\right) - E\left(C_{EHS}^{(n,i-1)^2}\right)}{E\left(C_{EHS}^{(n,i)^2}\right)}\right| \quad 3.8$$

where, $E\left(C_{EHS}^{(n,i)^2}\right)$ and $E\left(C_{EHS}^{(n,i-1)^2}\right)$ are the expected value after squaring lower coefficients $\left(C_{EHS}^{(n)}\right)$ during the i^{th} and $(i\text{-}1)^{th}$ iteration, respectively. Finally, sounds of interest (*SOI*) were reconstructed using $C_{SOI}^{(n)}$ coefficients (where n=1, …, 5) at every level (using both approximation and detail coefficients) as [235]:

$$SOI = IWT\left(\sum_{n=1}^{5} C_{SOI}^{(n)}\right) \quad 3.9$$

where, IWT represents the inverse wavelet transform. A reconstructed signal obtained in this step, as illustrated in Figure 3.5, demonstrated that most of the unwanted components are eliminated from $HSS[n]$ and coefficients of sound peaks of interest were used to localize sounds in recorded signals.

3.2.2 Localization of sound peaks of interest

A Shannon energy envelope retaining the physiological characteristics of signals and attenuating samples of low amplitude, much more than samples of high amplitude, was extracted using a 10ms window throughout the signal with 5ms of overlap window, as follows:

$$E_{\mathrm{S}} = \frac{-1}{N}\sum_{i=1}^{N} h^2[n](\mathrm{i}).\log h^2[n](i) \quad 3.10$$

where, N is the number of samples in the normalized signal ($h[n]$). The zero-crossing points and subsequent start and end-points of all peaks were extracted using the normalized Shannon

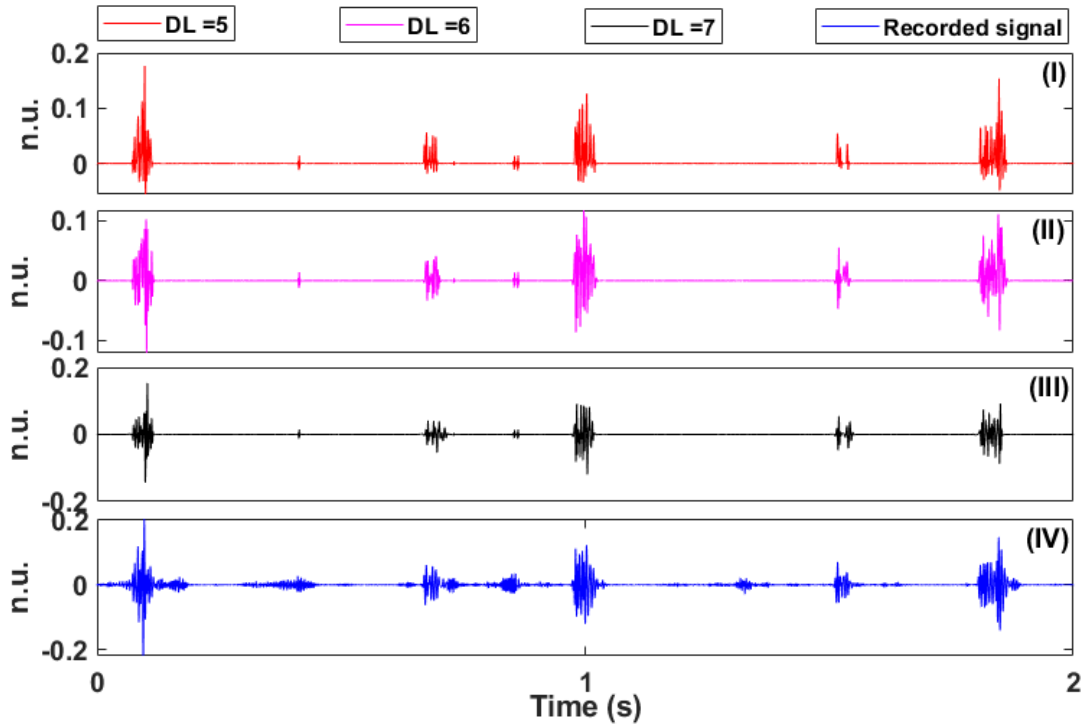


Figure 3.5. (I), (II) and (III) are the reconstructed recorded signals using adaptive sublevel filtering for different decomposition levels (DL) 5, 6 and 7, respectively. (IV) is the original recorded acoustic signal.

energy $E_{Snorm}[n]$ that places the signal below the baseline, expressed as:

$$E_{\mathrm{Snorm}}[n] = \frac{E_{\mathrm{s}}[n] - \overline{E_{\mathrm{s}}[n]}}{\sigma(E_{\mathrm{s}}[n])} \qquad 3.11$$

where, $E_{\mathrm{s}}[n]$ is the average Shannon energy, and $\overline{E_s[n]}$ and $\sigma(E_s[n])$ represent the mean and standard deviation of $E_{\mathrm{s}}[n]$, respectively. Start and end-points of peaks in combination with prior information of fundamental heart sounds from [208], [217], were used to localize sound peaks of interest as shown in Figure 3.6. The next step was to identify S_1 and S_2 among all the localized peaks using reference ECG signals from the RHSS database or using the reference locations of S_1 and S_2 available from [23].

3.2.3 Annotation of S_1 and S_2

This section discusses the approach used to create S_1 and S_2 labels for training and testing, using the gold-standard reference ECG signals, recorded simultaneously with heart sounds. For this purpose, separate algorithms were used to identify *R*-peaks and end-points of *T*-waves in ECG signals using the following steps:

1) High-frequency components were removed from recorded ECG signals (shown in

Figure 3.7(II)), using a zero-phase bandpass filtering, followed by a derivate filtering and the moving-window integration [241].

2) The Pan-Tompkins algorithm [242] was used to detect *R*-peaks. The performance of this method had been already tested with various databases, including the MIT-BIH Arrhythmia Database [243], [244].

3) *T*-waves in ECG signals were identified using the approach described in [245], as shown in Figure 3.7(V).

S_1 occurs just after the *R*-peaks and S_2 towards the end of *T*-waves [70]. Therefore, *R*-peaks and end-*T*-waves identified in an ECG recording (as shown in Figure 3.7) were used to mark S_1 and S_2 present among the sound peaks localized (as shown in Figure 3.6(b)) in a simultaneously recorded cardiac acoustic signal. Examples of ECG and PCG signals with different peaks marked are shown in Figure 3.8. Further, to estimate the positions of start and endpoints of each localized sound of interest (SOI), zero cross-over points were used. These start and endpoints were used to extract the segments of S_1 and S_2 which were identified using *R*-peaks and end-*T*-waves. This also provided enough tolerance to accurately locate the corresponding points in the time scale, and to estimate the duration of peaks precisely. The locations of S_1 and S_2 were subsequently used to mark the systole and diastole intervals, assisting the estimation of cardiac cycles as shown in Figure 3.8. Thus, a total of 16825 S_1 and 16808 S_2 labelled segments were extracted from the RHSS database. Note that ECG signals were only used as a reference to label S_1 and S_2 for creating dataset for training and testing, but they played no role in the evaluated performance of the approach discussed.

Additionally, 753 S_1 and 736 S_2 segments were extracted from the PASCAL database [23]. Sound peaks were identified using the steps discussed in sections 3.2.1 and 3.2.2, and S_1 and S_2

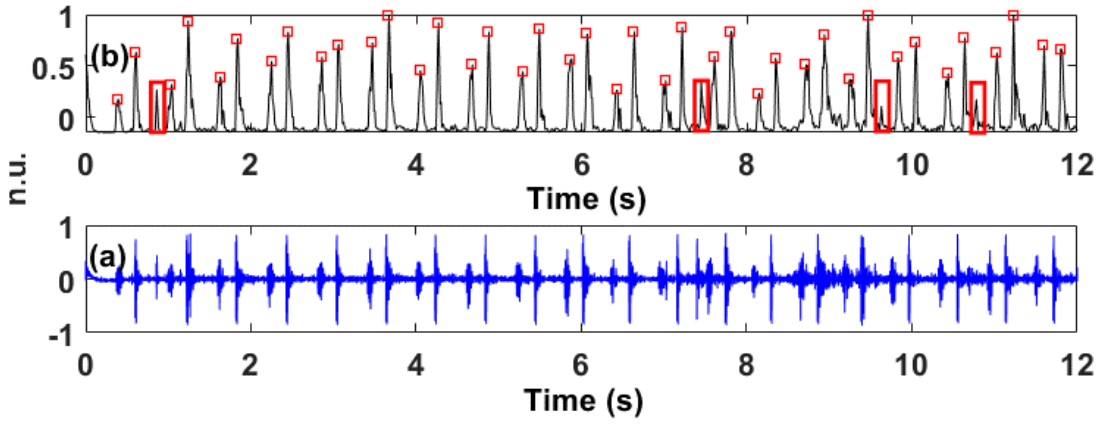


Figure 3.6. Localization of sound peaks of interest. (a) recorded heart sounds signal. (b) S_1 and S_2 sounds after discarding extra peaks based on priori information.

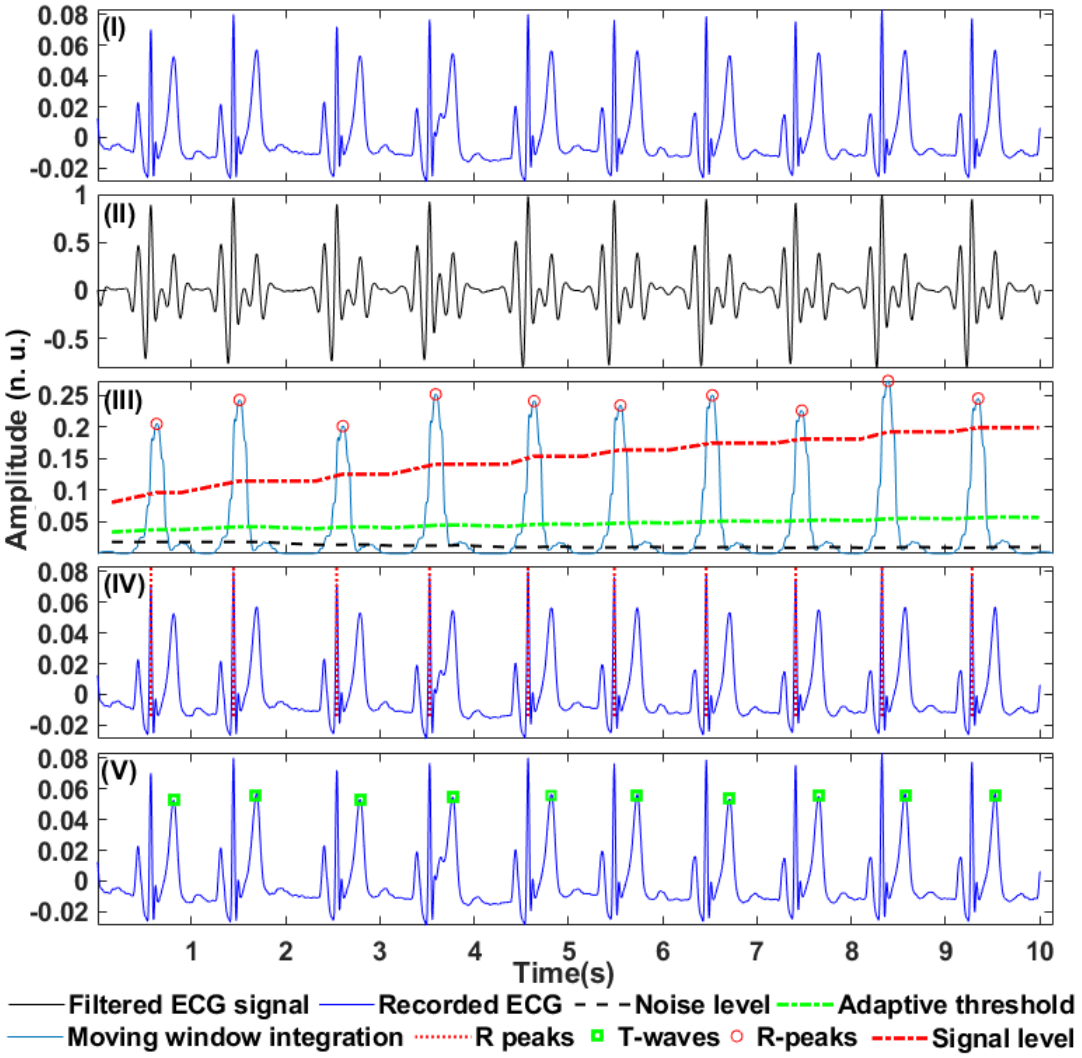


Figure 3.7. (I)-(IV) The various steps used in the identification of *R*-peaks in recorded ECG signal using the Pan-Tompkins algorithm, used as reference to mark S_1. (V) Identified *T*-waves in ECG signals used as reference to mark S_2.

were labelled using the location of S_1 and S_2 provided in the database. As an illustration, labelled S_1 and S_2 sounds in a recorded signal from the PASCAL database is shown in Figure 3.9; and the extracted S_1 and S_2 segments (8 segments of each sound) are shown in Figure 3.10 and Figure 3.11, respectively.

Overall, 17578 S_1 and 17544 S_2 were used to evaluate features and subsequently for training and testing the classifier. In addition to S_1 and S_2 sounds, localized sound peaks of interest that were neither identified as S_1 nor S_2, using the reference locations, were labelled as artifacts. Thus, a total of 3707 artifacts (random sounds from different sources) segments from the RHSS database and 335 artifacts segments from the PASCAL database were used in the study.

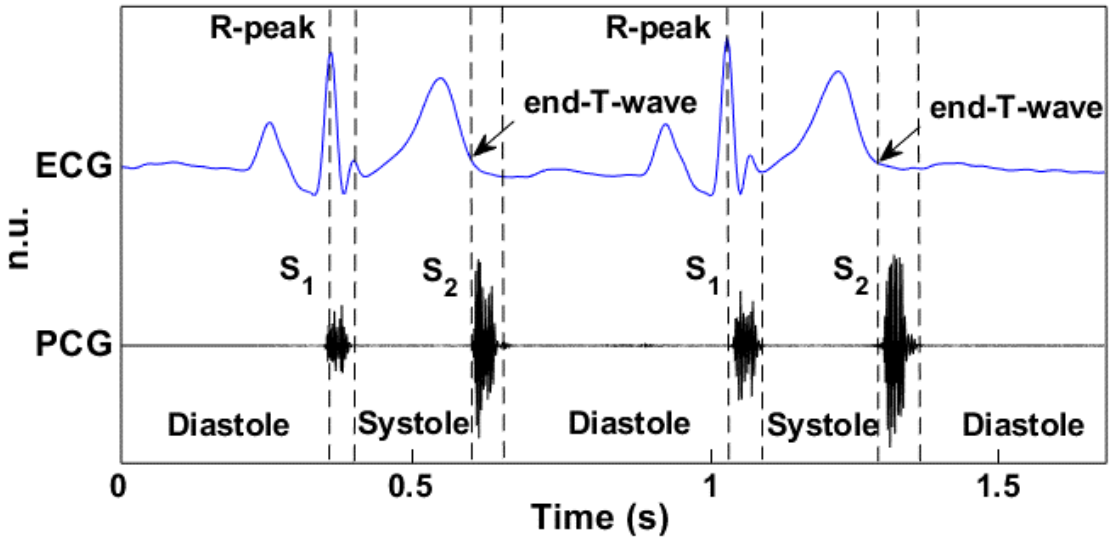


Figure 3.8. S_1, systole, S_2, and diastole intervals in a filtered cardiac acoustic signal labelled using *R*-peaks and end-*T*-waves locations identified in an ECG recording.

3.3 Feature Extraction & Selection

Due to the non-stationary and non-linear characteristics of heart sounds, several features containing a quantitative description of S_1 and S_2 were extracted from the labelled dataset to assist with the identification of different states (S_1 or S_2). Overall, 87 features were used in this study and their details can be found in Table 3.3.

Neighborhood component analysis (NCA), used to derive a set of most significant features from all the features extracted, allocated weights to each feature by maximizing the expected leave-one-out classification accuracy [219]. Features that corresponded to large feature weights (or above a specified threshold (δ)) were selected as optimum features. Out of the 87 features reported in Table 3.3, the normalized weights of only 14 selected features are shown in Figure 3.12. The 14 most relevant features, having weights more than 0.1, were selected for the

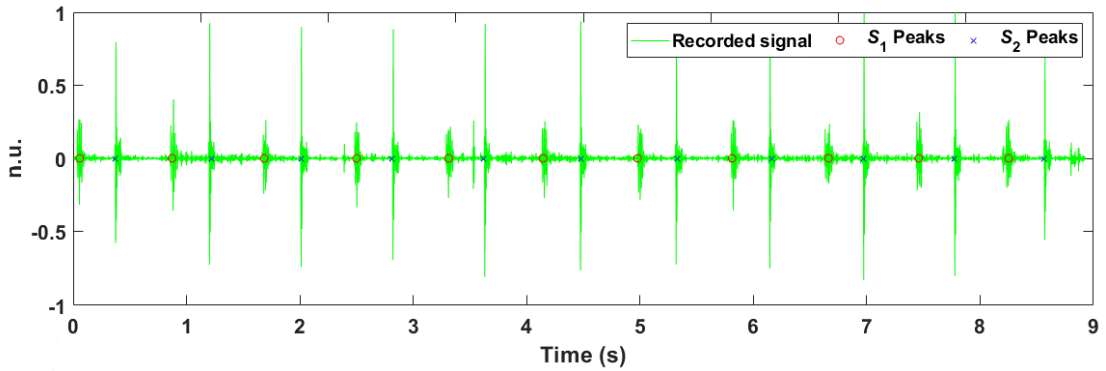


Figure 3.9. Labelled S_1 and S_2 sounds in a recorded signal from the PASCAL database.

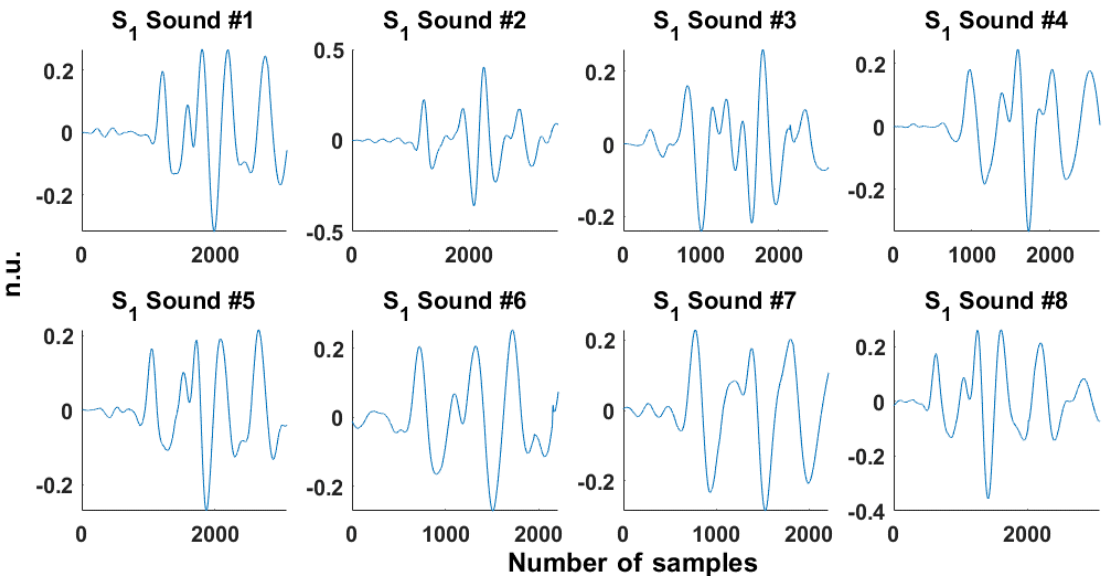


Figure 3.10. S_1 sound segments extracted from the signal shown in Figure 3.9.

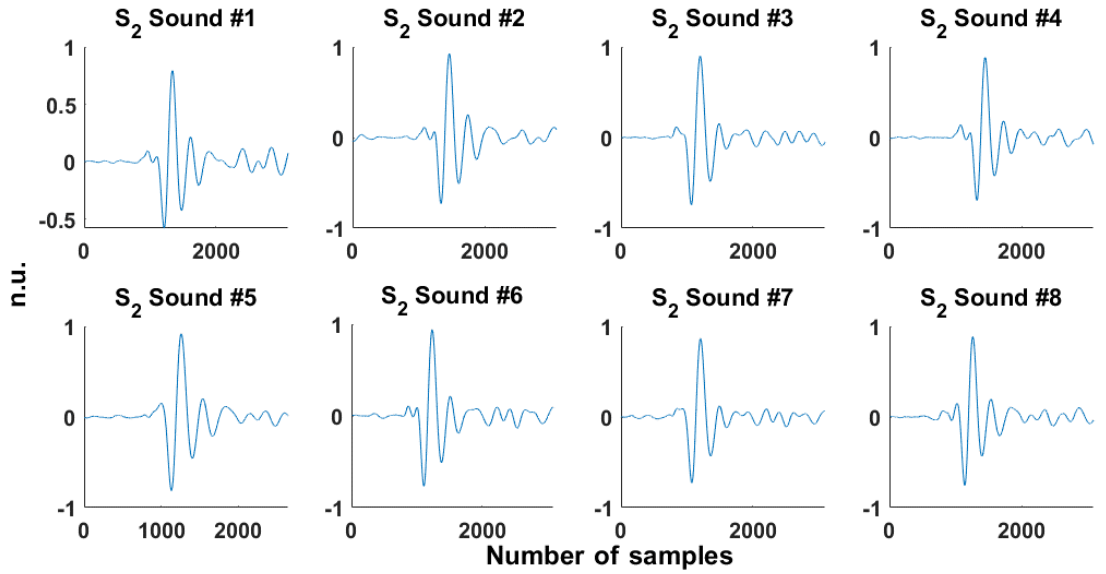


Figure 3.11. S_2 sound segments extracted from the signal shown in Figure 3.9.

optimized classifier model. These feature weights were evaluated from both the PASCAL database and RHSS database, and it was found that the highest-ranking features were almost the same for both the datasets. Of these, the most relevant feature was found to be F_{47}.

As shown in Figure 3.12, MFCCs features (F_{59} and F_{60}) were also selected, showing the ability of perceptual features to significantly discriminate S_1 and S_2 sounds. Among all statistical features, only F_{82} and F_{84}, representing the kurtosis and dominant frequency of each window were selected. The boxplot of the 14 selected features shown in Figure 3.13 reveals the intra-class variations for S_1 and S_2 sounds. MODWT and Hilbert-Huang transform-based

features are plotted in Figure 3.14, showed potential in differentiating S_1 and S_2 sounds, which

Table 3.3: Various features extracted and their description (F_1-F_{73}).

Feature#	Type	Feature	Description
$F_1 - F_{10}$	MODWT	Energies at different scales	9-levels of decomposition details and approximation coefficients produced 10 features
F_{11}	MODWT	Energy	Total energy of all the scales
$F_{12} - F_{21}$	MRA of MODWT matrix	Energies at different scales	10 features extracted from the MODWT matrix
F_{22}	MRA	Energy	Total energy of all the levels
$F_{23} - F_{25}$	Homomorphic envelope	Frequency and amplitude	The dominant frequency of the envelope, spectral value, and peak of the envelope
$F_{26} - F_{28}$	Hilbert spectrum	Frequency and amplitude	
$F_{29} - F_{34}$	EMD	Energy	Energy of first 5 IMFs and the residue
$F_{35} - F_{42}$	EMD	Frequency and magnitude	Maximum frequency in the spectrum and its corresponding value in the spectrum form 4 IMFs
$F_{43} - F_{50}$	Hilbert-Huang transform	Instantaneous frequency	IF and cluster of IF for each mode for 4 selected IMFs (8 features)
$F_{51} - F_{58}$	Hilbert-Huang transform	Instantaneous energy	IE and cluster of IEs for each mode for 4 selected IMFs (8 features)
$F_{59} - F_{73}$	Perceptual features	MFCCs	15 features from each segment
F_{74}	Statistical features	Mean	$\overline{h[n]} = \frac{1}{N}\sum_{i=1}^{N} h[i]$
F_{76}		Standard deviation	$std = \sqrt{\frac{\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2}{N-1}}$
F_{81}		Skewness	$skw = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^{3/2}}$
F_{82}		Kurtosis	$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^2}$
F_{87}		Zero-crossing rate	$Z_R = \frac{1}{2N}\sum_{i=1}^{N}\lvert sgn[h[i]] - sgn[h[i-1]]\rvert$
F_{83}	Frequency domain-based features	Spectral Entropy	$S_E = \frac{\sum_{i=1}^{N} P_i . \log P_i}{\log_2 N}$
F_{84}		Dominant frequency	Frequency corresponding to the maxima of power spectrum
F_{85}		Maximum frequency	$f_{max} = \max(frequency)$

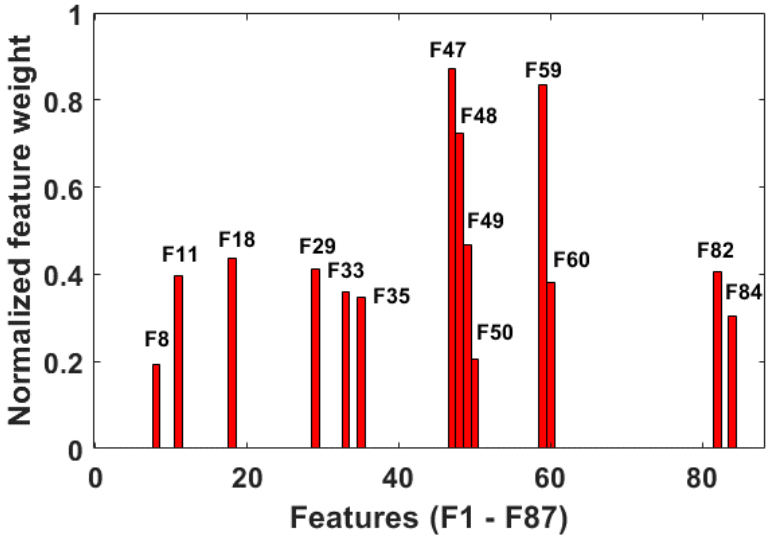


Figure 3.12. Normalized feature weights of the selected features obtained using the NCA based feature selection approach. Feature weights were evaluated for both the PASCAL database and the RHSS database.

made them suitable for training the classifier.

Additionally, to compute feature weights using NCA, the optimum value of name-value pair argument '*Lambda*' representing the regularization parameter was optimized against the minimum classification loss using cross-validation. In general, *Lambda* is equal to 1/N where, *N* is the size of the training data set. However, in this work, the optimized mean value of *Lambda* was found to be 0.005 corresponding to the minimum classification loss of 0.0268. These values were used in optimizing the classifier model using learning algorithms. The classifier's performance using these features is discussed in section 3.5.

3.4 Classifier Training and Testing

The approach is to develop a classifier using a training set that decodes the state of unseen test samples given their input features. The main problem with the supervised learning is that the dataset may not contain physiological measurement from varied population. This is really important as the characteristics of the signals vary widely from subject to subject. To address this, two different databases were used in the study, which included signals recorded in this study as well as signals available from an open source.

In this study, an 87-dimensional feature vector from $x^{(i)} \in R^n$ was used to train an ensemble classifier to interpret the class of $y^{(i)} \in R$; enabling the trained classifier to

potentially decode the state of unseen test samples given their input features.

3.4.1 Training and test datasets

In this study, the feature vector was evaluated from S_1, S_2 and artifacts segments, for each analysis window of 5 ms, producing the dimensions of feature matrix $X \in R^{159762 \times 14}$, from the RHSS database, where 159762 (52718 S_1, 55068 S_2, and 51976 artifacts) represents the total number of PCG instances. The hold-out approach and 10-fold cross-validation were used to split the data (feature matrix) into training and testing datasets. Thus, for hold-out, 80% of the data was used for the training feature matrix $X_{training} \in R^{127810 \times 14}$ and the remaining 20% of the data was used for the testing feature matrix $X_{testing} \in R^{31952 \times 14}$, where 127810 (42175 S_1, 44054 S_2 and 41581 artifacts) and 31952 (10543 S_1, 11014 S_2 and 10395 artifacts) are the number of PCG instances or windows used for training and testing, respectively. Here, 107786 FHS (86229 for training and 21557 for test set) represents the total number of S_1 and S_2 sounds (fundamental heart sounds) in the feature matrix. Similarly, the feature matrix $X \in R^{4198 \times 14}$, was evaluated from the PASCAL database, where 4198 (2039 S_1, 1751 S_2, and 408 artifacts) represents the total number of PCG. For hold-out, 80% of the data was used for the training feature matrix $X_{training} \in R^{3359 \times 14}$ and the remaining 20% of the data produced a testing feature matrix $X_{testing} \in R^{839 \times 14}$, where 3359 (1631 S_1, 1401 S_2 and 327 artifacts) and 839 (408 S_1, 350 S_2 and 81 artifacts) are the number of PCG instances used for training and testing, respectively. Here, 3790 FHS (3032 for training and 758 for test set) represents the total number of S_1 and S_2 sounds in the feature matrix. These features were extracted from an unbalanced number of segments from each class of S_1 and S_2. Hence, the random undersampling boosting (RUSBoost) algorithm designed to address the problem of training classifier on skewed data was used [246]. This algorithm undersamples the classes with more observations by taking only N observations from each class, where N is the number of observations from the class having the fewest observations.

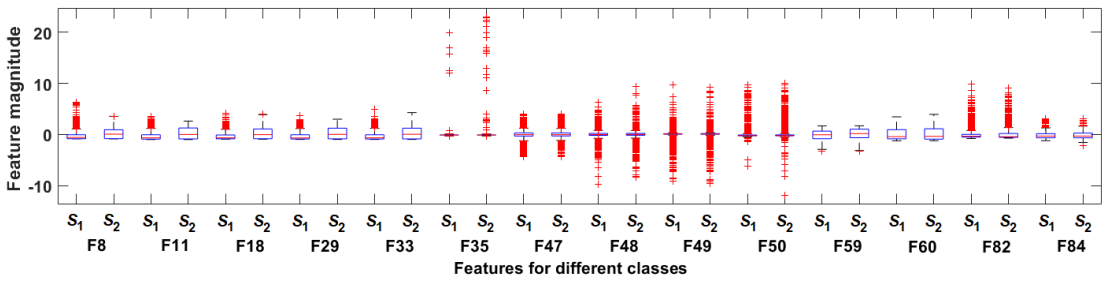


Figure 3.13. Boxplot of the top 14 features selected to differentiate S_1 and S_2 sounds identified using NCA based feature selection approach.

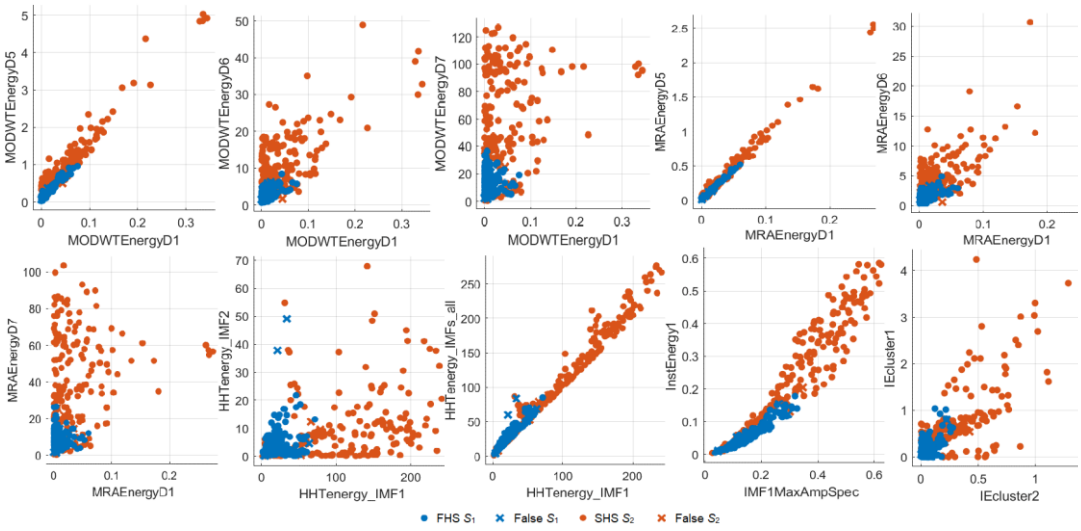


Figure 3.14. Key features of S_1 and S_2 sounds identified using NCA based feature selection approach. Plotted features are energy of different levels for MODWT (MODWTEnergy), energy computed using MRA (MRAEnergy), energy of IMFs using HHT (HHTenergy), instantaneous energy (InstEnergy) and IE using *k*-means clustering (IEcluster).

3.4.2 Classifier training

Recent works on identification of S_1 and S_2 sounds using individual machine learning classifiers have demonstrated promising results [75], [232], [233]. Hence, the ensemble machine learning algorithms were used with the objective of improving performance. The ensemble learning approach combines the prediction of several learners to derive a powerful model that performs better than any single classifier [68], [228]. This approach constructs a set of classifiers with different weights assigned using an adaptive ensemble algorithm and then classifies a new dataset by taking a weighted vote of their predictions. Thus, the weighted combination of multiple algorithms provides greater flexibility by reducing the variance of predictions to achieve good performance and robustness.

The approach used to train the classifier and to extract optimum parameters of the classifier using the input feature matrix to predict the class (S_1 or S_2 or artifacts) of PCG instance in normal and abnormal recordings is demonstrated in Figure 3.15. The classification task was divided into two sub-classification tasks. First, PCG instances were classified as fundamental heart sounds (FHS i.e. S_1 or S_2) or artifacts, and second, fundamental heart sounds were classified as S_1 or S_2.

Several algorithms were tested, and their performance evaluated using different metrics (sensitivity, precision, and F_1-score) to find the most suitable classification algorithm [70],

[233]. In this study, decision trees are used as weak learners due to their capacity to deal with high dimensional datasets. The boosting algorithm was used to optimize the weights and parameters of the weak learners at each iteration during the cross-validation [247]. Among several boosting algorithms (adaptive boosting (AdaBoost) [248], adaptive Logistic Regression (LogitBoost) [247], random undersampling (RUSBoost) [246], GentleBoost [247], BrownBoost [249] and XgBoost [231]), the RUSBoost algorithm was used to handle imbalanced data [246].

3.4.3 Hyperparameter optimization

The 10-fold cross-validated misclassification rates with respect to the number of trees, learning rate, and for a different number of splits are plotted in Figure 3.16. It can be observed from the figure that an optimum misclassification rate of 0.04348 is obtained for 71 number of trees, 9 maximum number of splits for a learning rate of 0.25, while tuning the number of trees, the learning rate (0.1, 0.25, 0.5, 0.75 and 1) and a maximum number of splits (1, 3, 9, and 27). Here, a deep classification tree and a stump tree (1-level decision tree) were used as a standard reference to compare the misclassification rate of the classifier.

3.5 Experimental Results

The performance of the proposed classifier has been assessed using widely accepted evaluation metrics used in heart sounds analysis and classification [70], [75], [89], [233]. These evaluation metrics are: recall or sensitivity $\left(S_E = \frac{TP}{TP+FN}\right)$, classification accuracy $\left(C_A = \frac{TP+TN}{TP+FP+FN+TN}\right)$, precision $\left(P_+ = \frac{TP}{TP+FP}\right)$, and F_1 score $\left(F_1 = 2\frac{P_+ \times S_E}{P_+ + S_E}\right)$. S_E was computed to estimate the true

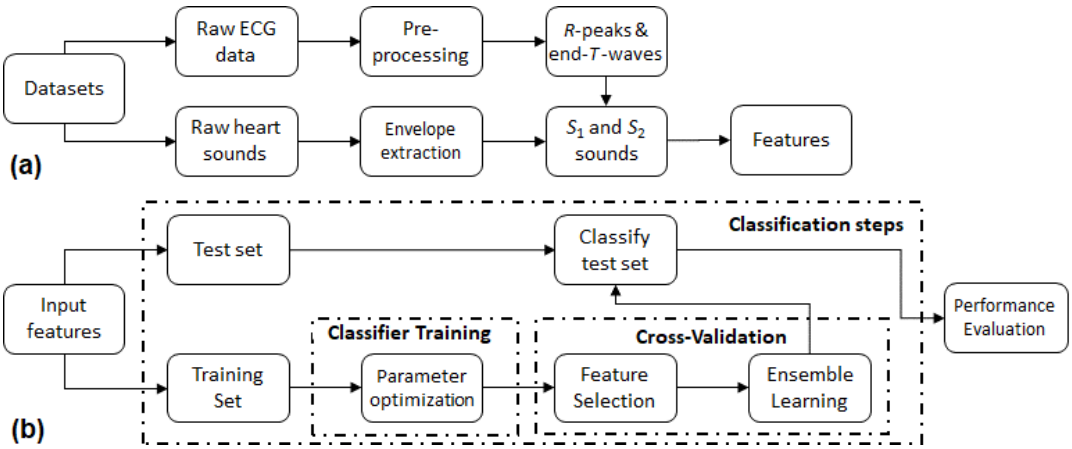


Figure 3.15. (a) Feature extraction from raw signals and (b) steps used for training and cross-validation of the classifier to extract hyperparameters.

positive rates. And F_1 score was computed to represent a single harmonic mean of S_E and P_+ [70]. Here, true positive (*TP*) represents the correct result, false positive (*FP*) is false alarm, false negative (*FN*) is missed detection and true negative (*TN*) is the correct absence of result. PCG instances were labelled as *TP*, *FP*, *FN*, and *TN* as defined in previous studies [75], [89], [220], [234]. Thus, any instance where the predicted class (S_1 or S_2 or artifacts) matched the ground truth (or actual class) was labelled as a true detection.

As the dataset contains recordings from three categories (S_1, S_2, and artifacts), the performance was evaluated using a 3 × 3 (three-class) confusion matrix, denoted as *P* [220]. Table 3.4 shows the interpretation of the confusion matrix *P*, where P_{ij} represents the number of PCG instances where the predicted class was *i* and the ground truth (or actual class) was *j*. Thus, the number of *TP*, *TN*, *FP*, and *FN*, for each class *k* were computed as follows:

$$TP = P_{kk} \tag{3.12}$$

$$TN = \sum_{i,j \neq k} P_{ij} \tag{3.13}$$

$$FP = \sum_{i \neq k} P_{ki} \tag{3.14}$$

$$FN = \sum_{i \neq k} P_{ik} \tag{3.15}$$

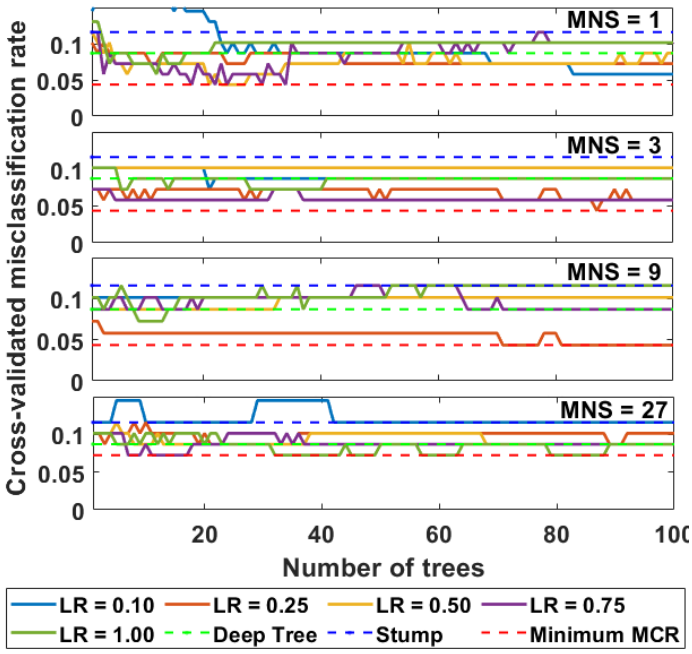


Figure 3.16. Key hyperparameter optimization using misclassification rate (MCR). MNS represents the maximum number of splits in the tree.

During segmentation, 37340 sound peaks of interest were localized in recordings from the RHSS database, out of which 16825 were identified as S_1 and, 16808 as S_2, using R-peaks and end-T-waves. Rest of the peaks (3707) were labelled as artifacts. Similarly, 1824 sound peaks of interest were localized in the recordings from the PASCAL database, out of which 753 were identified as S_1 and 736 as S_2, using the locations of S_1 and S_2 provided in the database. Rest 335 sound lobes were labelled as artifacts. Thus, all labelled S_1, S_2 and artifacts, from both the databases were used to evaluated features for training and testing as described in section 3.4.1. The number of observations from the different classes in the training and test sets for both the RHSS and PASCAL databases are also reported in the previous section.

The confusion matrices of the trained classifier for the different datasets are shown in Table 3.5 (FHS = Fundamental Heart Sounds (S_1 or S_2) and A = artifacts). For each dataset, two confusion matrices showing the classification of PCG instances as fundamental heart sounds (FHS i.e. S_1 or S_2) or artifacts, and subsequent classification of fundamental heart sounds as S_1 or S_2, are reported. Finally, 3 × 3 confusion matrices for both the databases (training and test sets) are shown in Table 3.6. These values were used to evaluate the performance of the classifier on the training (using 10-fold cross-validation) and test sets for both the RHSS and PASCAL databases as shown in Table 3.7. The scores illustrated in the table are the average values on all observations, counting each observation when it was in a held-out fold. In order to illustrate the robustness of the achieved results, the classifier's performance was evaluated using the training and test datasets of both the databases. From results shown in Table 3.7, it can be observed that the ensemble classifier with optimized hyperparameters achieved an average F_1 score of 90.65%, 89.43% and 92.08% using the PASCAL database and the RHSS database, in classifying S_1, S_2 and artifacts, respectively. The overall accuracy using both the databases was found to be 90.79%.

The deviations in the performance parameters (F_1 score, S_E, and P_+) reported for both the training and test datasets were found to be less than 5% from the mean values. Results achieved were also almost consistent for both the training and test datasets. For instance, achieved F_1 score of the classifier under the unseen data from the PASCAL database was found to be 91.12% and 89.30% for S_1 and S_2 sounds. Similarly, F_1 score of 89.00% and 88.50% were

Table 3.4: Three-class confusion matrix.

		Actual class		
		Artifacts	S_1	S_2
Predicted class	Artifacts	P_{11}	P_{12}	P_{13}
	S_1	P_{21}	P_{22}	P_{23}
	S_2	P_{31}	P_{32}	P_{33}

achieved under the unseen data from the RHSS database for S_1 and S_2 sounds, respectively. Furthermore, the method achieved an average S_E of 92.24% and 90.30% in classifying S_1 and S_2 sounds, respectively, using the training data. S_E was found to be 90.94% for S_1 and 89.25% for S_2, under the test data. These values were computed as an average value achieved using all the training and test data separately as shown in Table 3.7.

The classifier's performance in classifying S_1 and S_2 sounds recorded from different auscultation sites using extracted features was also studied. An average S_E of 90.82% and F_1 score of 89.43% was achieved, with very little deviation for S_1 and S_2 sounds, for the data recorded from four auscultation sites (aortic valve, tricuspid valve, mitral valve, and pulmonary valve). This leads to the conclusion that the extracted features are able to capture the characteristics helpful in the classification of S_1 and S_2 recorded from different auscultation sites. This also shows that the features selected to train the classifier are able to accurately identify S_1 and S_2 despite the differences due to physiological factors at varying locations.

Additionally, the performance of the algorithm was assessed in terms of confidence in classifying S_1 and S_2 sounds. The ensemble classifier scores were transformed into probabilities to find the likelihood of each observation, in both training and test datasets, being a member of

Table 3.5: Confusion matrices of the trained classifier for different datasets. (RHSS Training Set (A = Classification I and, B = Classification II), RHSS Test Set (C = Classification I and, D = Classification II), PASCAL Training Set (E = Classification I and, F = Classification II) and PASCAL Test Set (G = Classification I and, H = Classification II))

(A)

Predicted \ Actual	FHS	A
FHS	85466	3538
A	763	38043

(B)

Predicted \ Actual	S_1	S_2
S_1	38504	3166
S_2	3345	40451

(C)

Predicted \ Actual	FHS	A
FHS	21240	972
A	317	9423

(D)

Predicted \ Actual	S_1	S_2
S_1	9486	883
S_2	933	9938

(E)

Predicted \ Actual	FHS	A
FHS	2996	34
A	36	293

(F)

Predicted \ Actual	S_1	S_2
S_1	1520	135
S_2	97	1244

(G)

Predicted \ Actual	FHS	A
FHS	750	7
A	8	74

(H)

Predicted \ Actual	S_1	S_2
S_1	375	37
S_2	29	309

a class S_1 or S_2 [250], [251]. In this work, $P(S_1)$ and $P(S_2)$ (or $(1 - P(S_1))$) representing the probabilities of an observation X to belong to a class S_1 and S_2, respectively, were computed. These values were computed such that, if $P(S_1) < P(S_2)$ then the classifier was confident that X belongs to S_2 class, and vice versa. The '*logit*' function was used to map the ensemble classifier output score (representing the confidence of the classification into a class) into probability estimates. Figure 3.17 shows the $P(S_1)$ and $P(S_2)$ for all the observations in the training and test datasets, where $P(S_1) = 1$ represents the positive class (or S_1) and $P(S_2) = 0$ is the negative class (or S_2). To estimate the number of observations that were not classified as S_1 or S_2 confidently, probability values ($P(S_1)$ and $P(S_2)$) within the interval of ±5% and ±1%, from the mean decision boundary ($P(S_1) = P(S_2) = 0.5$) were considered (as shown in Figure 3.18). The value of ±5% was considered based on the confidence interval of 95% [252]. Thus, 773 and 3984 observations in the RHSS training set were found within the interval of ±1% and ±5% of the probability distribution from the mean decision boundary ($P(S_1) = P(S_2) = 0.5$), respectively.

Similarly, for the RHSS test set, 204 and 1112 observations were found within the determined interval of ±1% and ±5%, respectively, from the mean decision boundary. For the PASCAL database, 6 and 61 observations in the training set, and 2 and 17 observations in the test set, were found within the interval of ±1% and ±5% of the probability distribution from the mean probability, respectively. These numbers show the confidence of the algorithm in predicting the class of the observations. For instance, out of total 86229 FHS (both S_1 and S_2) in the RHSS training set, 4.62% (3984 observations) of the observations were classified as

Table 3.6: Confusion matrices of the trained classifier for three-classes from different datasets. RHSS Datasets (A = Training Set and B = Test Set) and the PASCAL Datasets (C = Training Set and D = Test Set)

(A)

		Actual class		
		Artifacts	S_1	S_2
Predicted class	Artifacts	38043	326	437
	S_1	1371	38504	3166
	S_2	2167	3345	40451

(B)

		Actual class		
		Artifacts	S_1	S_2
Predicted class	Artifacts	9423	124	193
	S_1	401	9486	883
	S_2	571	933	9938

(C)

		Actual class		
		Artifacts	S_1	S_2
Predicted class	Artifacts	293	14	22
	S_1	13	1520	135
	S_2	21	97	1244

(D)

		Actual class		
		Artifacts	S_1	S_2
Predicted class	Artifacts	74	4	4
	S_1	3	375	37
	S_2	4	29	309

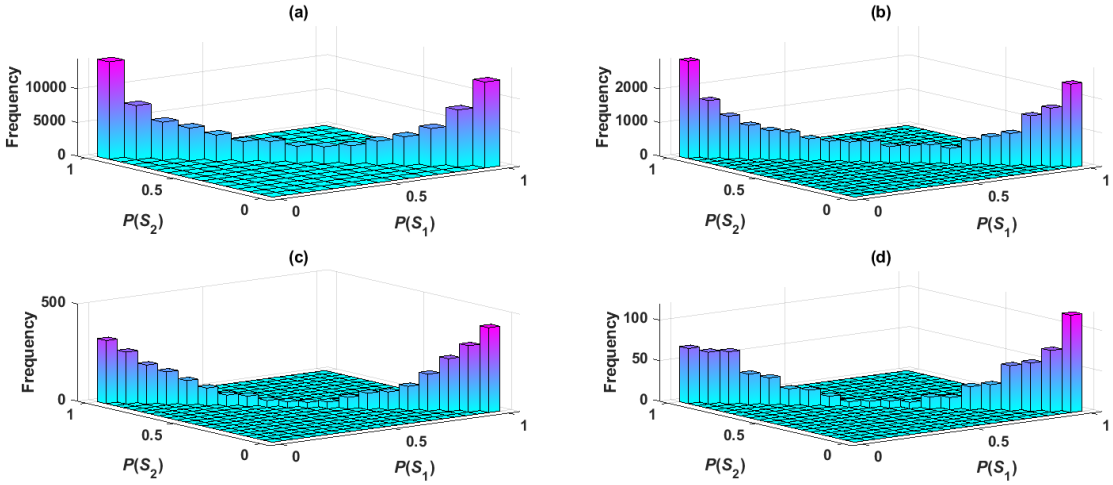


Figure 3.17. $P(S_1)$ and $P(S_2)$ for all the observations in: (a) RHSS training set (86229 observations), (b) RHSS test set (21557 observations), (c) PASCAL training set (3032 observations) and (d) PASCAL test set (758 observations). Frequency represents the number of observations having a given probability $P(S_1)$ or $P(S_2)$.

either S_1 or S_2 with probabilities less than ±5%. Similarly, for the PASCAL training set, out of 3032, 2.01% (61 observations) were classified as S_1 or S_2 with probabilities less than ±5%. Achieved results demonstrated that the trained classifier model was confident in predicting the class of observations based on the feature set used to train the classifier.

Table 3.7: Performance of the trained classifier using 10-fold cross validation.

		P_+ (%)	S_E (%)	F_1 (%)	C_A (%)
RHSS Dataset (training)	Artifacts	98.03	91.49	94.64	91.54 (overall)
	S_1	89.45	91.29	90.36	
	S_2	88.00	91.82	89.86	
Average		*91.826*	*91.53*	*91.62*	
RHSS Dataset (test)	Artifacts	96.74	90.64	93.59	90.28 (overall)
	S_1	88.07	89.97	89.00	
	S_2	86.85	90.23	88.50	
Average		*90.55*	*90.28*	*90.36*	
PASCAL Dataset (training)	Artifacts	89.05	89.60	89.32	91.00 (overall)
	S_1	91.12	93.19	92.14	
	S_2	91.33	88.79	90.04	
Average		*90.5*	*90.52*	*90.5*	
PASCAL Dataset (test)	Artifacts	90.24	91.35	90.79	90.34 (overall)
	S_1	90.36	91.91	91.12	
	S_2	90.35	88.28	89.30	
Average		*90.31*	*90.51*	*90.40*	

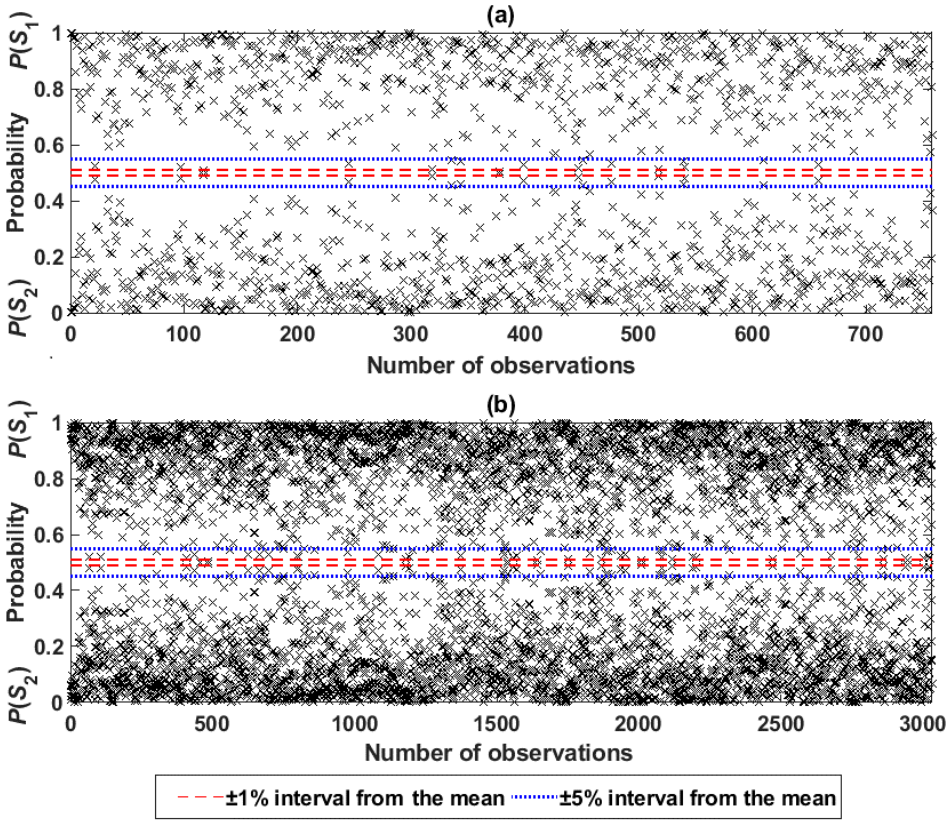


Figure 3.18. Probabilities ($P(S_1)$ and $(1 - P(S_1))$) for each observation and the intervals used to estimate the classifier's performance in: (a) PASCAL test set (758 observations) and (b) PASCAL training set (3032 observations).

3.6 Discussion

The performance comparison of the classifier model proposed in this chapter, with the previous works available in the literature is shown in Table 3.8. Widely accepted performance parameters (S_E and F_1 score) are compared [70], [233]. The achieved average S_E and F_1 score for S_1 and S_2 sounds under the training and test data were used for comparison.

It can be seen in Table 3.8 that the proposed approach achieved higher mean sensitivity for both S_1 and S_2 than the sensitivity achieved by authors in [71]. Similarly, the achieved overall average accuracy of 90.79%, was found to be higher than the average accuracy reported in [253], and [254]. This work was also compared to the recently proposed deep neural network (DNN) based approach by Chen *et al*. [75] that outperformed well-known classifiers, such as *K*-nearest neighbour, Gaussian mixture models, logistic regression and support vector machines. Although a same level of performance (average accuracy 90.79% versus 91.12% [75]) was achieved, the algorithm was tested with a small dataset of 460 S_1 and 460 S_2 sounds under controlled conditions. Similarly, authors in [235] achieved an average S_E of 93.24% for

S_1 sounds, however, the algorithm was tested using a small dataset of 30 recordings. To achieve robust performance, it is important to validate algorithms using a large database of recordings, and in real use scenarios. For instance, more recently, authors in [232] achieved an average sensitivity of 99.12% (and accuracy of 99.49%) in classifying S_1 and S_2 sounds, using a small on-line dataset comprised of 358 S_1 and S_2. However, the sensitivity of the approach dropped to 93.01% (and accuracy dropped to 92.23%) when tested with the data comprised of 93 S_1 and 93 S_2 sounds, recorded under controlled conditions. Furthermore, articles [233] and [70] achieved a higher F_1 score than the score achieved in this work, using recordings from different auscultation sites, however, the duration of recordings used to evaluate the performance of the algorithms was less than 5-minutes (30-40 seconds in [70] and 5-120 seconds in [233]). Unlike previous work [255] that suggests separate models for the classification of long-term and short-term recordings, the features extracted in this work were found suitable for varied duration of recordings. This was tested using short duration recordings obtained from the PASCAL

Table 3.8: Performance comparison with approaches for S_1 and S_2 identification.

Ref.	Method	Dataset	F_1 (%)	S_E (%)	Remarks
This work	Acoustic features with supervised learning	17578 S_1 and 17544 S_2	S_1: 90.65 S_2: 89.42	S_1: 91.59 S_2: 89.78	♥
[75]	DNN using MFCCs	460 S_1 and 460 S_2	F_1, S_E and A_{CC}: 91.12 (S_1 and S_2)		♦
[233]	DRNN	84400 beats (2874 records)	S_1: 97.5 S_2: 95.0	S_1: 97.8 S_2: 95.4	♥
[232]	Spectral and Convex Hull Features	358 S_1 and S_2 (online) & 93 S_1 and 93 S_2 (recorded)	-	99.12% (S_1 and S_2)	♦
[70]	Logistic Regression HSMM	12181 S_1 and 11627 S_2	S_1: 96.95 S_2: 94.29	-	♥
[235]	Adaptive algorithm	207 cycles (30 records)	-	S_1: 93.24 (Only S_1)	♦
[71]	HF (Time-domain intensity envelopes)	100 records	-	S_1: 80.3 S_2: 77.5	♦
[253]	Multi-band wavelet energy	326 beats (16 abnormal conditions)	A_{CC} S_1: 91.47, S_2: 88.95		♦
[254]	VMD and Shannon energy	53 cardiac periods (10 records)	A_{CC}: 90.0 (S_1 and S_2)		♦

Abbreviation: DNN: Deep Neural Network, DRNN: Deep Recurrent Neural Network, HSMM: Hidden Semi-Markov Model, VMD: Variational Mode Decomposition, HF: Homomorphic Filtering, A_{CC}: Accuracy.

Note: '♦' represents a small dataset of S_1 and S_2 segments used for validation. '♥' represents a large database of recordings from different auscultation sites.

database and long-term recordings of 5-minutes each, acquired during this work. The performance of the algorithms was also tested using the labelled segments of S_1 and S_2 sounds from the different auscultation sites (4217 S_1 and 4209 S_2 from the aortic valve site, 4289 S_1 and 4285 S_2 from the tricuspid valve site, 4171 S_1 and 4163 S_2 from the mitral valve site, and 4148 S_1 and 4151 S_2 from the pulmonary valve site). Thus, the approach demonstrated in this work can potentially be used to classify S_1 and S_2 sounds in signals recorded in real-uncontrolled conditions from four auscultation sites. Also, unlike previous studies (such as [75]) that demonstrated the classification of S_1 and S_2 sounds, this work also presented the classification of fundamental heart sounds (S_1 and S_2) and artifacts present in the recorded signals. Therefore, this work contributes to the development of algorithms used to distinguish S_1 and S_2 sounds in long-term continuous recorded signals by identifying features suitable for the recognition of S_1, S_2, and the artifacts.

This work extracted features using Maximal overlap discrete wavelet transform (MODWT) in which the coefficients are not decimated as in the case of DWT, producing the same number of wavelet coefficients as the number of sample observations at each level of transform offering several advantages over DWT [224]. The extracted coefficients of MODWT multiresolution analysis are associated with zero phase filters and are aligned with the original time series (time-invariant). The redundancy of the MODWT coefficients increases the effective degrees of freedom on each scale and provides an efficient wavelet variance estimator [223].

Overall, the approach presented in this study, based on a novel feature set, inherently addresses various real-world challenges in the classification of S_1 and S_2, which are important for different clinical applications. And it can be useful in the context of many cardiac related clinical applications. For example, accurate identification of S_1 and S_2 sounds is necessary for the estimation of heart rate variability. Additionally, the information is useful for the calculation of the duration of systole and diastole intervals, and the cardiac cycle. Any abnormal pattern/occurrence of S_1 or S_2 sounds in a cardiac cycle can be treated as a possible indicator of cardiac disorder. Another potential application of accurate localization of S_1 and S_2 sounds is the identification of cardiac abnormalities such as bradycardia, tachycardia, and first or second degree atrioventricular block, which are traditionally identified using the QRS complex and *R-R* peaks in ECG signals as demonstrated in [256]. As per the knowledge gathered from the reviewed articles, this is the first study which uses recordings of 5 minutes duration from the different sites in non-controlled real-world conditions.

4 An Approach for Automatic Identification of Fundamental and Additional Sounds

The research presented within this chapter is an edited version of research previously published in:

> *Amit Krishna Dwivedi, and Esther Rodriguez-Villegas, "An Approach for Automatic Identification of Fundamental and Additional Sounds from Cardiac Sounds Recordings", IEEE EMBC (2019), pp. 6685-6688.*

4.1 Introduction

The discussion in Chapter 3 underlined that the accurate localization of sound peaks of interest creates basis for measurement of vital physiological parameters for the identification of additional sounds, including sounds due to the splitting of sub-components of fundamental heart sounds – first heart sounds (S_1) and second heart sounds (S_2). As mentioned in the previous chapter, S_1 and S_2 sounds are the ones mainly observed in a phonocardiogram of a clean acoustic signal recorded from a healthy subject. S_1 is the result of the mechanical activities of the mitral and tricuspid valves; while S_2 results from the activities of the aortic and pulmonary valves. During normal cardiac conditions, S_1 appears as a single sound with discrete subcomponents M_1 and T_1; while S_2 appears as a single sound with discrete subcomponents A_2 and P_2, in the phonocardiogram. Although there are two discrete subcomponents in each of them (corresponding to the operation of the different valves), these are hard to distinguish because of the very small time interval existing between the occurrences of the individual mechanical cardiac events [17], [217]. Hence, S_1 sound (a single peak made of M_1 and T_1) and S_2 sounds (a single peak made of A_2 and P_2) are the fundamental heart sounds.

The maximum interval between the split sounds is generally shorter than 50ms [89]. However, the time difference widens when there is malfunctioning of the heart valves [257]. Thus, a variation in the existing systematic time difference, also termed as splitting, between the discrete sub-components potentially reflects the physiological changes including the malfunctioning of heart valves or mechanical activities such as during inspiration and expiration process. For instance, during inspiration, the two transient sub-components are

widely separated and come closure during expiration.

Clinically, the splitting of S_1 sounds into its subcomponents, because of reasons other than the respiratory cycle, may be evidence of atrial spectral defects (ASD) or right bundle branch block (RBBB). Similarly, the splitting of S_2 sounds may be an indicator of left bundle branch block (LBBB), ASD, RBBBs, left ventricular ectopic beats and pulmonary stenosis [16], [208], [258]. During cardiac abnormalities, apart from the split sounds, additional cardiac lub/dub sounds may be observed in the phonocardiogram, which can provide vital information about the cardiac conditions and assist in the early-stage detection of various cardiovascular diseases (CVDs). The presence of additional cardiac lub/dub sounds creates challenges in the identification of S_1 and S_2 sounds and subsequent estimation of cardiac cycle. Consequently, the automatic detection of fundamental and additional heart sounds has been a topic of great interest among researchers. However, as heart sounds are often overlapped with high-frequency acoustic signals unrelated to the operation of the heart, such as noise and artifacts, their correct interpretation is challenging. This also creates challenges when applying advanced signal processing methods to automate the process, especially, in real, non-controlled, environments where the number of disturbances present in the signal acquired is much higher.

This chapter presents an approach to automatically identify sound peaks of interest from cardiac auscultatory recordings without the need of any additional ECG or pulse carotid signal; identify normal fundamental cardiac sounds; and differentiate normal fundamental heart sounds recordings from the recordings with abnormal split sounds and additional lub/dub sounds in the systolic and diastolic intervals.

4.2 An Approach to Analyse Split of S_1 and S_2 Sounds

As heart sounds can potentially provide solutions to long term continuous monitoring suitable for wearable applications, various approaches have been suggested to identify sub-components of S_1 and S_2 sounds, and to estimate vital physiological parameters helpful in assessing the overall cardiac condition of a subject. One of the possible ways to diagnose cardiac abnormalities due to the splitting of the discrete sub-components of S_1 and S_2, is to evaluate the time interval between the sub-components M_1 and T_1 in S_1 sounds, and similarly, the time interval between the occurrence of sub-components A_2 and P_2 in S_2 sounds [259]. Here, the time (or split) interval is also termed as time delay between the discrete sub-components [16]. Thus, a quantitative measurement of split intervals of S_1 and S_2 would be of interest. For instance, the separation between A_2 and P_2 may be of fixed duration during both, inspiration, and expiration, indicating possibilities of atrial spectral defects (ASD). Proper identification of sub-components of S_1 and S_2 sounds can also assist in distinguishing third heart sound (S_3) or fourth

heart sound (S_4) which appears just before the start of S_2 sounds in a phonocardiogram. Thus, the split interval (or time delay) between the sub-components of S_1 and S_2 sounds needs to be investigated to ascertain the several possible cardiac conditions or abnormal behaviour of cardiac system. This demands accurate localization of sub-components of S_1 and S_2 sounds in cardiac acoustic recordings.

While, the mechanisms related to the origin of the two subcomponents of S_2 sounds are widely accepted, the mechanisms leading to the genesis of S_1 sounds are hard to determine and hence, the contribution of each of its components is still unknown [16], [17]. Additionally, M_1 and T_1 of S_1 sound lobes are more closely located in a phonocardiogram compared to the A_2 and P_2 of S_2 sounds. Due to this, the identification of split components of S_1 is even more complex compared to its counterpart [260]. That is why most of the studies are focused on the quantitative measurement of split of second heart sounds (S_2) in which the overlap of A_2 and P_2 is comparatively wider to identify. On the other hand, approach leading to the identification of S_1 splitting are still not well studied.

Previous studies identified the discrete subcomponents of S_1 and S_2 using instantaneous amplitude and instantaneous phase information [16], perceptual features with *k*-means algorithm [75], linear and non-linear transient chirp models [17], damped sinusoid models [261], [262], and matching pursuit methods [263]. Following a non-linear chirp based method, fundamental heart sounds (S_1 and S_2) are expressed mathematically as [217], [258]:

$$S_1(t) = A_{M_1}(t)\, sin\left(\varphi_{M_1}(t)\right) + A_{T_1}(t - \delta_{S1})\, sin\left(\varphi_{T_1}(t - \delta_{S1})\right) \quad 4.1$$

$$S_1(t) = A_{M_1}(t)\, sin\left(\varphi_{M_1}(t)\right) + A_{T_1}(t - \delta_{S1})\, sin\left(\varphi_{T_1}(t - \delta_{S1})\right) \quad 4.2$$

where, $A_{\mathrm{M}_1}(t)\, sin\left(\varphi_{\mathrm{M}_1}(t)\right)$ and $A_{\mathrm{T}_1}(t - \delta_{\mathrm{S1}})\, sin\left(\varphi_{\mathrm{T}_1}(t - \delta_{S1})\right)$ represent the amplitude and phase of M_1 and T_1 components of S_1 sounds, respectively. Likewise, $A_{\mathrm{A}_2}(t)\, sin\left(\varphi_{\mathrm{A}_2}(t)\right)$ and $A_{\mathrm{P}_2}(t - \delta_{\mathrm{S2}})\, sin\left(\varphi_{\mathrm{P}_2}(t - \delta_{\mathrm{S2}})\right)$ represent the amplitude and phase of A_2 and P_2 components of S_2 sounds, respectively. Additionally, δ_{S1} and δ_{S2} are used to indicate the delay (or split) between the subcomponents of S_1 and S_2, respectively.

In this work, consecutive S_1 and S_2 sounds were aligned and plotted on top of one another (as shown in Figure 4.1), to extract statistical information of sound lobes. These were extrapolated to obtain the characteristics of fundamental heart sounds in a cardiac signal. As S_1 and S_2 sounds are quasi-cyclic stationary with almost fixed duration in consecutive cycles for a particular subject [42], a proper estimation of the spectral width of lobes helps in assessing the time intervals of S_1 and S_2 and the delay (δ_{S1} and δ_{S2}), which can be compared with the normal intervals (estimated during normal cardiac conditions) to ascertain the splitting of heart sounds.

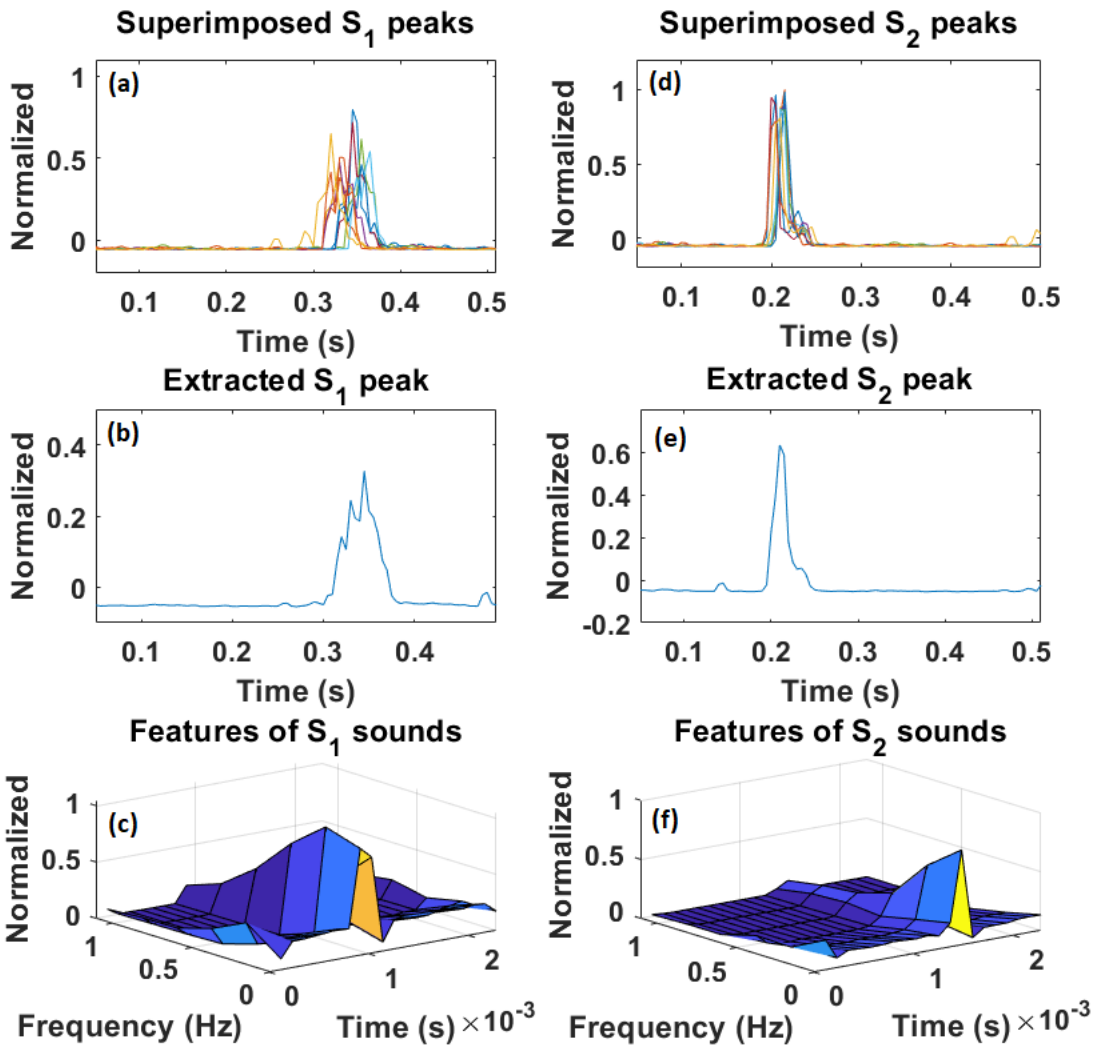


Figure 4.1. S_1 and S_2 sounds aligned to extract features based on the consistency of S_1 and S_2 sounds. (a) and (d) show 22 S_1 and S_2 sounds from the database, aligned and plotted on top of one another to extract S_1 and S_2 sounds as in (b) and (e), respectively. (c) and (f) show the extracted features of S_1 and S_2 sounds, respectively.

The algorithm presented in this chapter, extracted the onset and offset measurements of each sound lobes of interest, which then were used to accurately determine the beginning and end of S_1 and S_2 sounds based on the exact locations of S_1 and S_2 sounds available from the PASCAL Classifying Heart Sounds Challenge database [23]. Features to identify the split sounds were based on the consistency of S_1 and S_2 sounds. This was previously investigated by Tang *et al* [42]. Consecutive S_1 and S_2 sounds were aligned and plotted on top of one another to extract features. The onset and offset measurements obtained as the intersection of the Shannon envelope of each peak with the zero crossings were used to estimate the timing intervals of S_1 and S_2 sounds. This feature facilitated the identification of splitting sounds by evaluating the change in the intervals, with respect to the normal cardiac conditions, of the consecutive cardiac

cycles.

As the duration of S_1 and S_2 sounds (without splitting of subcomponents) is known to range between 80ms to 150ms during normal cardiac operation, sound lobes with shorter or longer durations were investigated as potentially abnormal heart sounds [208]. The duration and energy of adjacent peaks were also evaluated, as S_1 and S_2 sounds cannot have the same energy under normal cardiac conditions [89]. Hence, if two adjacent peaks were found to have the same energy, then one of the two sounds was considered as abnormal. The intervals between the two adjacent S_1 and S_2 peaks represent the systolic and diastolic intervals, where the diastolic period under normal cardiac operation is greater than the systolic period. The start of S_1 sound marks the beginning of the systolic period and extends up to the beginning of S_2 sound; while the diastolic period is marked as the starting of S_2 sound and extends up to the starting of S_1 sound. The energy of sound lobes identified in the systole or diastole intervals regions was also evaluated to distinguish the additional sound peaks from the fundamental heart sounds.

In case of splitting of S_1 or S_2, the interval between the two discrete consecutive sound lobes would be less than 50ms and both sound lobes would have different energies [89]. This is because one of the split components of S_1 or S_2 will have a lower intensity compared to the another component [89]. Furthermore, the consecutive sound lobes with time interval greater than 50ms, were investigated for abnormal/additional lub/dub sounds. Statistical features evaluated were the mean, standard deviation of skewness, mean absolute deviation, root mean square and root sum of squares of successive lobes and kurtosis from each analysis window. Since the possibility existed that split sounds of S_1 and S_2 might get confused with other abnormal heart sounds, perceptual features i.e. mel-frequency cepstral coefficients (MFCCs) were also extracted to capture the non-linear behaviour of the signals using higher order decomposition of transients. Steps followed to extract various features used in this work are described in detail.

4.3 Segmentation of Heart Sounds Signals

As discussed in Chapter 3, the first step in the development of the algorithm was to localize peaks in the recorded signal. The temporal localization of all peaks would, later on, be used to assist in the identification of S_1 and S_2 sounds as well as other abnormal sounds. This helped in extracting features to assist the classification of normal and abnormal recordings. The steps followed are discussed briefly as a detailed discussion has been already provided in Chapter 3.

4.3.1 Pre-processing

The amplitude of the heart sounds was normalized to a fixed scale of [-1, 1] dividing by its

absolute maximum:

$$Y[n] = \frac{y'[n]}{\max_n(|y'[n]|)} \tag{4.3}$$

where, $Y[n]$ was the normalized signal and $\max_n(|y'[n]|)$ represented the absolute maximum of the resampled and filtered signal y'[n]. Further, a low pass filter with a cut off frequency of 500 Hz was employed, to eliminate high-frequency components from the normalized signals, which are typical of artifacts. The frequency spectrum of the recordings from the database was used to estimate this cut-off frequency (as shown in Figure 4.2).

4.3.2 Identification of sound peaks of interest

As wavelets are rapidly decaying oscillations with zero mean, and well-suited for non-stationary signal analysis, wavelet analysis was chosen to obtain the time-frequency domain resolution of the cardiac sound signals. A 6th order Daubechies wavelet with 7-levels of decomposition and reconstruction was used. A windowing step was included to split the signal

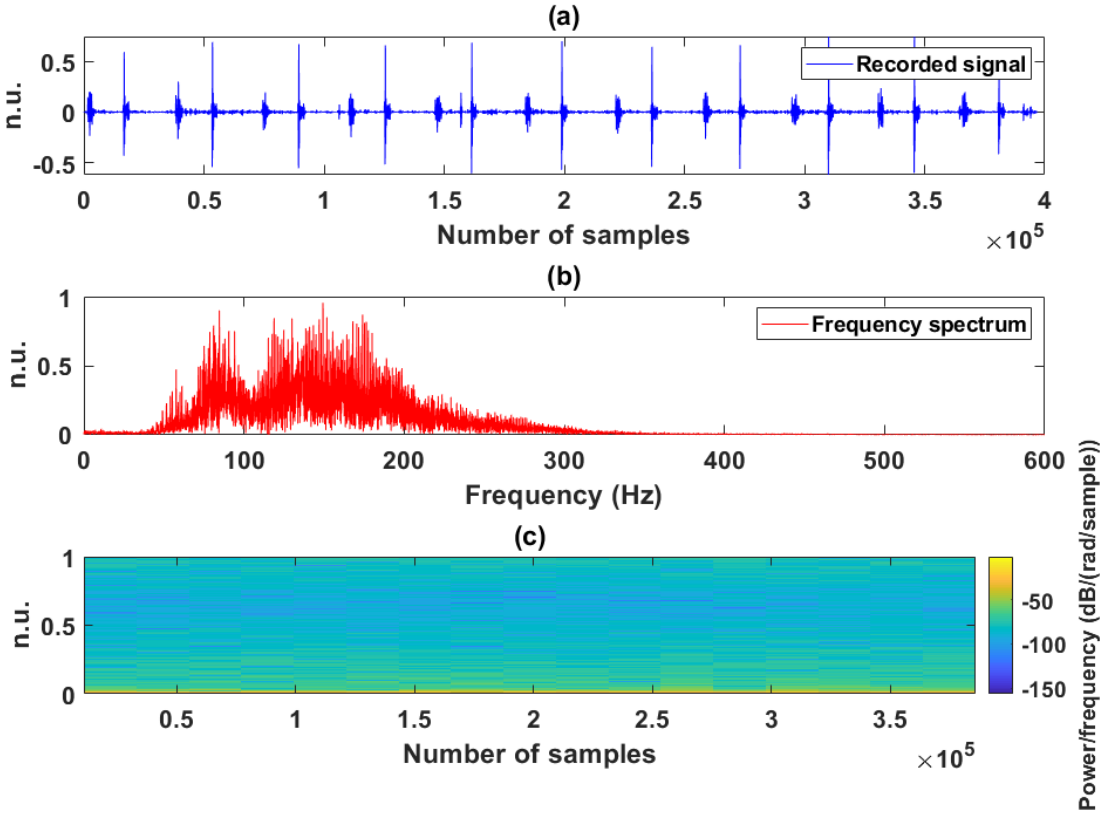


Figure 4.2. (a) Heart sounds signal from the PASCAL database. (b) Frequency spectrum of the signal using the Fourier analysis. This was used to estimate the cut-off frequency of the filter used to remove noise. (C) Spectrogram of the recorded signal.

into a sequence of frames for the analysis. A 10ms window was used to extract the envelope of the pre-processed signal using the average Shannon-energy (E_S):

$$E_S = \frac{-1}{N}\sum_{i=1}^{N} Y^2[n](i).logY^2[n](i) \quad 4.4$$

where, N represented the number of normalized samples in the selected window. The sound lobes of interest were localized using the zero crossing points of the normalized Shannon-energy $E_{Snorm}[n]$, extracted as:

$$E_{Snorm}[n] = \frac{E_s[n] - \overline{E_s[n]}}{\sigma(E_s[n])} \quad 4.5$$

where, $E_S[n]$ was the average Shannon energy and $\overline{E_S[n]}$ and $\sigma(E_s[n])$ represented the mean and standard deviation of $E_s[n]$, respectively. An adaptive wavelet-based filtering approach described in the Chapter 3 (Section 3.2.1) was used to separate non-stationary components in recorded signals. As an illustration, the Shannon energy envelope extracted using a recording from the PASCAL database is shown in Figure 4.3.

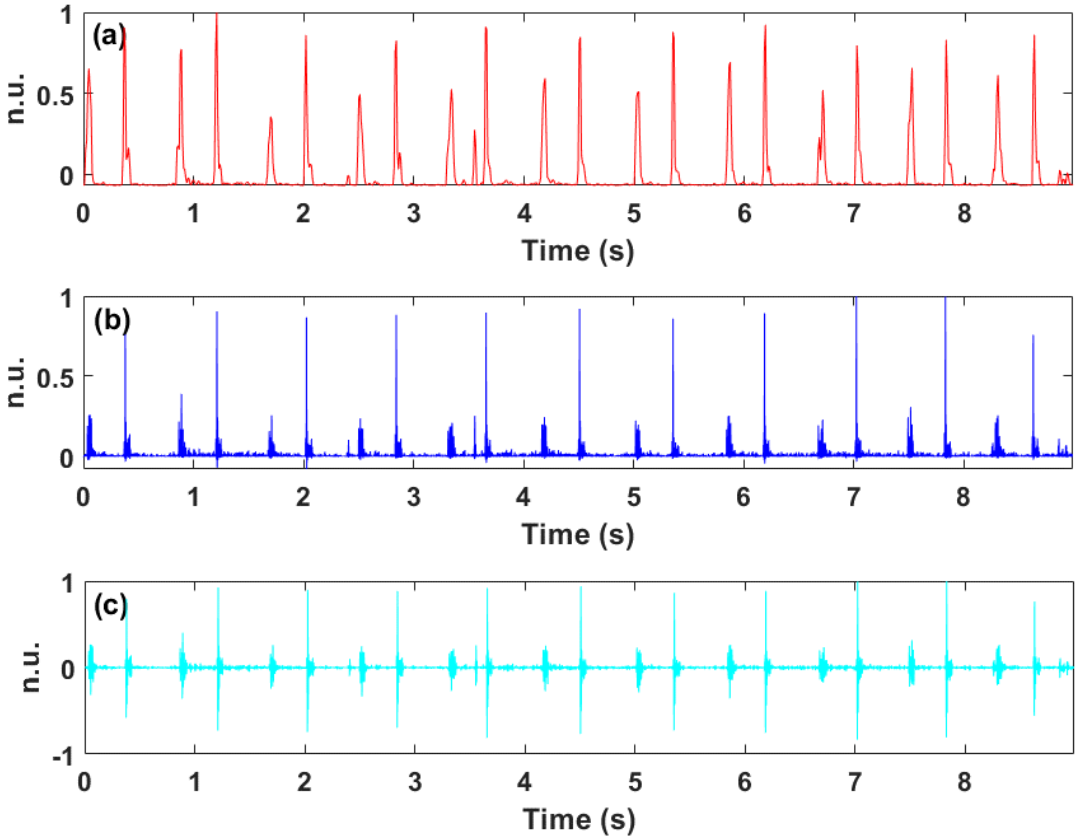


Figure 4.3. (a) Shannon energy envelope of the signal shown in (c) (from the PASCAL database). (b) shows the filtered wavelet-based coefficients to extract envelope.

Once the peaks of interest were found, the algorithm extracted the onset and offset measurements (start and end point) of each sound lobes of interest, which were then used to accurately determine the beginning and end of S_1 and S_2 sounds. The reference locations of S_1 and S_2 sounds available from the PASCAL Classifying Heart Sounds Challenge database [23] were used. An example of how the transients were localized is shown in Figure 4.4, in which the sound lobes of S_1 and S_2 were identified clearly in the signal obtained from the CHSC database [23]. These identified S_1 and S_2 sounds were only used to study the characteristics of the fundamental heart sounds and to help in the identification of potential features, as described in Section 4.2. Thus, the performance of the classifier is independent of the location of S_1 and S_2 sounds, and there is no need of auxiliary signals for the identification.

The algorithm presented in this chapter identified the sound peaks of interest (i.e. S_1 or S_2 sounds) that were used to extract features. An example of how the transients were localized is shown in Figure 4.5, in which sound lobes of S_1 and S_2 were identified clearly in the signal obtained from the CHSC database [23] that contained additional lub/dub sounds. Similarly, as an illustration, Figure 4.6 shows the localization of sound lobes in an acoustic signal recorded using an electronic stethoscope [213] in a real, non-controlled environment.

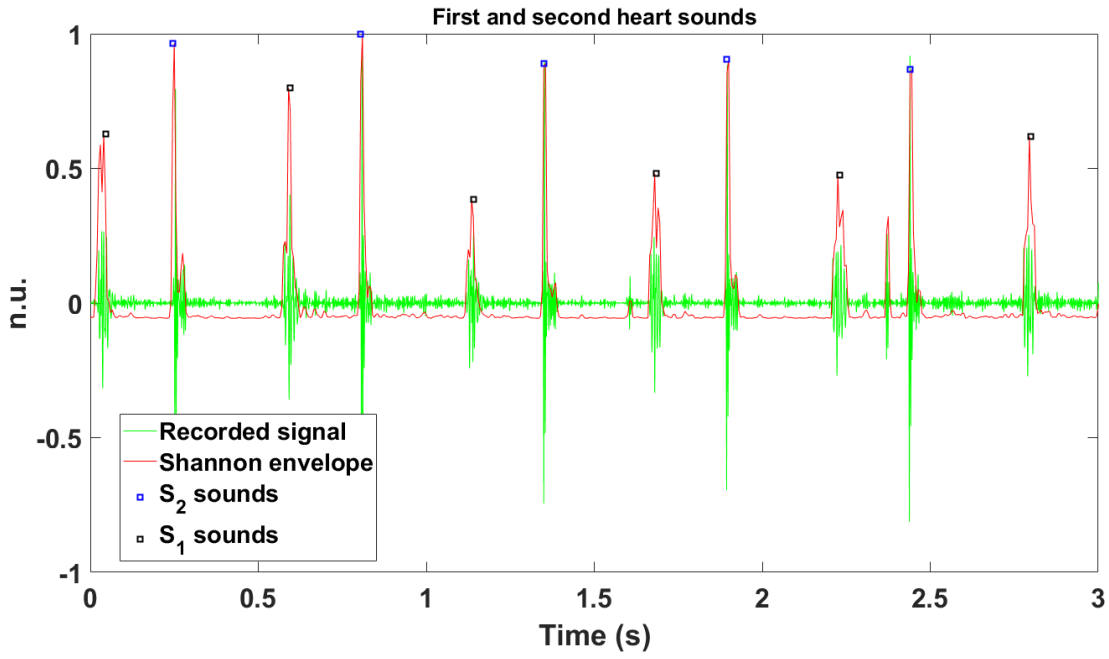


Figure 4.4. Localization of sound peaks of interest and S_1 and S_2 sounds identification using the reference locations.

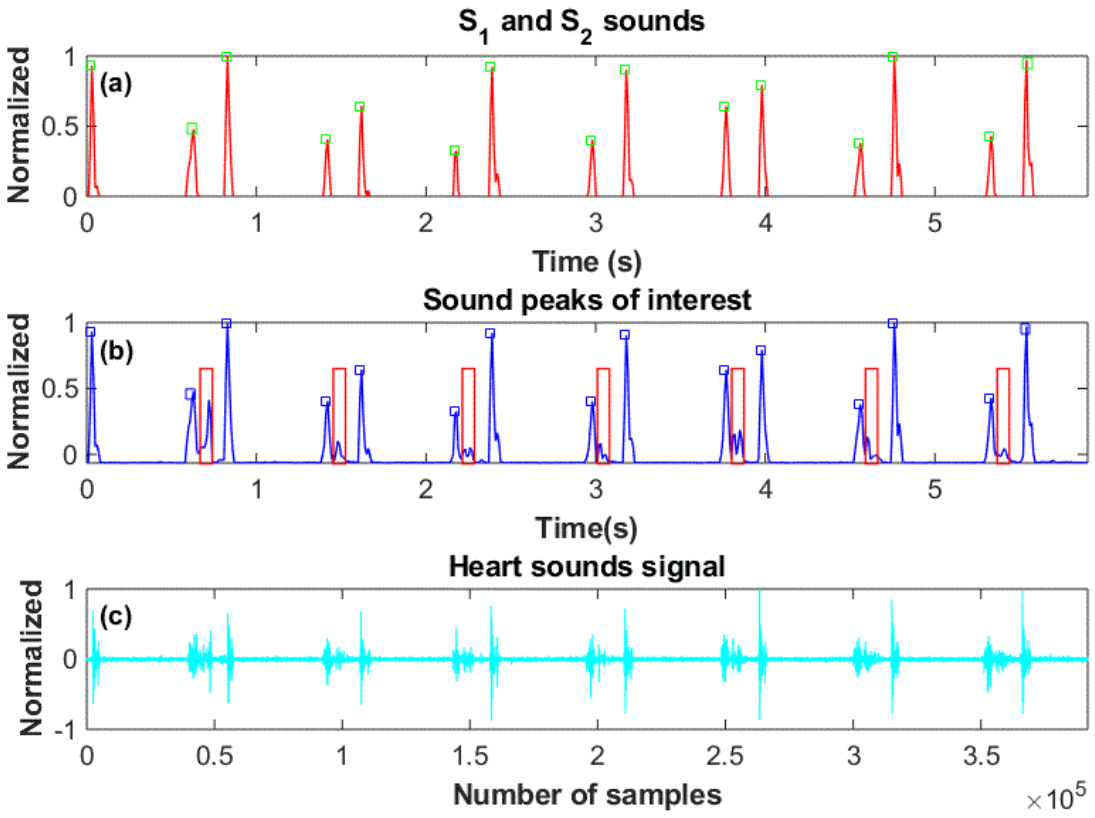


Figure 4.5. Identification of S_1 and S_2 sounds in the heart sounds signals with additional lub/dub sounds obtained from the database. (c) and (b) show the recorded signal and the identified sound lobes of interest, respectively. (a) shows the potential S_1 and S_2 sounds after discarding the identified abnormal sounds.

4.4 Feature Extraction & Selection

4.4.1 Feature extraction

Once sound lobes of interest in cardiac signals were localized using the wavelet filtering and Shannon energy estimation, features were extracted from labelled segments to try to differentiate recordings with normal fundamental heart sounds from additional heart sounds, corresponding to abnormalities. In Chapter 3, the features that demonstrated the greatest potential in differentiating fundamental heart sounds – S_1 and S_2, were used in this study for the classification of recordings with normal fundamental heart sounds and recordings with additional lub/dub sounds. Additionally, statistical features were also extracted from localised heart sounds. 11 features (denoted as F_{27}, …, F_{37}) represented the energy across the scales and scaling coefficients extracted using the Maximum Overlap Discrete Wavelet Transform (MODWT), and 11 features (denoted as F_{38}, …, F_{48}) were extracted to represent the energy

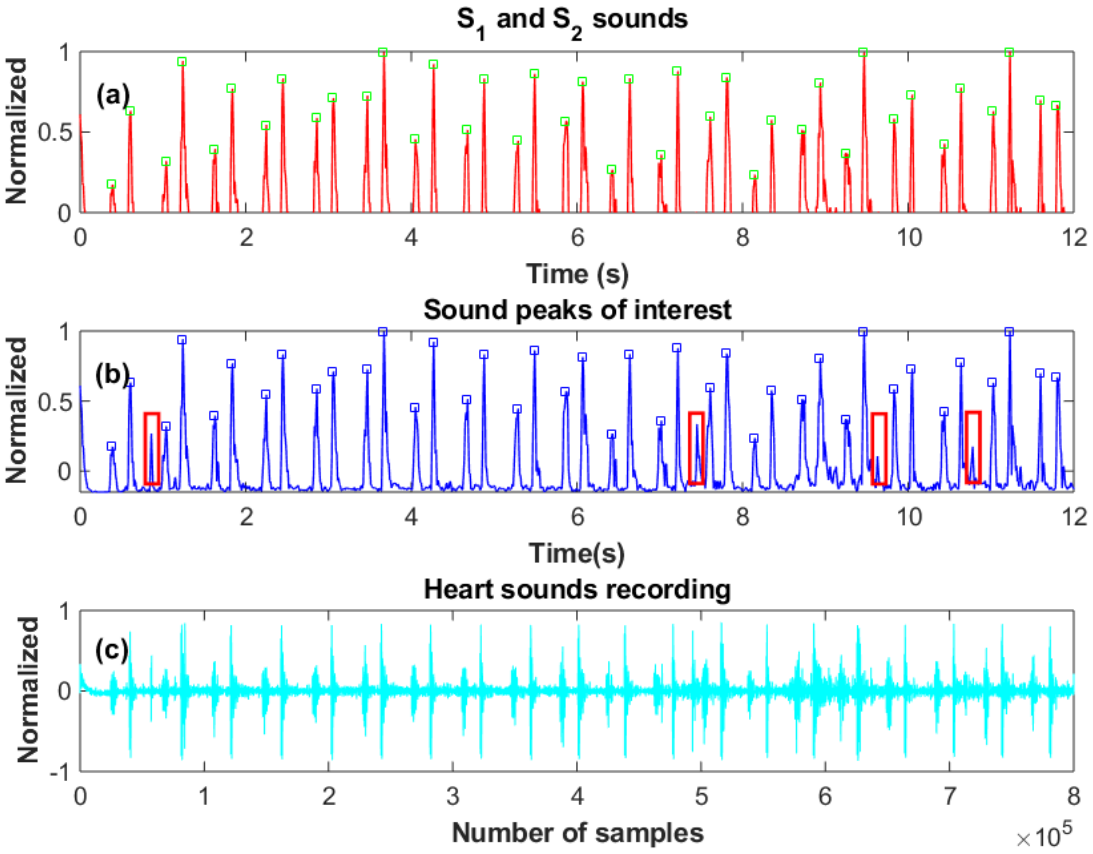


Figure 4.6. Identification of S_1 and S_2 sounds in a heart sounds signals recorded in an uncontrolled environment using an electronic stethoscope. (c) and (b) show the recorded signal and the identified sound lobes of interest, respectively. (a) shows the potential S_1 and S_2 sounds after discarding the identified abnormal sounds.

across the scales using the multiresolution analysis (MRA) of the MODWT matrix. Thus, a total of 22 features were extracted using the MODWT analysis of the signals. 3 features (F_{49}, F_{50} and F_{51}) representing the peak amplitude, maximum frequency, and its corresponding value (index) for each window using a first-order Butterworth low-pass filter with cut off frequency 10 Hz, were extracted using the homomorphic filtering. As an illustration, Figure 4.7 shows different envelopes extracted in this study for feature extraction. Apart from these, 3 features denoted as F_{52}, F_{53} and F_{54} were extracted to capture the peak amplitude, maximum frequency, and its corresponding value (index) in the spectrum, of the constructed Hilbert envelope for each window. 6 features (F_{55}-F_{60}) were the energy computed for the first 5 IMFs and the residual part. 2 features from each IMF were also extracted to represent the maximum frequency and its corresponding index value in the spectrum. This produced 8 features represented as F_{61}-F_{68}, from 4 selected IMFs (IMF2, IMF3, IMF4, and IMF5). Figure 4.8 shows how these IMFs were selected for the analysis. Later, a *k*-means cluster algorithm was used to cluster the

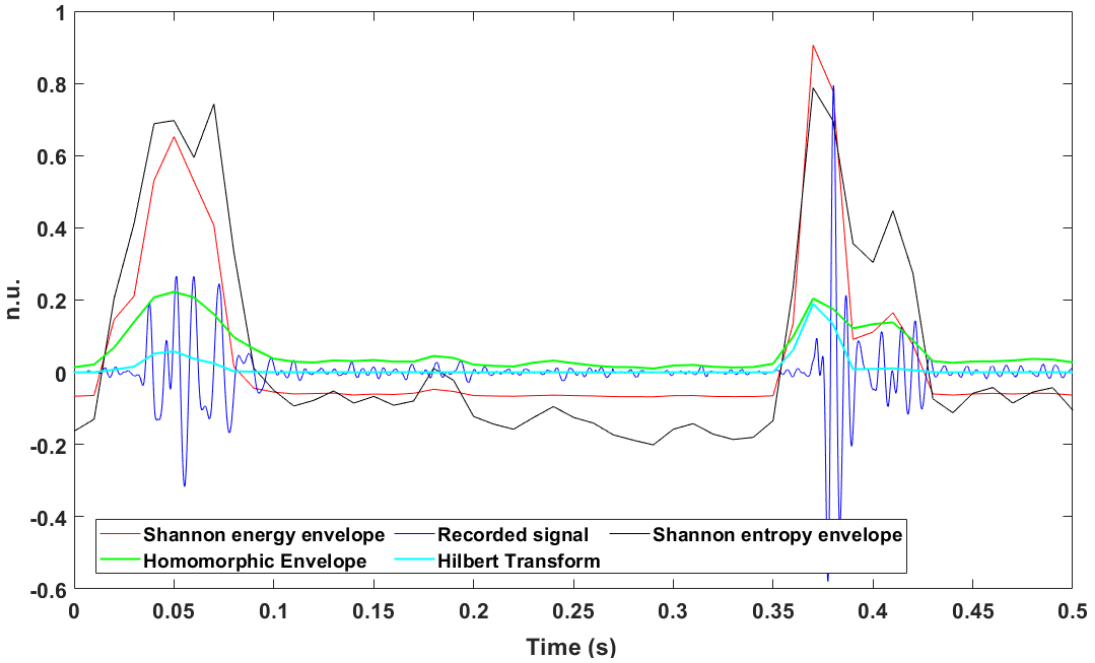


Figure 4.7. Envelopes extracted from a recorded signal.

instantaneous frequency and the instantaneous energy of the 4 modes (IMF2, IMF3, IMF4, and IMF5) that produced 2 features from each mode by keeping the number of modes equal to the number of clusters in the algorithm. As an illustration, the instantaneous frequency and instantaneous energy of the signal for different IMFs are shown in Figure 4.9. Here, IMF1 was eliminated as it corresponded to high-frequency components; and only IMF2, IMF3, IMF4, and IMF5, containing meaningful information, were selected for the feature extraction [215]. Furthermore, 15 Mel-frequency cepstral coefficients (MFCCs) features (F_{85} to F_{99}) were also computed. All extracted features, used in this study, are listed in Table 4.1 and Table 4.2, where, $h[n]$ and $\overline{h[n]}$ represent the signal and it's mean. N is the length of the signal. $H(i)$ is the discrete Fourier transform of $h[n]$.

4.4.2 Feature selection and reduction

The extracted features were statistically analysed to derive the top 10 features that achieved the highest weights and demonstrated the greatest ability to differentiate recordings with normal fundamental heart sounds from the additional lub/dub sounds in the systole or diastole interval. These features were selected as optimum features for the classification. Feature weights were computed using a diagonal adaptation of neighborhood component analysis (NCA) [227]. Weights were allocated to each feature by maximizing the expected leave-one-out classification accuracy [219]. As demonstrated in Chapter 3 [215], the name-value pair argument '*Lambda*' representing the regularization parameter was tuned against the minimum classification loss

using cross-validation to find the optimum value. The optimum value of *Lambda* was found to be 0.021×10^{-3} that corresponded to a minimum classification loss of 0.017.

The top 10 features identified using the weights from the neighborhood component analysis were: F_{25} (deterministic energy of S_1 and S_2 sounds), F_{26} (non-deterministic energy of the systole and diastole intervals), F_{32} (energy of MODWT decomposition coefficients at level 6), F_{43} (energy at level 6 computed using MRA of MODWT matrix), F_{61} (maximum frequency of

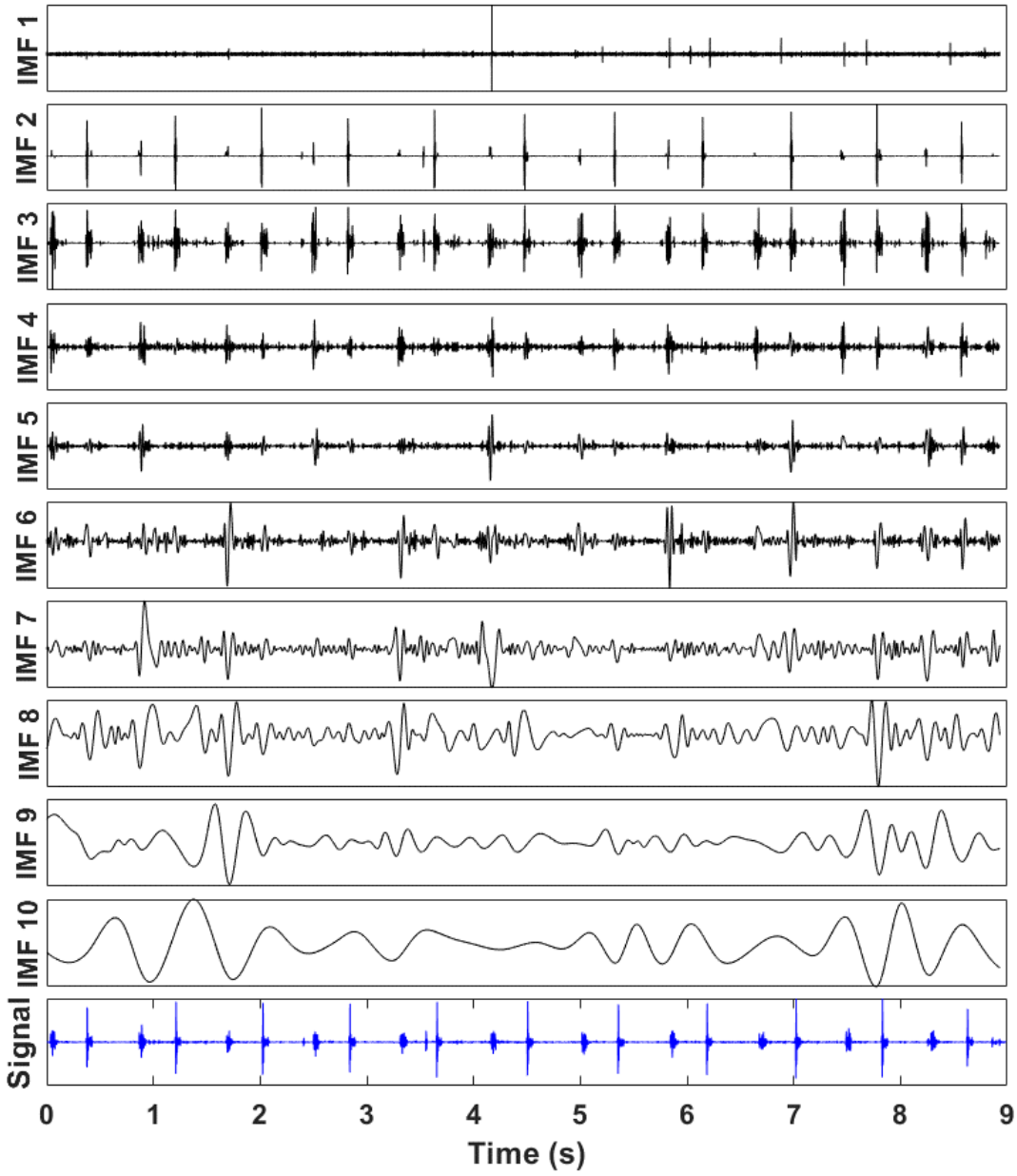


Figure 4.8. The different intrinsic mode functions extracted from a heart sounds recording using EMD. IMF1 contains high-frequency components, hence, only IMFF2, IMF3, IMF4 and IMF5 were used to obtain the cluster of the instantaneous frequency and instantaneous energy of the signal. (Y-axis is in the normalised unit.)

IMF1 using EMD), F_{63} (maximum frequency of IMF2 using EMD), F_{69} (instantaneous frequency of IMF2 using HHT), F_{73} (cluster of instantaneous frequencies of IMF2), F_{74} (cluster

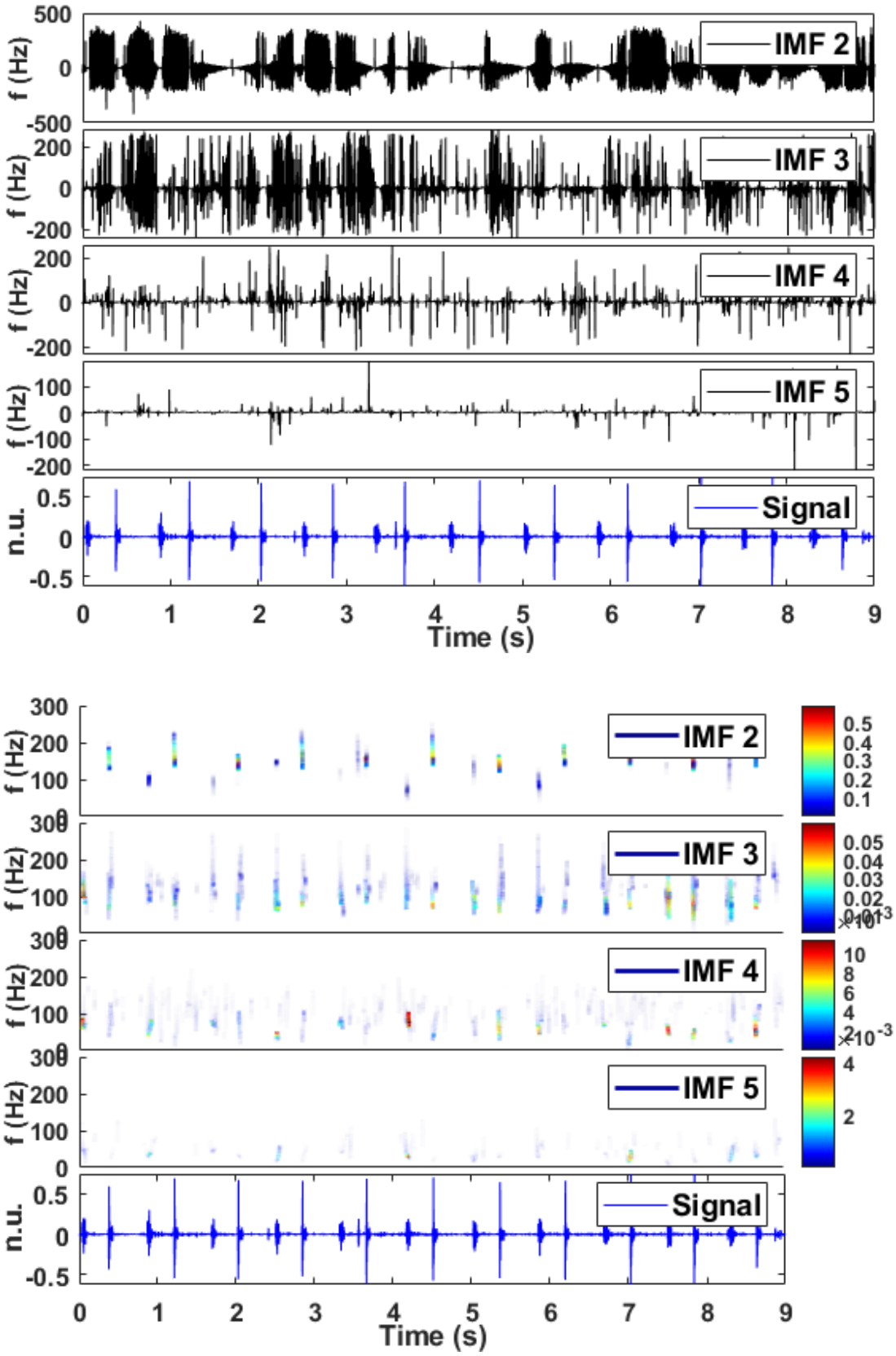


Figure 4.9. (a) The instantaneous frequency (first figure), and (b) instantaneous energy (second figure), for the different IMFs selected using a signal, for feature extraction.

of instantaneous frequencies of IMF3), and F_{75} (cluster of instantaneous frequencies of IMF4). Among all the features selected, F_{73} achieved the highest weight in classifying recordings with normal fundamental heart sounds, additional lub/dub sounds in systole or diastole interval, and artifacts.

Table 4.1: Various features extracted and their description (F_1- F_{36}).

Feature#	Type	Feature	Description
F_1	Time-domain features	P_{1T}	Mean of the absolute value of P_1 interval
F_2		P_{2T}	Mean of the absolute value of P_2 interval
F_3		P_{1T}/P_{2T}	Ratio of P_1 and P_2 intervals
F_4		$T_{P1\text{-}P2}$	Interval between the beginning of P_1 and P_2 sounds (Systole interval)
F_5		$T_{P2\text{-}P1}$	Interval between the beginning of P_2 and the next P_1 sounds (Diastole interval)
F_6		$T_{P1\text{-}P1}$	Interval between the two adjacent P_1 sounds
F_7		$T_{P2\text{-}P2}$	Interval between the two adjacent P_2 sounds
F_8		$T_{P1\text{-}P2}/T_{P2\text{-}P1}$	Ratio of Systole to diastole intervals
F_9		$T_{P1\text{-}P2}/P_{1T}$	Ratio of systole interval to P_1 interval
F_{10}		$T_{P2\text{-}P1}/P_{2T}$	Ratio of diastole interval to P_2 interval
F_{11}	Statistical features	Mean	$\overline{h[n]} = \frac{1}{N}\sum_{i=1}^{N} h[i]$
F_{13}		Standard deviation	$std = \sqrt{\frac{\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2}{N-1}}$
F_{18}		Skewness	$skw = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^{3/2}}$
F_{19}		Kurtosis	$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^2}$
F_{20}	Frequency domain-based features	Spectral Entropy	$S_E = \frac{\sum_{i=1}^{N} P_i . \log P_i}{\log_2 N}$
F_{21}		Dominant frequency	Frequency corresponding to the maxima of power spectrum
F_{22}		Maximum frequency	$f_{max} = \max(frequency)$
F_{24}	Statistical features	Zero-crossing rate	$Z_R = \frac{1}{2N}\sum_{i=1}^{N}\left\lvert sgn\left[h[i]\right] - sgn\left[h[i-1]\right]\right\rvert$
F_{25}		Energy	Deterministic energy of S_1 and S_2
F_{26}			Non-deterministic energy of the systole and diastole intervals
$F_{27} - F_{36}$	MODWT	Energies at different scales	9-levels of decomposition, details and approximation coefficients produced 10 features

Table 4.2: Various features extracted and their description (F_{37}-F_{99})

Feature#	Type	Feature	Description
F_{37}	MODWT	Energy	Total energy of all the scales
$F_{38} - F_{47}$	MRA of MODWT matrix	Energies at different scales	10 features extracted from the MODWT matrix
F_{48}	MRA	Energy	Total energy of all the levels
$F_{49} - F_{51}$	Homomorphic envelope	Frequency and amplitude	The dominant frequency of the envelope, spectral value, and peak of the envelope
$F_{52} - F_{54}$	Hilbert spectrum	Frequency and amplitude	
$F_{55} - F_{60}$	EMD	Energy	Energy of first 5 IMFs and the residue
$F_{61} - F_{68}$	EMD	Frequency and magnitude	Maximum frequency in the spectrum and its corresponding value in the spectrum form 4 IMFs
$F_{69} - F_{76}$	Hilbert-Huang transform	Instantaneous frequency	IF and cluster of IFs for each mode for 4 selected IMFs (8 features)
$F_{77} - F_{84}$	Hilbert-Huang transform	Instantaneous energy	IE and cluster of IEs for each mode for 4 selected IMFs (8 features)
$F_{85} - F_{99}$	Perceptual features	MFCCs	15 features from each segment

4.5 Classifier Training and Testing

The classifier was trained using the features obtained from the true known categories from the PASCAL database available from the Classifying Heart Sound Challenge (CHSC) database [23]. This helped in extracting robust global parameters of the model. The details of the recordings used to test the performance of the algorithm developed for the identification of additional sounds in the cardiac cycle are provided in Table 4.3. Among the 231 recordings with normal fundamental heart sounds, 111 recordings marked with the exact locations of S_1 and S_2 sounds were used to study the characteristics of fundamental heart sounds. This helped in identifying potential features to assist the classification of normal and abnormal recordings

4.5.1 Training and test datasets

The feature vector was evaluated for each analysis window of 4 seconds. The window size was selected to capture at least two complete cardiac cycle in a recorded signal. This produced a feature matrix of dimensions $X \in R^{4296 \times 10}$ using 290 recordings (231 normal, 19 extra heart sounds, and 40 artifacts). The hold-out approach and 10-fold cross-validation were used to split the data (PCG instances) into training and test datasets. Thus, for hold-out, 80% of the data was used for the training and the remaining 20% of the data was used as a test dataset. This produced

Table 4.3: Cardiac acoustic recordings used in this study.

Source		Subjects	No. of Recordings	Comments
The PASCAL database [23]	Dataset A	Normal	31	21 records with locations of S_1 and S_2
		Normal/abnormal	19	Extra heart sounds
		Normal/abnormal	40	Artifacts
	Dataset B	Normal	200	90 records with locations of S_1 and S_2

a training feature matrix $X_{training} \in R^{3437 \times 10}$ and test feature matrix $X_{test} \in R^{859 \times 10}$, where, 3437 and 859 were the number of PCG instances or windows used for training and testing from the database, respectively. The training set contained 1568 PCG instances of artifacts, 563 of extra heart sounds and 1306 of normal fundamental heart sounds. Similarly, the test set contained 392 PCG instances of artifacts, 140 of extra heart sounds and 327 of normal fundamental heart sounds. These instances were used as a ground truth to evaluate the performance of the algorithm.

4.5.2 Model training and parameter optimization

The final step of the algorithm was to take the extracted features and fed them to an appropriate classifier to interpret them. In this work, classifier mainly used for the heart sounds analysis, found during the literature review were investigated. It was found that the ensemble of classifiers are used to achieve improved classification accuracy by combining the prediction of several learners [229]. The ensemble learning approaches have been used in recent studies for successful classification of cardiac auscultatory recordings [68], [230], [231]. It is for this reason the ensemble approach was used to derive a powerful classifier model with the objective to achieve improved classification performance and robustness. The feature vector evaluated in the previous section was used to train the ensemble classifier and to extract global parameters of the classifier (these steps are described in Chapter 3). Decision trees were used as weak learners due to their capacity to deal with high dimensional datasets. The boosting algorithm was used to optimize the weights and parameters of the weak learners at each iteration during the cross-validation [247]. Among several boosting algorithms (adaptive boosting (AdaBoost) [248], adaptive Logistic Regression (LogitBoost) [247], random undersampling (RUSBoost) [246], GentleBoost [247], BrownBoost [249] and XgBoost [231]), the RUSBoost algorithm was used to handle imbalanced data and the LogitBoost algorithm to minimize the binomial deviance loss function, while fitting the model parameters, and because of its robustness against noise [264].

In this work, the number of observations from recordings with normal fundamental heart

sounds, additional lub/dub sounds, and artifacts, used for training were found unbalanced (1568 PCG instances of artifacts, 563 of extra heart sounds and 1306 of normal fundamental heart sounds). For this reason, the random undersampling boosting (RUSBoost) algorithm was used to address the problem of training classifier on skewed data [246]. Using the RUSBoost algorithm helps in handling imbalanced data and it is better compared to the use of '*Prior*' name-value pair argument or 'Cost' function in the classifier. As the datasets used in this study contained recordings from three categories (normal fundamental heart sounds, extra heart sounds and artifacts), the classifier model was trained to predict these three classes using the selected features.

4.6 Results

4.6.1 Evaluation metrics

The evaluation metrics used to assess the performance of the approach were: recall or sensitivity $\left(S_E = \frac{TP}{TP+FN}\right)$, classification accuracy $\left(C_A = \frac{TP+TN}{TP+FP+FN+TN}\right)$, precision $\left(P_+ = \frac{TP}{TP+FP}\right)$, and F_1-score $\left(F = 2\frac{P_+ \times S_E}{P_+ + S_E}\right)$. Here, true positive (*TP*) is the correct results, false positive (*FP*) represents false alarm, false negative (*FN*) represents missed detection and true negative (*TN*) is the correct absence of result. As the dataset contain recordings from three categories, the performance was evaluated using a 3 × 3 (three-class) confusion matrix, denoted as $\boldsymbol{P}$ [220]. Table 4.4 shows the interpretation of the confusion matrix, where P_{ij} represents the number of instances where the predicted class was i and the ground truth (or actual class) was j. The number of *TP*, *TN*, *FP* and *FN* for each class k were computed as described in Chapter 3 (Section 3.5) of this thesis, for a 3 × 3 (three-class) confusion matrix.

4.6.2 Results

Table 4.5 shows the confusion matrices of the trained classifier on the training (using 10-fold cross-validation) and test sets from the PASCAL database. The performance parameters evaluated using the confusion matrices are shown in Table 4.6. The scores illustrated in the

Table 4.4: Three-class confusion matrix.

		Actual class		
		Artifacts	Extra Sounds	Normal
Predicted class	Artifacts	P_{11}	P_{12}	P_{13}
	Extra Sounds	P_{21}	P_{22}	P_{23}
	Normal	P_{31}	P_{32}	P_{33}

Table 4.5: Confusion matrices of the trained classifier for (a) training datasets and (b) test dataset.

(A)

		Actual class		
		Artifacts	Extra	N
Predicted class	Artifacts	1554	2	4
	Extra	2	501	45
	Normal	12	60	1257

(B)

		Actual class		
		Artifacts	Extra	N
Predicted class	Artifacts	379	1	3
	Extra	0	118	25
	Normal	13	21	299

table are the average values on all observations, counting each observation when it was in a held-out fold. The trained classifier model achieved an average P_+ of 95.20%, S_E of 94.77%, F_1 of 94.97%, and C_A of 97.57% using the training dataset, and, an average S_E of 90.79%, P_+ of 90.41%, F_1 of 90.59%, and C_A of 95.11%, under the unseen test dataset, in classifying recordings of three different classes.

4.7 Discussion

This chapter presents an approach to automatically differentiate between normal fundamental sounds, artifacts, and additional abnormal sounds; where the latter include both splitting of fundamental, S_1 and S_2 sounds, and additional lub/dub sounds in the diastolic and systolic periods. It can be observed from the performance table that a few recordings with additional lub/dub sounds (i.e. extra heart sounds) were wrongly classified as normal recordings, and vice versa. This is mainly because of the resemblance in the characteristics of the normal fundamental heart sound and addition lub/dub sounds. Thus, it creates challenges in the identification of S_1 and S_2 sounds, as well as the estimation of cardiac cycle.

As fundamental heart sounds in a normal subject follow a specific sequence of S_1 followed

Table 4.6: Performance of the classifier using 10-fold cross-validation.

		P_+ (%)	S_E (%)	F_1 (%)	C_A (%)
Dataset (training)	Artifacts	99.61	99.10	99.35	99.42
	Extra heart Sounds	91.42	88.98	90.18	96.83
	Normal	94.58	96.24	95.40	96.48
Average		*95.20*	*94.77*	*94.97*	*97.57*
Dataset (test)	Artifacts	98.95	96.68	97.80	98.02
	Extra heart Sounds	82.51	84.28	83.38	94.53
	Normal	89.79	91.43	90.60	92.78
Average		*90.41*	*90.79*	*90.59*	*95.11*

by systole, and S_2 followed by diastole, this information was helpful in extracting the expected state and duration of heart sounds in a cardiac cycle [70]. The intervals between the preceding and following peaks in an extracted heart sounds envelope provided sufficient information. This included the duration (or interval) of the localized peaks and intervals between the adjacent peaks, that may vary with respect to the cardiac abnormality.

The Shannon envelope of split sounds was found to be wider than the envelope of S_1 or S_2 sounds. Also, in normal conditions, the minimum separation between S_1 and S_2 was found to be 50ms. Thus, if two fundamental peaks were detected within an interval of less than 50ms, the peak with larger energy corresponded to the fundamental heart sounds (may be S_1 or S_2) and the left-over peak may be the split component of FHSs or abnormal heart sound. This is because it was important to include features that considered the variation of the intervals of the peaks based on cardiac conditions. Hence, the start and end points of sound peaks of interest, extracted during the localization, were used to estimate the duration of peaks and various peak to peak intervals. Thus, the indices of localized peaks were found useful in estimating the energy between the adjacent peaks (deterministic and non-deterministic energy).

The importance of features studied in this chapter was also investigated using weights achieved by each feature used to train the classifier. The top 10 features were: F_{25}, F_{26}, F_{32}, F_{43}, F_{61}, F_{63}, F_{69}, F_{73}, F_{74}, and F_{75}. The details of these features can be found in Appendix F. Among these selected features, the features representing the energy between the intervals (F_{25} and F_{26}) signify the importance of accurate localization of sound peaks of interest. In addition to these, features extracted using the Hilbert-Huang transform were also found promising in classifying recordings.

The performance results achieved under the unseen test dataset were used to compare with the previous works that have used the same dataset to classify normal, artifacts and extra heart sounds recordings. Results shown in Table 4.7 demonstrate that the presented approach achieved improved performance in classifying normal sounds and artifacts compared to [45], [91], [102]. Except the performance achieved by [91] (achieved precision of 100% for extra heart sounds), this work also demonstrates improved results compared to [45], [102], for extra sounds classification. Furthermore, compared to the previous work [217], the performance (in terms of precision (P_+)) was improved from 81.48% to 91.42% (training set) and 82.51% (test set), in classifying recordings with extra heart sounds (additional lub/dub sounds) with added features. Overall, the achieved results in this chapter demonstrate that the presented approach can potentially classify recordings with normal fundamental heart sounds from additional sounds, including sounds from sub-components of S_1 and S_2, without using any auxiliary signals, such as ECG and/or carotid pulse. It can also be noted that the performance of the algorithm is independent of the locations of S_1 and S_2 obtained from the PASCAL database.

Table 4.7: Performance comparison with approaches for normal, extra heart sound and artifacts classification.

Ref.	Type of recordings	Classifier	P_+ (%)
This work	Normal	Ensemble classifier	89.79
	Extra Heart Sound		82.51
	Artifacts		98.95
[45]	Normal	SVM	76.00 (best result)
	Extra Heart Sound		44.00
	Artifacts		94.00
[102]	Normal	SVM (SS-TD)	83.00 (best result)
	Extra Heart Sound		43.00
	Artifacts		80.00
[91]	Normal	SVM	79.00 (best result)
	Extra Heart Sound		100.00
	Artifacts		94.00

The locations of S_1 and S_2 were used only to study the variations in the intervals of S_1 and S_2, and, subsequently, in the systole and diastole intervals. This led to the identification of potential features for the classification.

Thus, using the features, the splitting of fundamental sounds is classified without extracting the discrete subcomponents of S_1 (M_1 and T_1) or S_2 (A_2 and P_2) sounds. This method could have a potential application in the automatic detection of cardiac abnormalities. It should be noted that, although, the proposed approach performed well on signals acquired in non-controlled environments, other features to assist with the extraction of subcomponents of S_1 and S_2 sounds could lead to further improvements. This can be further improved by training the classifier using features extracted from a large database (of split sounds) and with signals obtained from different subjects population, including wider age ranges, and in real use scenarios.

Apart from interpreting the class of acoustic recordings, this work could also find application in estimating the heart rate and heart rate variability using the S_1 or S_2 sounds. The heart rate variability could be estimated using the mid-points of S_1 (S_{1M}) and S_2 (S_{2M}) described in this chapter [265]. Additionally, the duration of systole and diastole could also be determined once S_1 and S_2 are accurately identified, this leading to an accurate estimation of cardiac cycle.

5 Classification of Normal and Abnormal Heart Sounds from Recorded Signals

5.1 Introduction

A comprehensive review of the approaches suggested by researchers to try to address the challenges involved in automatic analysis and classification of cardiac sounds recordings was presented in Chapter 1 [208]. Although some existing approaches achieved good classification performance, most of them have been validated with limited duration of recordings and protocols followed during signal acquisition were not always fully specified. For instance, information related to the subjects, such as gender and age, recording conditions and auscultation sites was not found for the database used in [45], [91]. This is important considering there can be significant variations in the features extracted using signals recorded from different auscultation sites resulting in varying classification performance [235]. Consequently, achieving an improved classification accuracy in interpreting cardiac acoustic signals acquired from different auscultation sites, particularly during continuous monitoring in real-world conditions, still remains an open research question. The objective of the work described in this chapter was to find potential features that could help in the identification and subsequent classification of normal and abnormal events in cardiac acoustic signals, taking into account the changes in characteristics with varying subject populations, including wider age ranges, auscultation sites, and in real use scenarios for long term monitoring.

5.2 Methods

A machine-learning based approach is proposed to try to address the challenges in the automatic classification of cardiac recordings as normal and abnormal. Several features were extracted from labelled normal and abnormal recordings originally collected as part of a clinical study as described in Appendix G (and Chapter 2), as well as recordings available from three large public databases. These include the PhysioNet Computing in Cardiology (CinC) Challenge 2016 database [22] (referred as Database1 in this study), the PASCAL database available from the Classifying Heart Sound Challenge 2011 (CHSC) [23] (referred as Database2 in this study),

and a public database with recordings collected from different sources available from [216] (referred as Database3 in this study). The number of recordings and their categories available from these databases are reported in Table 5.1.

The top 10 features from the signals in each database were identified and subsequently used to develop a model to classify normal and abnormal recordings. An overview of the approach used in this work together with the different stages of processing is shown in Figure 5.1. Finally, the proposed method's performance metrics were evaluated and compared with the recent state-of-the-art.

5.3 Pre-processing and Segmentation

Recordings were normalized to a fixed scale of [-1, 1]. As most of the information is present in low frequency band (below 500 Hz) (as discussed in Chapter 3), a low pass 4th order Butterworth filter with a cut off frequency of 500 Hz was used to eliminate high frequency components [217]. Further to remove the unwanted components such as low frequency noise

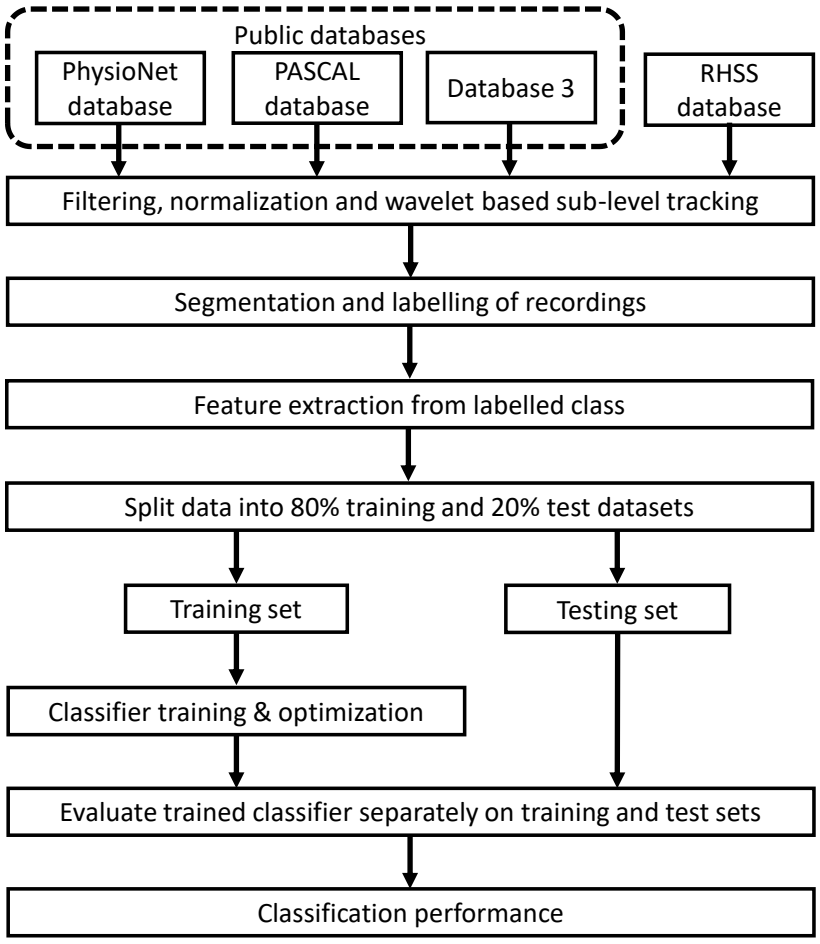


Figure 5.1. Block diagram of the overall heart sounds classification system.

and artifacts superimposed on recorded signals, a wavelet decomposition and reconstruction based adaptive sub-level filtering described in Chapter 3 was applied [215]. This is because the wavelet transform quantifies the non-uniform distribution of energy and provides adaptive resolution. The signals were decomposed using discrete wavelet transform and an adaptive threshold was used to selectively filter wavelet coefficients that corresponded to non-stationary part of the signal, before reconstructing them using an inverse wavelet transform [235]. The detailed and approximation coefficients were extracted by decomposing signals into 7-levels using db5 wavelet.

The reconstructed signals were used to localize cardiac sounds of interest by segmenting the peaks in the recordings. The accurate segmentation of recordings is an important step for the analysis; however, it is challenging to identify sounds of interest, especially in the presence of various intra-cardiac events. This is one the reasons, few research articles have suggested approaches to classify heart sounds signals without performing segmentation [56], [91], [230], [231]. Though this reduces the computational burden and avoid the segmentation error to propagate to the next level of analysis, accurate segmentation not only helps in the identification of S_1 and S_2 and subsequent determination of systole and diastole intervals, but also leads to

Table 5.1: Cardiac acoustic recordings used in this work.

Source	Subjects	No. of Recordings	Comments
Recorded heart sounds signals (RHSS database)	11 normal and 10 abnormal	14 males and 7 females (21 subjects)	5-minutes of recording from each auscultation site
PhysioNet data (Database 1) [22]	764 normal and abnormal subjects	3153	Labelled normal and abnormal recordings
The PASCAL database [23] (Datasets A and B) (Database 2)	Normal and abnormal	31	Normal
		34	Murmurs
		19	Extra heart sounds
		40	Artifacts
	Normal and abnormal	320	Normal
		95	Murmurs
		46	Extrasystole
Public database (Database 3) [216]	Normal and abnormal	200	Normal
		200	MR
		200	AS
		200	MVP
		200	MS

*MS: Mitral stenosis, AS: Atrial stenosis, MVP: mitral valve prolapse, MR: Mitral regurgitation.

basis for measurements helpful in assessing the overall cardiac conditions of the cardiovascular system. This work used the segmentation approach described in Chapter 3 which is based on the wavelet transform and Shannon energy estimation in combination with the characteristics/features extracted from heart sounds to help in the localization of sound peaks of interest. The details of various features extracted, and the steps followed to localize sound peaks of interest can be found in Chapter 3 (Section 3.2.2). An example of how the transients were localized is shown in Figure 3.6, in which the sound lobes of interest were identified clearly in the signal recoded during this study.

5.4 Feature Extraction & Selection

Due to the non-stationary and non-linear characteristics of heart sounds signals, several features containing quantitative and meaningful description of normal and abnormal recordings were extracted to help in accurately distinguishing normal and abnormal recordings. These features were: statistical time and frequency domain features (F_1-F_{24}), deterministic and non-deterministic energies (F_{25}-F_{26}), MODWT-based features (F_{27}-F_{48}), homomorphic envelogram features (F_{49}-F_{51}), Hilbert-Huang transform based features (F_{52}-F_{84}), and MFCCs features (F_{85}-F_{99}). They were already described in Chapter 2.

A feature vector containing 99 features (reported in Appendix F) was evaluated in this work. Furthermore, feature weights were computed using a diagonal adaptation of neighborhood component analysis (NCA) to derive a set of most significant features that could potentially best represent the signal [227]. NCA allocated weights to each feature by maximizing the expected leave-one-out classification accuracy [219]. As demonstrated in [215], the name-value pair argument '*Lambda*' representing the regularization parameter was tuned against the minimum classification loss using cross-validation to find the optimum value. The optimum value of *Lambda* was found to be 8.47×10^{-4}. This corresponded to a minimum classification loss of 0.0116.

Features weights were analysed to investigate which ones showed the greatest ability to differentiate normal and abnormal recordings from the different databases. The top 10 informative and non-redundant features with the highest weights were selected as optimum features from each database. The top 10 features identified using the RHSS database, the Database1 and the Database3 were: F_{25}, F_{26}, F_{37}, F_{59}, F_{61}, F_{69}, F_{73}, F_{74}, F_{75}, and F_{76}. For Database2, the top 10 features were: F_{25}, F_{26}, F_{32}, F_{37}, F_{56}, F_{59}, F_{73}, F_{74}, F_{75}, and F_{76}. Only features F_{32} and F_{56} identified in the Database2 were different from the top 10 features identified using the other databases. Thus, the 12 highest ranking feature which covered all the databases to train a classifier model to classify normal and abnormal recordings, were selected. These

features were - F_{25} (deterministic energy of S_1 and S_2 sounds), F_{26} (non-deterministic energy of the systole and diastole intervals), F_{32} (energy of MODWT decomposition coefficients at level 6), F_{37} (total energy of all the detail and approximation coefficients using MODWT), F_{56} (energy of IMF2 extracted using EMD), F_{59} (energy of all 5 IMFs using EMD), F_{61} (maximum frequency of IMF1 using EMD), F_{69} (instantaneous frequency of IMF2 using HHT), F_{73} (cluster of instantaneous frequencies of IMF2), F_{74} (cluster of instantaneous frequencies of IMF3), F_{75} (cluster of instantaneous frequencies of IMF4), and F_{76} (cluster of instantaneous frequencies of IMF5). The normalized weights of all these selected top 12 features obtained using NCA for the different databases used in this work are shown in Table 5.2. Although the selected features were almost the same, the different databases produced different weights for each feature. Among all the features selected, F_{25} was highest ranked by the RHSS database, the Database1, and the Database3. The selected feature vector with reduced number of features was used to train the classifier to assist the classification of normal and abnormal recordings.

Significant differences were observed between the normal and abnormal class of recordings for the selected features. The median and interquartile range (IQR) (or mid-spread) values obtained for the selected features using the RHSS database and the PhysioNet database, shown in Table 5.3, reveal the intra-class variations in the normal and abnormal recordings. For instance, abnormal recordings showed significantly reduced energy of MODWT decomposition coefficients at level 6 (denoted as F_{32}): 0.542(0.17) versus 1.02(0.259) (median (IQR)) for the RHSS database, and 4.53(14.19) versus 8.33(18.52) (median (IQR)) for the PhysioNet database. Thus, intra-class variations, measured as median and interquartile range, were observed in all the features selected from the normal and abnormal recordings.

The variation in the classifier's performance with respect to the features used for training the classifier was also studied. A trade-off between the computational cost and the number of features used for its training was observed. The computational cost (assessed in terms of the training time) of the system was reduced significantly with the reduction in feature vector used

Table 5.2: The top 10 features selected from the different databases.

RHSS	1	0.622	-	0.002	-	0.261	0.001	0.001	0.434	0.003	0.432	0.004
Database 1	1	0.003	-	0.349	-	0.843	0.012	0.769	0.405	0.036	0.001	0.274
Database 2	0.541	0.080	-	0.190	-	0.414	0.020	0.012	0.002	1	0.326	0.493
Database 3	1	0.519	0.057	0.006	0.337	0.219	-	-	0.0746	0.187	0.185	0.210
Features	F_{25}	F_{26}	F_{32}	F_{37}	F_{56}	F_{59}	F_{61}	F_{69}	F_{73}	F_{74}	F_{75}	F_{76}

1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0

to train the classifier. For instance, the system took 21.398 seconds for training with 99-dimensial feature vector, which was reduced to 8.680 seconds with 12-dimensional feature vector, while keeping the same number of observations. Thus, a promising classification performance was achieved using the 12 top features with reduced computational cost.

5.5 Classification of Cardiac Acoustic Signals

The classifier was trained using the top 12 features extracted from the four different databases listed in Table 5.1. A known set of input data (i.e. feature set from the recorded signals) and know responses to the data (i.e. class of the recorded signals) were used by the supervised

Table 5.3: Intra-class variations of the top 10 features selected using different databases.

Features	DB	RHSS		DB 1	
	Class	N	AN	N	AN
F_{25}	Med	228.6	261.1	183.4	198.8
	IQR	537.2	283.7	108.9	96.3
F_{26}	Med	308.5	414.3	213.8	247.2
	IQR	984.1	496.4	188.1	200.7
F_{32}	Med	1.02	0.542	8.33	4.53
	IQR	0.259	0.17	18.52	14.19
F_{37}	Med	211.6	111.4	66.17	15.32
	IQR	586.8	224.5	103.4	65.15
F_{56}	Med	32	8.662	16.37	5.93
	IQR	261.1	50.24	38.3	27.23
F_{59}	Med	224.5	117.6	62.63	17.52
	IQR	625.3	240.9	104.1	91.28
F_{61}	Med	161	174.1	59.57	41.13
	IQR	69.27	68.78	41.02	30.39
F_{69}	Med	109.2	97.95	69.02	58.99
	IQR	38.38	20.28	68.43	77.81
F_{73}	Med	192.4	182.9	196.5	171.2
	IQR	208.6	169.2	320.9	321.1
F_{74}	Med	105.9	100.1	79.66	64.96
	IQR	81.45	63.82	136.9	141.5
F_{75}	Med	70.56	53.44	45.05	35.98
	IQR	85.86	137.9	165.2	149.1
F_{76}	Med	15.5	13.2	15.04	19.24
	IQR	47.83	49.85	142.1	141.2

Med: median, IQR: interquartile range, DB: database, DB1: PhysioNet database (Database1).
N: normal and AN: abnormal or pathological

learning algorithm to train a model to potentially decode the class of unseen test data given their feature set.

5.5.1 Training and test datasets

As the characteristics of cardiac signals vary widely from subject to subject, it was important to obtain physiological measurements of the true labels from a varied population. Hence, datasets available from the different sources; together with the data recorded in this study from subjects with and without cardiac diseases, including wider age groups and auscultation sites, were used to obtain the results presented.

Different features were extracted from the recorded signals with an analysis window of 5s, producing the dimensions of feature matrix $X \in R^{4104 \times 12}$ from the RHSS database (1560 normal and 2544 abnormal); $X \in R^{13016 \times 12}$ from the Database 1 (PhysioNet database) (9858 normal and 3158 abnormal); $X \in R^{5271 \times 12}$ from the Database 2 (PASCAL database) (3245 normal and 2026 abnormal); and $X \in R^{1466 \times 12}$ from the Database 3 (287 normal and 1179 abnormal). The hold-out approach and 10-fold cross-validation were used to split the data (feature matrix obtained using recordings from different datasets) into training and testing datasets. Thus, for hold-out, 80% of the data was used for the training feature matrix $X_{training} \in R^{3284 \times 12}$ and the remaining 20% of the data was used for the test feature matrix $X_{test} \in R^{820 \times 12}$, where 3284 (1248 normal and 2036 abnormal segments) and 820 (312 normal and 508 abnormal segments) were the number of PCG instances or windows used for training and testing from the RHSS database, respectively. Similarly, the training feature matrix $X_{training} \in R^{10413 \times 12}$ from the Database 1 (7887 normal and 2526 abnormal segments); $X_{training} \in R^{4217 \times 12}$ from the Database 2 (2596 normal and 1621 abnormal segments); and $X_{training} \in R^{1173 \times 12}$ from the Database 3 (229 normal and 944 abnormal segments), were used for training. And the test feature matrix $X_{test} \in R^{2603 \times 12}$ from the Database1 (1971 normal and 632 abnormal segments); $X_{test} \in R^{1054 \times 12}$ from the Database2 (649 normal and 405 abnormal segments); and $X_{test} \in R^{293 \times 12}$ from the Database3 (58 normal and 235 abnormal segments), were used for testing.

The number of normal and abnormal segments in the databases were unbalanced. For instance, the Database3 produced a training feature matrix $X \in R^{1173 \times 12}$ containing 944 abnormal and 229 normal segments. As the normal class was underrepresented, the random undersampling boosting (RUSBoost) algorithm designed to address the problem of training classifier on skewed data was used [246].

5.5.2 Model training

In this work, the ensemble classification approach that combines the prediction of several learners to derive a powerful classifier model was used with the objective of improving performance [68], [228]. This approach constructs a set of classifiers with different weights assigned to the weak learners using an adaptive ensemble algorithm and then classifies a new dataset by taking a weighted vote of their predictions. The weighted combination of multiple algorithms reduces the variance of predictions to achieve good performance and robustness.

The selected features evaluated in Section 5.4, were used to train the classifier and optimum parameters of the classifier were extracted using the steps shown in Figure 5.2. The decision trees were used as weak learners due to their capacity to deal with high dimensional datasets. Both bagging and boosting algorithms were tested, and their performance were analysed to find the best suitable algorithm for the classification. As the true labelled class were normal and abnormal, the model was trained separately for each database to predict the normal or abnormal class based on the input feature matrix.

5.5.3 Hyperparameter optimization

Key hyperparameters (learning rate, number of learning cycles, number of weak learners/trees and the maximum number of splits) were optimized to improve the performance of the classifier model. This was achieved by analysing the performance of the key parameters with respect to the performance of the overall classifier during the training stage. To optimize the hyperparameter of the decision tree, the misclassification rate was computed iteratively for the different training sessions using 10-fold cross-validation. The cross-validated misclassification rate with respect to the number of trees, learning rate, and number of splits were analysed as described in [215]. Thus, the optimum misclassification rate of 0.0563 was obtained for number of trees = 93, maximum number of splits = 243 for a learning rate of 0.50, while tuning the number of trees (150), the learning rate (0.1, 0.25, 0.5, 0.75 and 1) and the maximum number

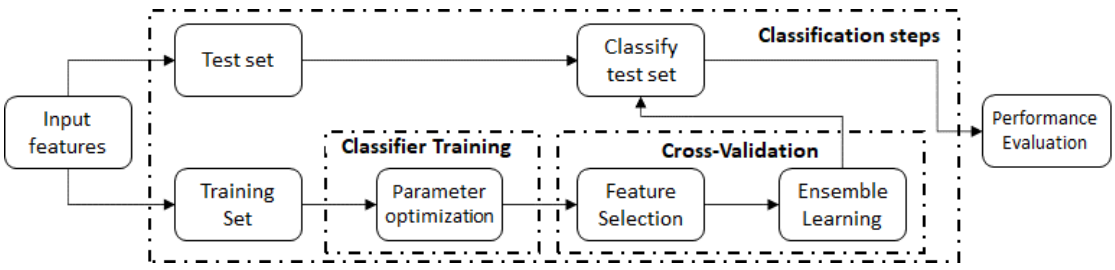


Figure 5.2. Steps used for training and cross-validation of the classifier to extract optimized hyperparameters.

of splits (1, 3, 9, 27, 81, 243, 729, 2187 and 6561).

5.6 Experimental Results

The performance of the proposed approach was assessed using the following metrics: recall or sensitivity $\left(S_E = \frac{TP}{TP+FN}\right)$, specificity $\left(S_P = \frac{TN}{FP+TN}\right)$, classification accuracy $\left(C_A = \frac{TP+TN}{TP+FP+FN+TN}\right)$, precision $\left(P_+ = \frac{TP}{TP+FP}\right)$, and F_1 score $\left(F_1 = 2\frac{P_+ \times S_E}{P_+ + S_E}\right)$. Here, true positive (*TP*) represents the number of abnormal PCG instances that are correctly classified as abnormal, false positive (*FP*) is the number of normal PCG instances that are incorrectly classified as abnormal, false negative (*FN*) is the number of abnormal PCG instances that are incorrectly classified as normal, and true negative (*TN*) is the number of normal PCG instances correctly classified as normal [75]. Thus, true positive (*TP*) represents the number of true abnormal recordings detected, and true negative (*TN*) is the number of true normal recordings. S_E and S_P were computed to estimate the true positive and true negative rates, respectively. F_1-score was used to represent the harmonic mean of S_E and $P+$ [70]. Finally, F_1-score, S_E, S_P and C_A were used to compare the performance of the proposed approach with the recent works.

The confusion matrices of the trained classifier for the different datasets are shown in Table 5.4 (AB = Abnormal recording segments, and N = Normal recording segments). These values were used to evaluate the performance of the classifier using 10-fold cross-validation on both the training and test feature matrices, for the RHSS database, Database 1, Database 2, and Database 3, as shown in Table 5.5. The scores illustrated in the table are the average values on all the observations, counting each observation when it was in a held-out fold. Using the training feature sets, the ensemble classifier achieved an average classification accuracy of 95.16% (RHSS database), 93.03% (Database 1), 99.41% (Database 2) and 97.44% (Database 3). Similarly, an average classification accuracy of 90.24% (RHSS database), 93.08% (Database 1), 88.43% (Database 2), and 96.59% (Database 3), were achieved using the test sets. From the average classifier's performance metrics, it can be observed that the classifier achieved almost same true positive and true negative rates for both the abnormal and normal class (S_E of 92.383% and S_P of 93.611%), across all the databases. Thus, the classifier was able to differentiate the normal and abnormal recordings.

Furthermore, the ensemble classifier's outputs were used to estimate the probabilities for each observation in training and test sets, for being a member of the normal or abnormal class. This was helpful in estimating the confidence of the classifier on the predicted class. The probability that an observation belongs to a normal class was denoted as *P*(*N*), whereas the abnormal class was denoted as *P*(*AN*). The number of observations falling within an interval of

Table 5.4: Confusion matrices of the trained classifier for different datasets.

(RHSS Database (A = Training Set and, B = Test Set), Database 1 (C = Training Set and, D = Test Set), Database 2 (E = Training Set and, F = Test Set), and Database 3 (G = Training Set and, H = Test Set)) (N= Normal Recordings and AN = Abnormal Recordings)

(A)

		Actual	
		AN	N
Predicted	AN	1976	99
	N	60	1149

(B)

		Actual	
		AN	N
Predicted	AN	471	43
	N	37	269

(C)

		Actual	
		AN	N
Predicted	AN	2154	354
	N	372	7533

(D)

		Actual	
		AN	N
Predicted	AN	543	91
	N	89	1880

(E)

		Actual	
		AN	N
Predicted	AN	1610	14
	N	11	2582

(F)

		Actual	
		AN	N
Predicted	AN	338	55
	N	67	594

(G)

		Actual	
		AN	N
Predicted	AN	928	14
	N	16	215

(H)

		Actual	
		AN	N
Predicted	AN	228	3
	N	7	55

±1% and ±5% from the mean probability ($P(N) = P(AN) = 0.5$) were estimated. Hence, if an observation with $P(N)$ (or $P(AN)$) was found within the interval ±1% or ±5%, this means that the observation was not classified confidently (i.e. weakly classified). The performance of the classifier excluding these observations (observations falling with the interval of ±1% and ±5% from the mean probability) is shown in Table 5.6 and Table 5.7, for the RHSS database and the PhysioNet database. Please note, the PhysioNet database (Database1) was used as a benchmark to measure the performance of the proposed classifier. For the RHSS training set, 20 observations (4 TP, 7 FP, 5 FN, 4 TN) in ±1% interval and 86 observations (17 TP, 25 FP, 13 FN and 31 TN) in ±5% interval were found. The achieved classification accuracy the classifier after excluding these observations from the RHSS training set was found to be 95.50% (for ±1% interval) and 96.22 (for ±5% interval). Thus, this signifies that there were very few observations with low prediction values ($P(N)$ or $P(AN)$), attesting to the fact the classifier was confident in classifying each observation.

Table 5.5: Performance of the classifier using 10-fold cross-validation.

		S_E (%)	S_P (%)	P_+ (%)	F_1 (%)	C_A (%)
RHSS Database	Training set	97.05	92.07	95.23	96.13	95.16
	Test set	92.72	86.22	91.63	92.17	90.24
Database 1	Training set	85.27	95.51	85.89	85.58	93.03
	Test set	85.92	95.38	85.65	85.78	93.08
Database 2	Training set	99.32	99.46	99.14	99.23	99.41
	Test set	83.46	91.53	86.01	84.71	88.43
Database 3	Training set	98.31	93.89	98.51	98.41	97.44
	Test set	97.02	94.83	98.70	97.85	96.59
Average		*92.383*	*93.611*	*92.595*	*92.482*	*94.172*

5.7 Discussion

This chapter presented a supervised learning approach to classify normal and abnormal cardiac auscultatory recordings. In order to validate the approach, various important factors that largely impact the physiology of heart sounds including the different auscultation sites used to record signals and the duration of recordings, were considered.

The electrical and mechanical activities of a human heart are not constant and vary significantly with the subjects' cardiac conditions. Hence, a classifier model trained using features extracted from signals acquired from a varied subject population, including those with different cardiac conditions, plays an important role in determining the performance of the classifier. For this reason, 99 features were extracted from the different databases containing true labelled normal and abnormal acoustic signals recorded in both clean and noisy environments. Later, a feature selection approach was used to determine the top 12 features

Table 5.6: Classification performance of the classifier model after excluding observations within ±1% of the mean probability.

Database	TP	FP	FN	TN	S_E (%)	S_P (%)	*PPV* (%)	F_1-score (%)	C_A (%)
RHSS Database	1972	92	55	1145	97.29	92.56	95.54	96.41	95.50
	469	41	36	265	92.87	86.60	91.96	92.41	90.51
Database 1	2139	341	358	7530	85.66	95.67	86.25	85.96	93.26
	540	88	88	1877	85.99	95.52	85.99	85.99	93.21
Average					90.452	92.587	89.935	90.192	93.12

Table 5.7: Classification performance of the classifier model after excluding observations within ±5% of the mean probability.

Database	TN	FP	FN	TN	S_E (%)	S_P (%)	PPV (%)	F_1-score (%)	C_A (%)
RHSS Database	1959	74	47	1118	97.66	93.79	96.36	97.00	96.22
	462	37	31	259	93.71	87.50	92.59	93.15	91.38
Database 1	2096	311	314	7478	86.97	96.01	87.08	87.03	93.87
	521	74	79	1870	86.83	96.19	87.56	87.20	93.99
Average					91.292	93.372	90.897	91.095	93.865

showing the greatest ability to differentiate normal and abnormal recordings.

Among all the features investigated, the sum of energies of the deterministic events - S_1 and S_2 (F_{25}), was identified as the most promising feature using the RHSS database, the Database 1, and the Database 3. The sum of energies of the non-deterministic events (F_{26}) was also found to be ranked in top 10 performing features based on the obtained weights. This is mainly because an abnormal recording would have more chaotic behaviour than a normal clean recording. A normal cardiac cycle contains a specific sequence of S_1 followed by a systole interval, and S_2 followed by a diastole interval. However, there may be additional sounds present in case of a cardiac abnormality resulting in the variation of the duration of peaks as well as the intervals between the adjacent peaks. Thus, the variation in the duration of intervals to preceding and following peaks in the extracted heart sounds envelope provided significant information for the classification of cardiac auscultatory recordings. Sounds other than fundamental heart sounds were used as an indicator to various cardiac abnormalities. For instance, murmurs produced due to turbulent flow of blood would be present in either the systolic or diastolic interval of a cardiac cycle, amounting to non-deterministic energy helpful in the classification. This can also be confirmed from the weights obtained using the Database 3 containing 80% of recordings from subject having mitral stenosis (MS), mitral regurgitation (MR), aortic stenosis (AS) and mitral valve prolapse (MVP). Both deterministic and non-deterministic features (F_{25} and F_{26}) were ranked first and second, respectively in classifying normal and abnormal recordings.

To evaluate F_{25} and F_{26}, accurate localization of sound peaks of interest (including fundamental heart sounds) was an essential step. The heart sounds were segmented using the approach described in Chapter 3. This helped in the identification of the start and end locations of sound peaks of interest, which were subsequently used to estimate the intervals between the peaks to help in computing the deterministic and non-deterministic energy. This also leads to

the conclusion that the proper identification of start and end points of peaks is helpful for feature extraction.

Other features extracted using Hilbert Huang transform (F_{73}, F_{74}, F_{75}, and F_{76}) and MODWT (F_{32} and F_{37}) also ranked highly for the classification of recordings. Among them, F_{74} (cluster of instantaneous frequencies of IMF3 using HHT) achieved the highest weight using the Database 2. The amplitude and instantaneous frequency extracted using Hilbert-Huang transform have previously demonstrated successful results [72]. However, in addition to these features, the *k*-means clustering algorithm was used to obtain a single value of the instantaneous frequency and instantaneous energy as features. The instantaneous frequencies of the IMF2, IMF3, IMF4 and IMF5 (denoted as F_{73}, F_{74}, F_{75}, and F_{76}) extracted using the *k*-means clustering algorithm were ranked among the top 10 features. Thus, in this work, the IMFs that were suitable for evaluating instantaneous frequency and energy-based features, were identified. Although some of the features reported in this study have been already discussed in the literature and explored for the classification of cardiac acoustic recordings, to the best of author's knowledge, the MODWT based features have never been used before for the classification of normal and abnormal recordings.

Currently, the most extensive, standardized and widely used database of heart sounds recordings available in literature is the PhysioNet CinC Challenge 2016 database [20], [22]. This database acts as a benchmark for the development and performance evaluation of algorithms for classification of recordings as normal and abnormal. Hence, most of the recent works on the classification of heart sounds as normal and abnormal used this database to validate the performance of the algorithms [8], [266]. However, the length of recordings varied from 5 seconds to 122 seconds. Similarly, the length of recordings available from the other public databases were mostly less than one-minute. Hence, it was important to validate the algorithms using larger data recorded continuously in real, non-controlled environment. To address this, signals in the RHSS database were continuously recorded for 5 minutes from four auscultation sites (Aortic, mitral, pulmonary, and tricuspid), from a varied subject population including normal and cardiac patients, in real non-controlled settings. Thus, this work can be used as a proof of concept to develop approaches that can assist doctors in automatically examining large data and assessing the overall cardiac condition of a subject, helpful in the early-stage diagnosis of cardiac events that are of pathological/clinical importance.

The trained ensemble classifier using the selected features achieved an average classification accuracy of 94.172% in classifying acoustic recordings as normal and abnormal using the different databases. Also, the robustness of the achieved results was demonstrated using the different datasets in Table 5.5. From the achieved results it can also be concluded that the

extracted informative and non-redundant features were well-suited for the non-stationary heart sounds signal analysis.

Achieved results in this work were compared with the recent works as illustrated in Table 5.8. Most of the recent proposed algorithms were validated using the PhysioNet CinC database. Hence, the top 5 challenge scores achieved by [103], [154], [267]–[269], as the outcome of the 2016 PhysioNet/Computing in Cardiology (CinC) Challenge published in [266], were used for the comparison. A modified accuracy $\left(MA_{CC} = \frac{S_E+S_P}{2}\right)$ was used to compare the results as it gives a single mean of sensitivity and specificity. Using this database, Potes *et al* [154] achieved the highest Challenge score with MA_{CC} of 86.02%, extracting 124 time-frequency features fed to an AdaBoost classifier and a deep-learning based classifier. Thus, the approach presented in this work achieved an improved MA_{CC} of 90.65% compared to [154]. This was also applicable for the MA_{CC} achieved by [103], [267]–[269]. The presented approach also achieved an improved MA_{CC} compared to Tang *et al* [270], which was not a part of the PhysioNet Challenge. Similarly, an improved overall classification accuracy (C_A) of 93.08% was achieved compared to Singh and Majumder [230]. Although a higher MA_{CC} of 92.9% was achieved by Arora *et al* [231] using the PhysioNet database, the overall MA_{CC} achieved in this work across all the datasets was found to be higher. Using the PASCAL database, the overall performance

Table 5.8: Performance comparison with approaches for normal and abnormal classification of recordings

Ref.	Dataset	S_E (%)	S_P (%)	C_A (%)	MA_{CC}
This work	All databases (average)	92.38	93.61	94.172	92.995
	PhysioNet data (test set)	85.92	95.38	93.08	90.65
[154]	PhysioNet data (test data)	94.24	77.81	-	86.02
[103]	PhysioNet data (test data)	86.91	84.90	-	85.90
[267]	PhysioNet data (test data)	87.43	82.97	-	85.20
[268]	PhysioNet data (test data)	86.39	82.69	-	84.54
[269]	PhysioNet data (test data)	88.48	80.48	-	84.48
[230]	PhysioNet data	94.08	91.95	92.47	-
[270]	PhysioNet data	88.00	87.00	-	88.00
[231]	PhysioNet data	94.5	91.3	-	92.9
[68]	PhysioNet data (test data)	79.6	80.6	-	80.1
[45]	PASCAL Dataset-B	34.00	90.00 (best)	-	-
[102]	PASCAL Dataset-B	49.00	84.00	-	-
[91]	PASCAL Dataset-B	34.00	95.00	-	-

achieved in articles [45], [91], [102] was found to be less than the performance achieved in this work. From the results it can be observed that the approach discussed in this work achieved an improved overall classification accuracy in interpreting the class of signals.

6 Non-invasive Approach for Abnormality Detection using Heart Sounds Recordings

6.1 Introduction

Cardiovascular diseases due to heart valve disorders (or valvular heart diseases) are characterized by the damage or defect in one of the four heart valves: the aortic valve, mitral valve, tricuspid valve, and pulmonary valve. Examples of these diseases include aortic stenosis (AS), mitral regurgitation (MR), aortic regurgitation (AR), mitral stenosis (MS), and others.

As discussed in Chapter 1, depending on the cardiac condition of a subject, abnormal recordings may contain a variety of additional pathological sounds such as third heart sounds (S_3), fourth heart sounds (S_4) and murmurs, along with the fundamental heart sounds – S_1 and S_2. Murmurs present in a recorded signal are mainly produced due to the turbulence in blood flow or defect in any of the four heart valves; where latter are usually categorized as stenosis or regurgitation [99]. Stenosis occurs when heart valves do not open completely, while regurgitation occurs when valves do not close completely. In auscultatory recordings, depending on the types of murmurs, they can be heard either in systolic or diastolic interval of a cardiac cycle. Stenosis of the aortic or pulmonary valve results in systolic murmurs, while regurgitation of either of these valves leads to diastolic murmurs. Similarly, stenosis of the mitral or tricuspid valves results in diastolic murmurs, while regurgitation of either of these valves leads to systolic murmurs. These murmurs reflect the pathological characteristics of the underlined heart valve and can be used to diagnose cardiac diseases [271]. Accurate identification of murmurs due to heart valve disorders can potentially help in marking cardiac events that are of pathological/clinical importance. Thus, their early detection can help in reducing the mortality rate due to cardiovascular diseases. This chapter presents a non-invasive approach to automatically detect heart valve disorders and determine the class of valvular heart disease using cardiac auscultatory recordings.

6.2 Methods

Features were extracted from labelled recordings, based on the information available from the

databases, and subsequently employed to develop an automated approach for non-invasive detection of heart valve disorders, namely the mitral stenosis (MS), mitral regurgitation (MR), mitral valve prolapse (MVP), and aortic stenosis (AS). There recordings were obtained from two public databases: a public database reported in [216] (represented as DB1 in this study) and the PASCAL database available from the Classifying Heart Sound Challenge (CHSC) [23]. The details of these databases are provided in Chapter 2. The number of recordings and their categories available from these databases is reported in Table 6.1. As an illustration, Figure 6.1 shows the different types of signals from the database DB1 containing three cardiac cycles in each. As before, the three major steps used in this approach were pre-processing and segmentation, followed by the feature extraction and selection to train a machine learning-based classifier. The flow-chart of these steps is shown in Figure 6.2.

The very first step prior to feature extraction was to remove high-frequency components, caused by noise or interference. A low pass 4^{th} order Butterworth filter with a cut off frequency of 500 Hz was applied to eliminate high-frequency components. This is because most of the information related to heart sounds and murmurs is present in low frequency band (below 500 Hz) [70], [272]. Subsequently, and for the same purpose, but to deal with unwanted low-frequency components overlapped in the frequency band of interest, wavelet decomposition and reconstruction based adaptive sub-level filtering approach was applied, as described in Chapter 3. The wavelet transform quantifies the non-uniform distribution of energy and provides adaptive resolution. Recordings were decomposed using a 7-level db5 discrete wavelet transform (DWT) to extract the detailed and approximation coefficients, and a thresholding criterion was used to selectively filter wavelet coefficients that corresponded to non-stationary part of the signal [235]. Filtered wavelet coefficients were used to reconstruct the signal using an inverse wavelet transform. Recordings were then normalized to a fixed scale of [-1, 1].

Table 6.1: Cardiac acoustic recordings used in this work.

Source		Subjects	No. of Recordings	Comments
Public database from [216]		Normal and abnormal	200	Normal
			200	MR
			200	AS
			200	MVP
			200	MS
The PASCAL database [23]	Dataset A	Normal	31	Normal
		Abnormal	34	Murmurs
	Dataset B	Normal	320 (120 noisy)	Normal
		Abnormal	95 (29 noisy)	Murmurs

*MS: Mitral stenosis, AS: Atrial stenosis, MVP: mitral valve prolapse, MR: Mitral regurgitation.

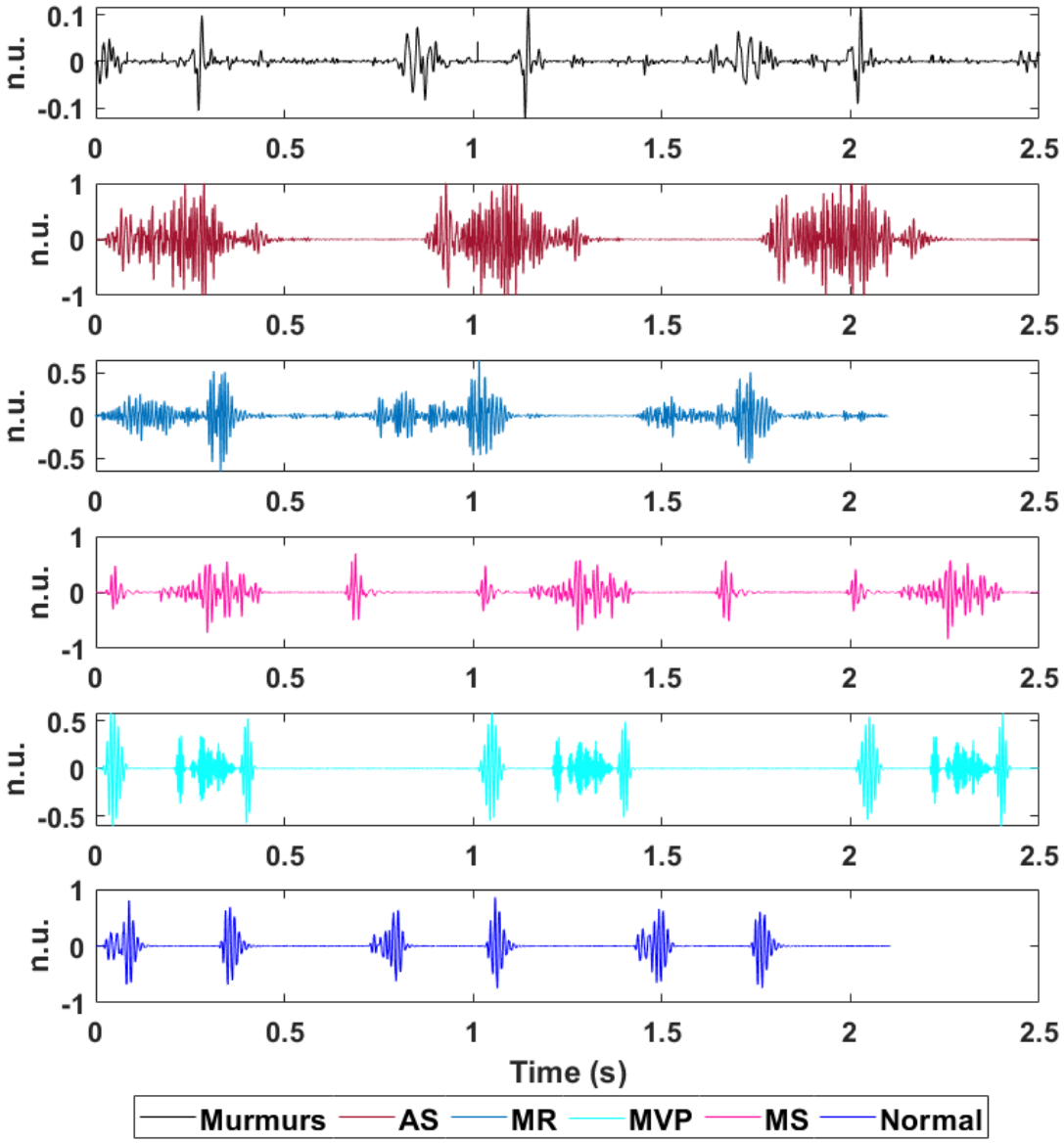


Figure 6.1. Different types of recordings used in this study. Murmur recording from the PASCAL database; and aortic stenosis (AS), mitral regurgitation (MR), mitral valve prolapse (MVP), mitral stenosis (MS), and normal recordings from a public database.

The pre-processed and reconstructed signals were used to localize cardiac sounds of interest, and to subsequently segment the recordings as described in Chapter 3. An example of how the transients were localized is shown in Figure 4.6, in which the sound lobes of interest were identified clearly in the signal recorded during this study. These peaks potentially represent fundamental heart sounds; however, there may be some peaks present which are neither S_1 nor S_2.

6.3 Feature Extraction & Selection

Features were extracted from segmented recordings, taking into account the characteristics of heart sounds that vary significantly with the recording conditions and the type of recording under investigation. Due to the non-stationary behaviour of the recorded signals, several features containing quantitative description were extracted and subsequently used to differentiate normal recordings from different class of abnormal recordings.

6.3.1 Feature extraction

Any physiological alteration in the cardiovascular events associated with the contraction and relaxation of the atria and ventricles influences the occurrence of S_1 and S_2 sounds, produced due to the closure of atrioventricular and semilunar valves, respectively [218]. Subsequently, the duration of the systolic interval (starting at the beginning of S_1 and lasting until the beginning of S_2), as well as the diastolic interval (starting at the beginning of S_2 and lasting until the beginning of S_1) may vary significantly as a function of different cardiovascular conditions and so will the sound generated in the process [217]. This is the reason why statistical time and frequency domain features were considered in this work. 10 features (F_1 to F_{10}) extracted were

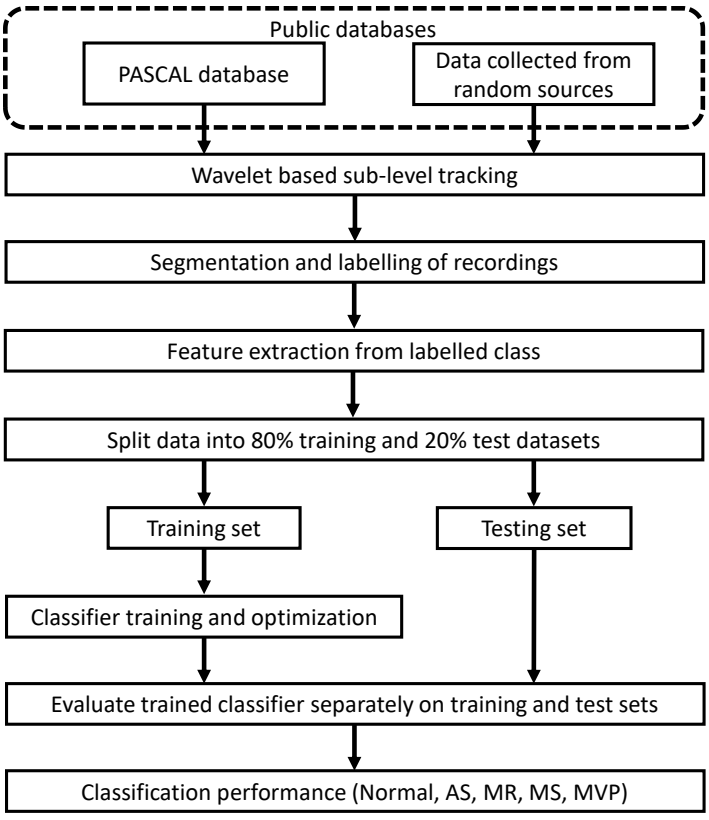


Figure 6.2. Block diagram of the overall heart valve disorders classification system.

the mean of the absolute value of P_1 sound intervals (P_{1T}), the mean of the absolute value of P_2 sound intervals (P_{2T}), the ratio of subsequent P_1 and P_2 intervals (P_{1T}/P_{2T}), the interval between the beginning of subsequent P_1 and P_2 sounds ($T_{P1\text{-}P2}$), the interval between the beginning of P_2 and the next P_1 sounds ($T_{P2\text{-}P1}$), the interval between the two adjacent P_1 sounds ($T_{P1\text{-}P1}$), the interval between the two adjacent P_2 sounds ($T_{P2\text{-}P2}$), the ratio of $T_{P1\text{-}P2}$ to $T_{P2\text{-}P1}$ intervals ($T_{P1\text{-}P2}/T_{P2\text{-}P1}$), the ratio of $T_{P1\text{-}P2}$ interval to P_1 interval ($T_{P1\text{-}P2}/P_{1T}$), and the ratio of $T_{P2\text{-}P1}$ interval to P_2 interval ($T_{P2\text{-}P1}/P_{2T}$). Here, P_1 and P_2 represent the first and second sound peaks of interest localized in a recorded signal, occurring in an alternate manner, respectively.

In addition to these features, 14 features (represented as F_{11}-F_{24}) were the signal's (denoted as $h[n]$) mean $\left(\overline{h[n]}\right)$ (F_{11}), the median (F_{12}), standard deviation (F_{13}), mean absolute deviation (F_{14}), quantile (F_{15} and F_{16}), interquartile range (F_{17}), skewness (F_{18}), kurtosis (F_{19}), spectral entropy (F_{20}), dominant frequency (f_D) (F_{21}), maximum frequency for each sound lobes (f_{max}) (denoted as F_{22}), the ratio of dominant frequencies (F_{23}), and zero-crossings (F_{24}). The detailed description and methods used to evaluate these features (F_1 to F_{24}) can be found in Chapter 2 (Section 2.2).

Apart from these statistical time and frequency domain-based features, 2 features (F_{25} and F_{26}) were based on the deterministic and non-deterministic energy; 22 features (F_{27} to F_{48}) extracted using multiresolution analysis were based on Maximum Overlap Discrete Wavelet Transform (MODWT); 3 features (F_{49}, F_{50} and F_{51}) represented the peak amplitude, maximum frequency extracted from homomorphic envelope and the index value corresponding to the maximum frequency; 3 features (F_{52}, F_{53} and F_{54}) represented the peak amplitude, maximum frequency, and the index value corresponding to the maximum frequency in the spectrum using the extracted Hilbert envelope; 6 features (F_{55} to F_{60}) were the energy computed for the first 5 intrinsic mode functions (IMFs) and the residual part; 8 features (F_{61} to F_{68}) from 4 selected IMFs (IMF2, IMF3, IMF4, and IMF5) were extracted to represent the maximum frequency and its corresponding index value in the spectrum; 8 features (F_{69} to F_{76}) were the instantaneous frequency and the cluster of instantaneous frequencies for each selected mode (IMF2, IMF3, IMF4, and IMF5); 8 features (F_{77} to F_{84}) were the instantaneous energy and the cluster of instantaneous energies for each selected mode (IMF2, IMF3, IMF4, and IMF5); and 15 features (F_{85} to F_{99}) were the mel-frequency cepstral coefficients (MFCCs). The methods used to evaluate these features can be found in Chapter 2. Additionally, the description of all the features studied in this work is provided in Appendix F.

As an illustration, the 4 IMFs selected (IMF2, IMF3, IMF4 and IMF5) used to extract instantaneous frequency and instantaneous energy of a murmur recording are shown in Figure 6.3. Compared to the instantaneous frequency and instantaneous energy of a normal signal

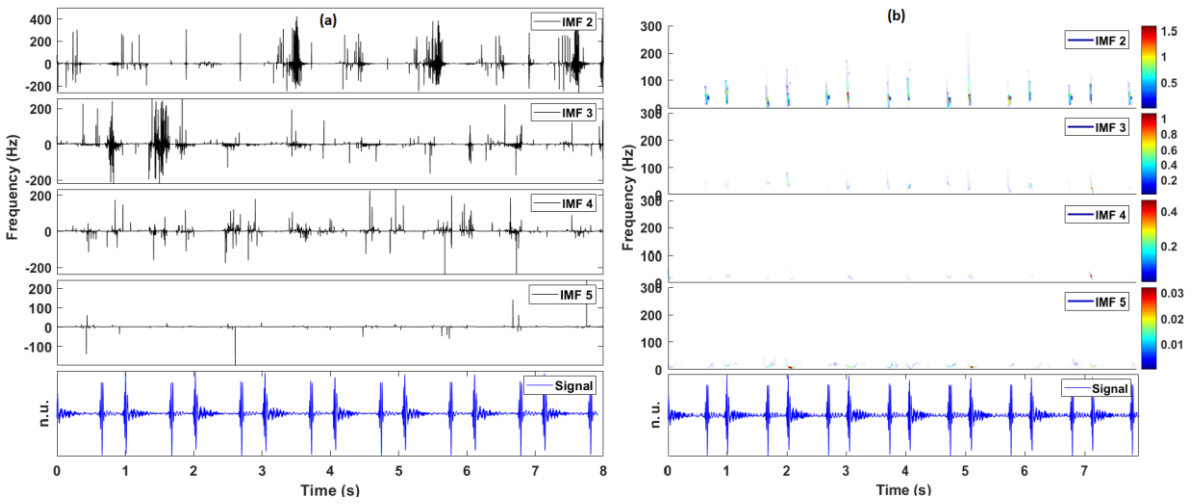


Figure 6.3. (a) The instantaneous frequency, and (b) instantaneous energy for IMF 2, IMF 3, IMF 4 and IMF 5, selected for feature extraction. Murmur recording obtained from the PASCAL database is used here.

shown in Figure 2.8, it was found that the information in murmurs recordings was sedimented in lower IMFs. This can also be observed from the IMFs extracted for the signals recorded from subjects with AS (shown in Figure 6.4), MVP (in Figure 6.5), MR (in Figure 6.6), and MS (in Figure 6.7). In these figures, values in the y-axis are represented in the normalized unit.

6.3.2 Feature selection and reduction

The extracted features were statistically analysed to identify the top 10 features that achieved the highest weights and demonstrated the greatest ability to differentiate the normal from the target pathological sounds from the different databases. These optimum features were selected for the classification. Feature weights were computed using a diagonal adaptation of neighborhood component analysis (NCA) [227]. Weights were allocated to each feature by maximizing the expected leave-one-out classification accuracy [219]. As demonstrated in [215], the name-value pair argument '*Lambda*' representing the regularization parameter was tuned against the minimum classification loss using cross-validation to find the optimum value. The optimum value of *Lambda* was found to be 0.48×10^{-3}; corresponding to a minimum classification loss of 0.02741.

The top 10 features identified using the PASCAL database (DB2 and DB3) were: F_{25}, F_{26}, F_{37}, F_{59}, F_{61}, F_{69}, F_{73}, F_{74}, F_{75}, and F_{76}. The normalized weights of these selected features are shown in Table 6.2. Although the top 10 features identified using the noisy and clean recordings from the PASCAL database were the same, a slight variation in the weights achieved was observed. This signifies that the presence of noise in the signal impacts the weights of the

extracted features. Furthermore, the top 10 features which showed the greatest ability to differentiate normal recordings from each cardiac abnormality (aortic stenosis (AS), mitral

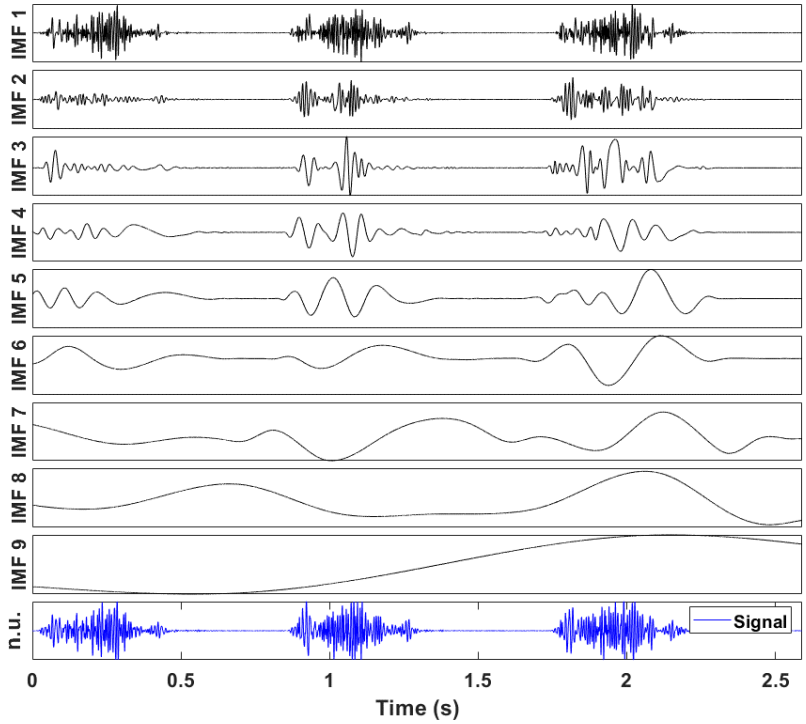


Figure 6.4. Extracted 9 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with aortic stenosis (AS).

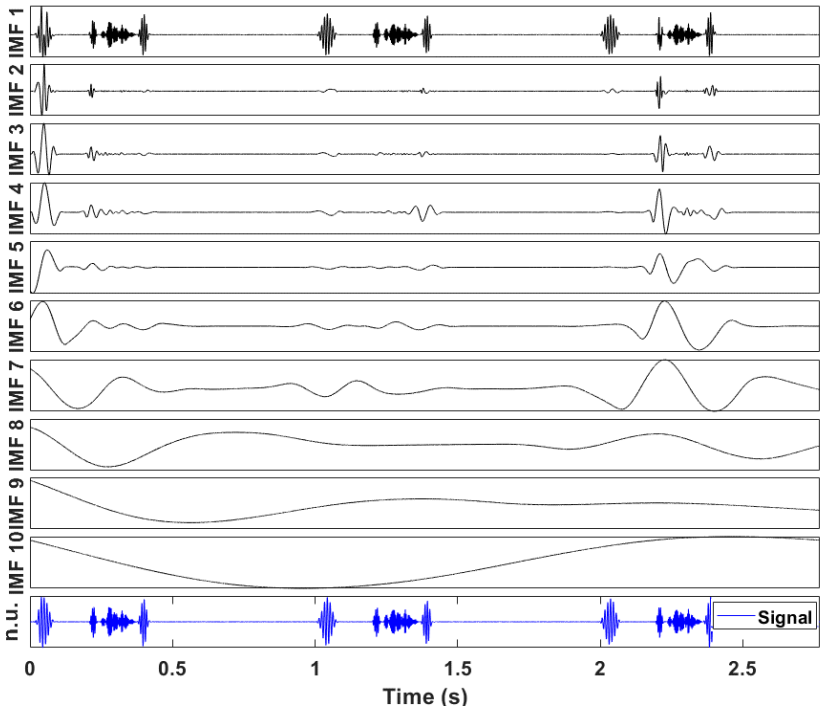


Figure 6.5. Extracted 10 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral valve prolapse (MVP).

stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP)) were also identified.

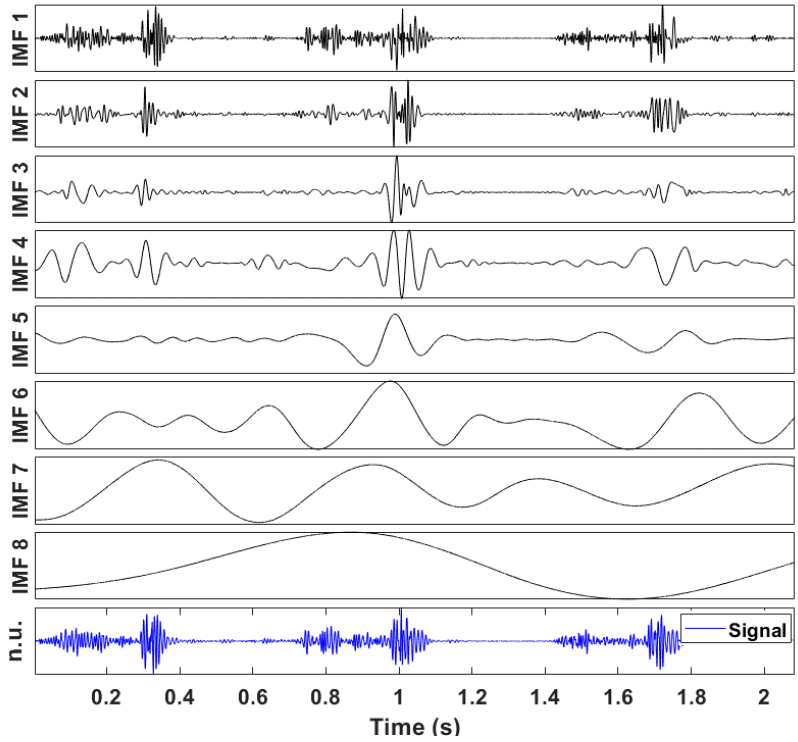


Figure 6.6. Extracted 8 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral regurgitation (MR).

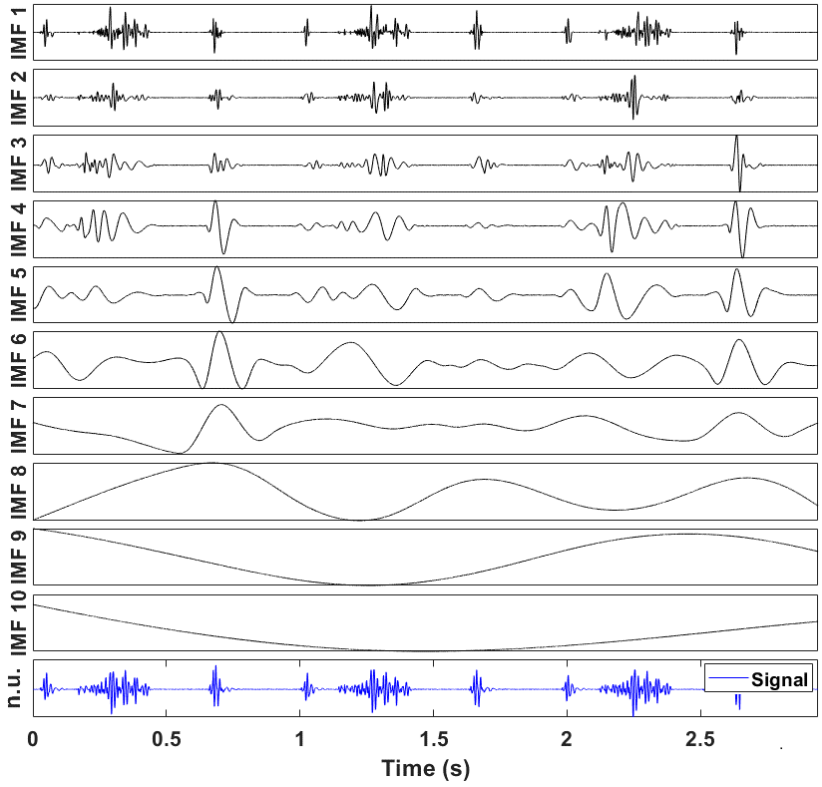


Figure 6.7. Extracted 10 intrinsic mode functions (IMFs) using empirical mode decomposition, of a signal recorded from a subject with mitral stenosis (MS).

Among the top 10 features identified using different datasets, 7 features (F_{25}, F_{26}, F_{37}, F_{59}, F_{73}, F_{74}, and F_{75}) were common for all the datasets (as shown in Table 6.2). Among them, features F_{75}, F_{25} and F_{73} showed the greatest ability to classifying normal and murmurs recordings. To achieve the best fit using the different databases used for the classification, a set of 13 features was selected. These 13 features were: F_{25} (deterministic energy of S_1 and S_2 sounds (represented as P_1 and P_2)), F_{26} (non-deterministic energy of the systole and diastole intervals (i.e. peak to peak intervals)), F_{32} (energy of MODWT decomposition coefficients at level 6), F_{37} (total energy of all the detail and approximation coefficients using MODWT), F_{43} (energy at level 6 computed using MRA of MODWT matrix), F_{56} (energy of IMF2 extracted using EMD), F_{59} (energy of all 5 IMFs using EMD), F_{61} (maximum frequency of IMF1 using EMD), F_{69} (instantaneous frequency of IMF2 using HHT), F_{73} (cluster of instantaneous frequencies of IMF2), F_{74} (cluster of instantaneous frequencies of IMF3), F_{75} (cluster of instantaneous frequencies of IMF4), and F_{76} (cluster of instantaneous frequencies of IMF5). Among all the features selected, F_{25} achieved the highest weight in classifying normal sounds and murmurs using the PASCAL database. Similarly, F_{75} achieved the highest weight in classifying pathological recordings (AS, MS, MR, and MVP). In this work, the most significant

Table 6.2: The top 10 features selected from the different databases

Features	DB1	DB2	DB3
F_{25}	0.4908	1	0.8702
F_{26}	0.2531	0.91087	0.4001
F_{32}	0.0089	-	-
F_{37}	0.619	0.66235	0.0069
F_{43}	0.0703	-	-
F_{56}	0.008	-	-
F_{59}	0.431	0.00858	0.1825
F_{61}	-	0.44177	0.9201
F_{69}	-	0.02283	0.1256
F_{73}	0.2223	0.15113	1
F_{74}	0.0645	0.84221	0.7515
F_{75}	1	0.058229	0.501
F_{76}	-	0.03466	0.0118

Note: DB1: the database obtained from [216]. DB2: the PASCAL database excluding noisy recordings, DB3: the PASCAL database with noisy recordings.

features representing the characteristics of the recorded signal were identified to assist the classification of abnormal acoustic recordings. Reducing the number of features helps in cutting the computational cost required to train the classifier. In general, larger the number of features, higher computational power is required to process. This has been discussed in the results section.

The selected features evaluated for the different types of recordings showed significant difference in the statistical values. The median and interquartile range (IQR) (or mid-spread) values computed for the selected features shown in Table 6.3 reveal the intra-class variations helpful for the classification. For instance, abnormal recordings had significantly higher energy of deterministic events (denoted as F_{25}): 175.6 (514.0) versus 119.9 (145.2) (median (IQR)) for the noisy data (DB 3); and 184.2 (534.2) versus 121.8 (180.3) (median (IQR)) for the clean data (DB 2). Similarly, pathological recordings (AS (292.5 (326.0)), MR (126.2 (271.7)), MS (93.24 (291.1)), and MVP (84.74 (222.4))) achieved higher deterministic energy values against the normal recordings (27.94 (242.1)). As shown in Table 6.3, the selected features achieved different median values for pathological recordings, and hence were found to be statistically significant for the classification. Also, features analysed using the noisy, and clean datasets from the PASCAL database, demonstrated slight variations in the weights. This signifies that the features evaluated from a signal may vary if noise is superimposed during data collection. Hence, the approach presented was validated using the different datasets (both clean and noisy) to extract robust set of features.

6.4 Classification of Pathological Recordings

In this work, the classifier was trained using the statistical measurements extracted from the true known states (normal, murmurs, AS, MS, MR and MVP) using the databases reported in Table 6.1. This helped in extracting robust global parameters of the model. A know set of input data (i.e. feature set from the recorded signals) and know responses to the data (i.e. class of the recorded signals) were used by the supervised learning algorithm to train a model to potentially decode the class of unseen test data given their feature set.

6.4.1 Training and test datasets

As the extracted features vary widely with respect to the recorded signal, recordings obtained from the two large databases were used. Among them, the PASCAL database contains clean recordings (dataset is represented as DB2), and noisy recordings (dataset is represented as DB3). The feature vector evaluated for an analysis window of 5 seconds produced a feature matrix of dimensions $X \in R^{2561 \times 13}$ using 331 recordings (97 murmurs and 234 normal clean recorded signals) from the PASCAL database (DB2); feature matrix $X \in R^{3032 \times 13}$ using 480

recordings (129 murmurs and 351 normal recordings with noisy signals) from the PASCAL

Table 6.3: Intra-class variations of features selected using different databases.

Features	DB	DB1					DB2		DB3	
	Class	AS	MR	MS	MVP	N	M	N	M	N
F_{25}	Med	292.5	126.2	93.24	84.74	27.94	184.2	121.8	175.6	119.9
	IQR	326	271.7	291.1	222.4	242.1	534.2	180.3	514	145.2
F_{26}	Med	145.4	66.34	145.9	67.33	11.21	228.6	105.9	223.4	114.2
	IQR	232.3	131.1	378.1	193.7	184.4	647.6	238.7	628.6	185.9
F_{32}	Med	19.95	8.949	36.32	37.68	89.32	0.056	0.034	0.064	0.154
	IQR	52.9	9.23	25.8	67.9	25.7	0.38	0.57	0.68	6.72
F_{37}	Med	42.63	32.5	52.85	64.77	116.7	32.31	2.826	32.28	8.402
	IQR	186	10.07	36.11	84.18	32.46	166	36.7	155.4	47.83
F_{43}	Med	15.23	5.597	29.14	31.05	78.04	0.014	0.016	0.028	0.068
	IQR	41.95	7.067	21.59	64.31	21.09	0.102	0.207	0.273	4.97
F_{56}	Med	8.411	6.401	17.75	13.08	73	2.001	0.01	2.339	0.095
	IQR	44.61	7.406	29.15	71.09	89.02	63.2	5.084	59.13	15.41
F_{59}	Med	47.96	35.25	67.48	76.3	142.5	28.72	2.194	29.66	7.528
	IQR	224.4	20.01	50.33	109.6	61.18	173.3	33.31	165.8	48.16
F_{61}	Med	13.79	21.6	11.35	13.24	9.887	937	679.9	928.2	399.9
	IQR	5.371	9.521	2.319	6.958	2.441	294.6	862	794.6	856.6
F_{69}	Med	13.38	18.59	10.01	15.22	5.186	328.1	202.5	325.1	192.7
	IQR	8.179	5.792	7.763	12.19	3.53	160.6	66.07	167.5	77.29
F_{73}	Med	40.68	24.25	13.43	50.29	11.36	222.4	185.5	228.8	190.4
	IQR	258.8	178.7	247.7	272.9	247.6	416.9	351.1	387.7	329.1
F_{74}	Med	6.762	5.752	5.184	18.79	2.231	186.5	168.9	179	164.8
	IQR	48.57	26.17	35.07	61.88	37.68	333.6	199.7	296.7	193.2
F_{75}	Med	13.02	10.61	9.221	14.89	5.965	169.3	114.2	149.4	104.4
	IQR	35.37	101.5	83.9	33.2	79.22	424.4	331.1	411.2	311.6
F_{76}	Med	21.13	62.5	46.92	20.41	72.78	34.7	19.47	52.54	19.23
	IQR	112.3	183.5	179.2	128.6	200.7	208.7	140.2	205.7	137.6

Med: median, IQR: interquartile range, DB: database, DB1: the database obtained from [216]. DB2: the PASCAL database without noisy signals, DB3: the PASCAL database with noisy recordings. N: normal, M: murmurs, AS: Aortic stenosis, MS: Mitral stenosis, MR: Mitral regurgitation, MVP: Mitral valve prolapse

database (DB3); and feature matrix $X \in R^{3452 \times 13}$ using 1000 recordings (200 AS, 200 MS, 200 MR, 200 MVP and 200 normal) from the database obtained from [216] (represented as DB1). The length of the analysis window was also varied to 4 seconds to extract features. This produced a feature matrix $X \in R^{3250 \times 13}$ using 331 recordings from the database DB2; $X \in R^{3859 \times 13}$ using 480 recordings from the database DB3; and $X \in R^{4469 \times 13}$ using 1000 recordings from the database DB1.

The hold-out approach and 10-fold cross-validation were used to split the data (recorded signals) into training and test datasets. Thus, for hold-out, 80% of the data was used for the training and the remaining 20% of the data was used as a test set. For the 5 second window, this produced a training feature matrix $X_{training} \in R^{2762 \times 13}$ and test feature matrix $X_{test} \in R^{690 \times 13}$, where 2762 and 690 were the number of PCG instances or signal windows used for training and testing from the database DB 1, respectively. Similarly, the training set $X_{training} \in R^{2049 \times 13}$ and test set $X_{test} \in R^{512 \times 13}$ from the database DB 2; and training set $X_{training} \in R^{2426 \times 13}$ and test set $X_{test} \in R^{606 \times 13}$ from the database DB 3 were obtained. For the 4 second window, the training set $X_{training} \in R^{3576 \times 13}$ and test set $X_{test} \in R^{893 \times 13}$ from the database DB 1; training set $X_{training} \in R^{2600 \times 13}$ and test set $X_{test} \in R^{650 \times 13}$ from the database DB 2; and training set $X_{training} \in R^{3088 \times 13}$ and test set $X_{test} \in R^{771 \times 13}$ from the database DB 3, were evaluated. The number of recordings used to create the training and test dataset for the two different analysis windows is provided in Table 6.4.

The number of observations from the normal and pathological recordings used for training were found unbalanced. For instance, out of 2426 PCG instances or observations obtained for training using the database DB3, 1088 belonged to murmur recordings and the rest 1388 were from normal recordings. For this reason, the random undersampling boosting (RUSBoost) algorithm was used to address the problem of training classifier on skewed data [246]. The RUSBoost algorithm helps in handling imbalanced data and it is better compared to the use of '*Prior*' name-value pair argument or 'Cost' function in the classifier.

6.4.2 Model training

The feature vector evaluated in Section 6.3.1 was used to train the classifier and to extract global parameters of the classifier using the steps described in Chapter 3 (Section 3.4.2) (shown in Figure 5.2). Decision trees were used as weak learners due to their capacity to deal with high dimensional datasets. The boosting algorithm was used to optimize the weights and parameters of the weak learners at each iteration during the cross-validation [247]. Among several boosting algorithms (adaptive boosting (AdaBoost) [248], adaptive Logistic Regression (LogitBoost)

[247], random undersampling (RUSBoost) [246], GentleBoost [247], BrownBoost [249] and XgBoost [231]), the RUSBoost algorithm was used to handle imbalanced data and the LogitBoost algorithm to minimize the binomial deviance loss function, while fitting the model parameters, and because of its robustness against noise [264]. As the datasets used in this work contained normal and pathological recordings of different cardiac abnormalities, the classifier model was trained to predict the class based on the extracted features.

6.4.3 Hyperparameter optimization

Hyperparameters were optimized to achieve the best performance of the trained classifier model. This was achieved by statistically analysing the key parameters (learning rate, method, number of learning cycles, number of weak learners/trees and the maximum number of splits) with respect to the performance of the overall classifier in the training stage. The misclassification error was computed iteratively for different training sessions using 10-fold cross-validation. An optimum misclassification rate of 0.0138 was obtained for 30 number of trees, 3451 maximum number of splits for a learning rate of 0.75, while tuning the number of trees, the learning rate, and the maximum number of splits. In this study, a deep classification tree and a stump tree (1-level decision tree) were used as a reference to compare the misclassification rate during the training sessions.

Table 6.4: Recordings used to create training and test datasets.

Analysis window	Dataset	#PCG instances	Type of PCG instances
(a)	DB 1 (training)	3576	805 AS, 648 MR, 677 MS, 742 MVP, 704 Normal
	DB 1 (test)	893	201 AS, 161 MR, 169 MS, 185 MVP, 177 Normal
(b)	DB 1 (training)	2762	626 AS, 487 MR, 516 MS, 589 MVP, 544 Normal
	DB 1 (test)	690	157 AS, 123 MR, 128 MS, 147 MVP, 135 Normal
(c)	DB 2 (training)	2600	1294 murmurs, 1306 Normal
	DB 2 (test)	650	323 murmurs, 327 Normal
(d)	DB 2 (training)	2049	1023 murmurs, 1026 Normal
	DB 2 (test)	512	256 murmurs, 256 Normal
(e)	DB 3 (training)	3088	1376 murmurs, 1712 Normal
	DB 3 (test)	771	343 murmurs, 428 Normal
(f)	DB 3 (training)	2426	1088 murmurs, 1338 Normal
	DB 3 (test)	606	272 murmurs, 334 Normal

(a), (c) and (e) are the observations using 4 seconds window, while (b), (d) and (f), using 5 seconds window.

6.5 Experimental Results

6.5.1 Evaluation metrics

The evaluation metrics used to assess the performance of the approach were: recall or sensitivity $\left(S_E = \frac{TP}{TP+FN}\right)$, classification accuracy $\left(C_A = \frac{TP+TN}{TP+FP+FN+TN}\right)$, precision $\left(P_+ = \frac{TP}{TP+FP}\right)$, and F_1-score $\left(F = 2\frac{P_+ \times S_E}{P_+ + S_E}\right)$. S_E was computed to estimate the true positive rate and F_1-score to represent the harmonic mean of S_E and $P+$ [70]. For a 2 ×2 confusion matrix, true positive (*TP*) represents the number of abnormal PCG instances that are correctly classified as abnormal, false positive (*FP*) is the number of normal PCG instances that are incorrectly classified as abnormal, false negative (*FN*) is the number of abnormal PCG instances that are incorrectly classified as normal, and true negative (*TN*) is the number of normal PCG instances correctly classified as normal [75].

As recordings from the PASCAL datasets (DB 2 and DB 3) were labelled as normal recordings and murmurs, the performance metrics was assessed as a binary classification (2 × 2 confusion matrix). However, as the dataset (DB 1) contained recordings from 5 different classes (normal, AS, MS, MR, and MVP), the performance was evaluated using a 5 × 5 (five-class) confusion matrix, denoted as ***P*** [220]. Table 6.5 shows the interpretation of the confusion matrix, where P_{ij} represents the number of instances where the predicted class was *i* and the ground truth (or actual class) was *j*. The number of *TP*, *TN*, *FP* and *FN* for each class *k* were computed as described in Chapter 3 (Section 3.5). Thus, the evaluation matrices were computed as follows:

$$S_E = \frac{P_{kk}}{P_{kk} + \sum_{i \neq k} P_{ik}} \tag{6.1}$$

$$C_A = \frac{P_{kk} + \sum_{i,j \neq k} P_{ij}}{P_{kk} + \sum_{i,j \neq k} P_{ij} + \sum_{i \neq k} P_{ki} + \sum_{i \neq k} P_{ik}} \tag{6.2}$$

$$P_+ = \frac{P_{kk}}{P_{kk} + \sum_{i \neq k} P_{ki}} \tag{6.3}$$

6.5.2 Results

Table 6.6 shows the confusion matrices of the trained classifier on the training (using 10-fold cross-validation) and test sets from the databases DB1, DB2, and DB3 (DB1: the database obtained from [216], DB2: the PASCAL database with clean recordings, and DB3: the PASCAL database with noisy recordings). The performance parameters evaluated using the confusion matrices for the different datasets are shown in Table 6.7. The scores illustrated in

Table 6.5: Five-class confusion matrix.

		Actual class				
		AS	MR	MS	MVP	Normal
Predicted	AS	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}
	MR	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}
	MS	P_{31}	P_{32}	P_{33}	P_{34}	P_{35}
	MVP	P_{41}	P_{42}	P_{43}	P_{44}	P_{45}
	Normal	P_{51}	P_{52}	P_{53}	P_{54}	P_{55}

the table are the average values on all observations, counting each observation when it was in a held-out fold. As the performance of a classifier might have varied with respect to the window size, two windows (4 seconds and 5 seconds) were used to analyse the results. For instance, with an analysis window of 4 seconds, the classifier model achieved an average S_E of 92.29%, P_+ of 92.68%, F_1 of 92.43%, and C_A of 96.95% under the unseen test dataset from DB1. While using the same dataset, the algorithm achieved an average S_E of 95.51%, P_+ of 95.83%, F_1 of 95.64%, and C_A of 98.26%, with an analysis window of 5 seconds. For the normal versus murmurs classification using the PASCAL database, the classification accuracy for the different window sizes was almost the same. Thus, achieved results demonstrated that the approach is suitable for the analysis of recordings with a variable window size or duration.

The trained classifier was also tested with the noisy recordings from the PASCAL database (DB 3) to evaluate the performance parameters in classify signals recorded in real-world conditions. A very slight reduction in the classification accuracy (C_A) was observed in case of noisy recordings (from 92.31% to 91.70% for the 4 seconds window, and from 92.19% to 91.42% for the 5 seconds window, using test datasets DB2 and DB3). This leads to the conclusion that the approach is suitable for real world application where the recorded signals have noise superimposed.

The trade-off between the classifier's performance and the number of features used for its training was also investigated. This helped in assessing the computational cost of the approach with respect to the features used for the training. It was found that the computational cost (assessed in terms of the training time) of the system was reduced significantly with the 13-dimensional feature vector (F_{25}, F_{26}, F_{32}, F_{37}, F_{43}, F_{56}, F_{59}, F_{61}, F_{69}, F_{73}, F_{74}, F_{75}, and F_{76}). The system took 25.315 seconds for training with a 99-feature vector, whilst this was reduced to 12.637 seconds with a 13-feature vector, while having the same number of observations. Thus, the feature selection helped in minimizing the computational cost of the system.

Table 6.6: Confusion matrices of the trained classifier for different datasets.

(A = DB 1 Training Set, B = DB 1 Test Set, C = DB 1 Training Set, D = DB 1 Test Set, E = DB 2 Training Set, F = DB 2 Test Set, G = DB 2 Training Set, H = DB 2 Test Set, I = DB 3 Training Set, J = DB 3 Test Set, K = DB 3 Training Set, L = DB 1 Test Set) A, B, E, F, I, and J using a 4 seconds window and rest using a 5 seconds window

(A)

		Actual class				
		AS	MR	MS	MVP	Normal
Predicted	AS	772	27	23	14	13
	MR	14	615	6	21	0
	MS	9	1	638	11	0
	MVP	10	5	9	693	11
	Normal	0	0	1	3	680

(B)

		Actual class				
		AS	MR	MS	MVP	Normal
Predicted	AS	188	11	9	7	2
	MR	7	144	2	8	0
	MS	3	1	152	2	0
	MVP	3	5	6	167	1
	Normal	0	0	0	1	174

(C)

		Actual class				
		AS	MR	MS	MVP	Normal
Predicted	AS	622	2	2	1	0
	MR	1	484	0	2	1
	MS	1	0	513	0	0
	MVP	2	1	1	585	1
	Normal	0	0	0	1	542

(D)

		Actual class				
		AS	MR	MS	MVP	Normal
Predicted	AS	154	7	1	2	0
	MR	0	113	1	5	1
	MS	1	0	121	0	0
	MVP	2	3	5	139	1
	Normal	0	0	0	1	133

(E)

		Actual	
		M	N
Predicted	M	1282	15
	N	12	1291

(F)

		Actual	
		M	N
Predicted	M	299	26
	N	24	301

(I)

		Actual	
		M	N
Predicted	M	1365	13
	N	11	1699

(G)

		Actual	
		M	N
Predicted	M	1016	9
	N	7	1017

(H)

		Actual	
		M	N
Predicted	M	235	19
	N	21	237

(J)

		Actual	
		M	N
Predicted	M	312	33
	N	31	395

(K)

		Actual	
		M	N
Predicted	M	1073	21
	N	15	1317

(L)

		Actual	
		M	N
Predicted	M	249	29
	N	23	305

Table 6.7: Performance of the trained classifier using 10-fold cross-validation.

	Dataset	Type	S_E (%)	P_+ (%)	F_1 (%)	C_A (%)
(a)	DB 1 (training)	AS	95.90	90.93	93.34	96.92
		MR	94.91	93.75	94.31	97.93
		MS	94.23	96.81	95.50	98.32
		MVP	93.39	95.19	94.28	97.65
		Normal	96.59	99.41	97.97	99.22
	DB 1 (test)	AS	93.53	86.63	89.94	95.30
		MR	89.44	89.44	89.44	96.19
		MS	89.94	96.20	92.96	97.42
		MVP	90.27	91.75	91.00	96.30
		Normal	98.30	99.42	98.85	99.55
(b)	DB 1 (training)	AS	99.36	99.20	99.27	99.67
		MR	99.38	99.18	99.28	99.75
		MS	99.41	99.80	99.60	99.86
		MVP	99.32	99.15	99.23	99.67
		Normal	99.63	99.81	99.71	99.89
	DB 1 (test)	AS	98.08	93.90	95.94	98.12
		MR	91.87	94.16	93.00	97.54
		MS	94.53	99.18	96.79	98.84
		MVP	94.55	92.67	93.60	97.25
		Normal	98.52	99.25	98.88	99.57
(c)	DB 2 (training)	N & M	99.07	98.84	98.96	98.96
	DB 2 (test)	N & M	92.57	92.00	92.28	92.31
(d)	DB 2 (training)	N & M	99.32	99.12	99.22	99.22
	DB 2 (test)	N & M	91.80	92.52	92.16	92.19
(e)	DB 3 (training)	N & M	99.20	99.06	99.13	99.22
	DB 3 (test)	N & M	90.96	90.43	90.70	91.70
(f)	DB 3 (training)	N & M	98.62	98.08	98.35	98.52
	DB 3 (test)	N & M	91.54	89.57	90.55	91.42

(a), (c) and (e) are results using 4 seconds window, while (b), (d) and (f) are results using 5 seconds window. N: normal, and M: murmurs.

From the results tabulated in Table 6.7, it can be observed that the approach discussed in this work performs well in interpreting the class of signals based on the extracted features. It is to be noted that the classification performance achieved using the dataset DB1 was higher compared to the performance achieved using the PASCAL datasets (DB2 and DB3). This is mainly because the DB1 contains clean recordings, where files/records with noisy signals were excluded during the data acquisition [216].

6.6 Discussion

This chapter presented an approach to classify heart valve disorders using cardiac auscultatory recordings. To demonstrate the suitability and applicability of the approach in classifying pathological recordings obtained from varied controlled and uncontrolled conditions, the approach was validated using recordings obtained from the two large open access databases. This also helped in extracting physiological measurements from different pathological classes, which were subsequently used to train a machine learning based classifier to establish a classification boundary. Unlike previous studies that demonstrated either classification of normal and murmurs ([45], [91], [102]), or classification of normal and pathological sounds (AS, MR, MS, and MVP) [216], [271]; the approach described was tested for both types of classification. The obtained results lead to the conclusion that the features discussed are robust and suitable for different datasets. Furthermore, the experimental results in this work demonstrate that the features used to train the classifier performed well using both, clean and noisy recordings. This was validated using the PASCAL dataset DB3 that contains noisy recordings acquired in real uncontrolled conditions. Overall, the test dataset achieved an average classification accuracy of 91.70% and 91.42% for different window sizes, using the approach described.

As the electrical and mechanical activities of a human heart are not constant and vary widely with the pathological conditions, feature weights were evaluated using recordings from the subjects with different pathological conditions. The extracted features were statistically analysed to identify the top 10 features showing the best capabilities in differentiating normal recordings from different pathological recordings. From the weights illustrated for the top 10 features using different datasets in Table 6.2, it can be observed that the selected features achieved different weights for the different databases used in this study. Also, the variations in values (median and interquartile range) of the selected features (shown in Table 6.3) was observed for the different databases.

Among the top 13 features selected based on weights, 7 of them were found common using both the databases. Features (F_{32}, F_{37} and F_{43}) extracted using MODWT and the coefficients of MODWT multiresolution analysis (MRA) were found useful for the identification of S_1 and S_2 sounds. The deterministic and non-deterministic energy-based features (F_{25} and F_{26}) were identified among the top 13 features using recordings from all the databases. This is mainly because murmurs produced due to turbulence of blood flow, may be observed in a systole or diastole interval. Thus, the non-deterministic energy content present between the peak to peak intervals helped in distinguishing murmurs from normal recordings. This is also applicable for stenosis and regurgitation murmurs (such as AS, MS and MR) which leads to non-deterministic

information gathered in the intervals between the peaks. Thus, this work suggests features that can be potentially used for the classification of pathological conditions resulting in extra sound lobes between the peaks, such as S_3 and S_4 sounds.

While most of the approaches in the literature are available to classify heart sounds recordings as normal and abnormal, the presented approach is extended to identify recordings that belonged to different heart valve disorders, in relation to long-term continuous monitoring. Due to non-availability of large databases of recordings from the different pathological conditions, the method was validated using recordings from subjects with AS, MR, MVP, MS, and murmurs. The features discussed in this work could be also explored for the detection of other heart valve disorders such as aortic regurgitation (AR). Thus, this work is of importance for the future development of approaches for automatic detection of cardiac abnormalities using acoustic signals with the objective to provide continuous monitoring.

To highlight the novelty of this work, approaches reported in the literature to classify recordings from the databases used in this work, are compared in Table 6.8. Most of the articles used the PASCAL database to classify normal and murmurs [45]. Recently, the database DB1 containing pathological sounds (AS, MS, MR and MVP) was made available by [216]. The approaches to classify these pathological recordings included SVM [216], and random forest classifier [271]. To compare, the mean classification accuracy (C_A) achieved in this work using the different test sets: (C_A of 98.26% (P_+ of 95.83%) using the DB 1, and 92.31% (P_+ of 92.00%)

Table 6.8: Performance comparison with approaches for pathological classification using the databases used in this study

Ref.	Dataset	Classifier	P_+ (%)	C_A (%)
This work	DB1	Ensemble classifier	95.83	98.26
	The PASCAL database		92.00	92.31
[271]	DB1	Random forest	-	95.13
[286]	DB1	Multiclass composite classifier	-	98.33
[216]	DB1	SVM	-	97.90
[45]	The PASCAL database	SVM	91.00	-
[102]	The PASCAL database (Dataset-B)	SVM	83.00	-
[91]	The PASCAL database	SVM	91.00	-
[146]	198 recordings (open data)	SVM	-	93.42

using the PASCAL database DB2), was used. The achieved results demonstrate that this work outperforms other recent works, under the same databases.

Overall, the informative and non-redundant features evaluated in this work demonstrated potential in classifying acoustic signals of different cardiac conditions. Thus, the approach described can be used for possible detection and identification of pathological conditions using cardiac acoustic recordings in real use scenarios.

7 Conclusions

7.1 Contributions

This thesis presents four different algorithms to segment and classify fundamental heart sounds – S_1 and S_2; distinguish normal fundamental heart sounds recordings from abnormal recordings with additional lub/dub sounds; classify normal and abnormal recorded signals; and classify recordings from subject with heart valve disorders. These algorithms were validated using signals collected in this study as well as signals obtained from the public databases, to present a robust approach that can used for automatic analysis and classification of cardiac auscultatory recordings in real-world conditions. The suitability of the feature set in classifying both, short duration and long duration recordings was also tested.

Chapter 1 presented an in-depth review and critical analysis of all the existing approaches for the segmentation, feature extraction and classification of heart sounds signals used for automatic continuous monitoring and early-stage detection of various cardiovascular diseases. This chapter also included a detailed discussion on available public heart sounds databases and various approaches used by researchers to record heart sounds in reviewed articles. This was followed up with a discussion about the limitations and shortcomings of the existing approaches. From the literature review, it was concluded that a lot of work has been done in this field to tackle challenges in the continuous monitoring of the cardiac system. However, there is still need for intelligent automated algorithms to ease the task of classification with a high precision. It was also found that the majority of algorithms reviewed, do not meet the level of clinical accuracy expected while predicting the class of abnormality; and the results achieved cannot be extrapolated for long term monitoring. Few articles reviewed identified heart sounds using a specified threshold criterion and, in some studies, assumptions related to the time interval and the duration of peaks were used for the analysis. Moreover, most of the algorithms need to be validated using a large, standardized dataset that contains heart sounds recorded for longer duration, from different auscultation sites and for use in real conditions.

The literature review also revealed that the localization of sound peaks of interest and subsequent identification of fundamental heart sounds – S_1 and S_2, is an important initial step to achieve the task of automatic analysis and classification of recordings. To achieve this,

Chapter 3 proposed an approach to automatically segment and identify S_1 and S_2, and artifacts, using a combination of potential informative features. These included features based on MODWT, which have not previously been used for heart sounds analysis. This chapter also contributed to the derivation of the top 14 features that showed the greatest ability in differentiating fundamental heart sounds from artifacts. Unlike previous studies, for the first time, the performance of the algorithm was validated using a large database of heart sounds signals recorded continuously for 5 minutes at four different auscultation sites. The achieved results demonstrated the ability of the algorithm in classifying S_1 and S_2 sounds, as well as artifacts, with high sensitivity and F_1 score at different auscultation sites, without using any external reference or assumptions related to the systolic or diastolic timing intervals; showing its potential for use in systems for long-term monitoring in real-world conditions. The accurate identification of S_1 and S_2 provides the basis for measurements of many useful clinical attributes of cardiac acoustic signals and can potentially help in monitoring the overall cardiac health for longer duration. This includes the evaluation of heart rate and heart rate variability, and extraction of features helpful for further classification of cardiac acoustic signals as normal or abnormal recordings.

In Chapter 4, an approach to automatically segment heart sounds into cardiac events (including S_1 and S_2 as discussed in Chapter 3) and from there differentiate between normal fundamental sounds and additional abnormal sounds; where the latter included both splitting of fundamental, S_1 and S_2 sounds, and additional lub/dub sounds in the diastolic and systolic periods, was presented. Abnormal recordings with additional lub/dub sounds were detected without extracting the discrete subcomponents of S_1 (M_1 and T_1) or S_2 (A_2 and P_2) sounds. Apart from the features discussed in Chapter 3, 12 features were identified based on the normal S_1 and S_2 sounds, which were used subsequently to train an ensemble classifier. The top 10 features which showed the greatest ability in classifying abnormal additional lub/dub sounds were also identified. This method could have a potential application in the automatic detection of cardiac abnormalities that relate to the split of S_1 and S_2 sounds. It should be noted that, although, the proposed approach performed well on signals acquired in non-controlled environments, other features to assist with the extraction of subcomponents of S_1 and S_2 sounds could lead to further improvements.

Once the features that demonstrated the greatest ability in differentiating S_1 and S_2 sounds; and artifacts, were identified, they were used to classify recorded signals as normal and abnormal. Chapter 5 presented an automatic machine learning-assisted approach to classify long-term, continuously recorded cardiac auscultatory signals from different possible auscultation sites in real uncontrolled conditions. A novel feature vector was proposed to train an ensemble classifier to interpret the class of acoustic signals. The top 10 features which

showed the greatest ability to differentiate normal PCG signals from pathological signals were also identified using different databases. To demonstrate the robustness of the approach, results achieved were tested using signals recorded in this study; together with the signals obtained from the three large, widely accepted public databases. The proposed approach demonstrated accurate classification of recordings with an average classification accuracy of 94.17%, using signals collected in this work as well as signals obtained from the public databases. Based on the results achieved, the approach can be potentially used to automatically examine large data collected during long-term monitoring of cardiovascular system using a wearable device.

In Chapter 6, abnormal cardiac murmurs that are produced due to heart valve disorders such as aortic stenosis, mitral stenosis, mitral regurgitation, and mitral valve prolapse, were identified from cardiac acoustic recordings. A detailed characterization study of these pathological conditions was presented. In first-of-its-kind study, the top 13 most significant acoustic features, showing the greatest ability in classifying pathological (aortic stenosis, mitral stenosis, mitral regurgitation, mitral valve prolapse, and murmurs) recordings were extracted and analysed. Based on the dataset used in this study, it can be concluded that the presented approach can be used to classify recordings under controlled and uncontrolled conditions, and for different recording durations. This was tested using two large databases containing clean and noisy recordings obtained in real conditions. These findings are of great importance for the further development of automatic approaches for classification of recordings with abnormal sounds, such as S_3 and S_4, and heart valve disorders.

Overall, this thesis presented algorithms that can assist the automatic identification of cardiac abnormalities using cardiac auscultatory recordings. A supervised learning-based approach was utilized to classify heart sounds signals using true features extracted from labelled segments. A novel feature set that showed the greatest ability in differentiating normal recordings from abnormal recordings of different cardiac conditions, acquired in real settings was also presented. The results achieved in this work were compared with the recent works in this field and it was found that the presented algorithms showed improved performance, measured in terms of classification accuracy. Due to the underlining importance of long term monitoring for automatic early-stage detection of cardiac abnormalities, such as atrial fibrillation, in which events happen scattered in time; apart from signals available from the different open access databases, the approach presented was tested using signals recorded from subjects with and without cardiac diseases, in real-world conditions for 5 minutes from four auscultation sites. As features may vary with the signals they are tested on, the robustness of the approach was tested using different databases with varying recording conditions. The algorithms created in this thesis can potentially be used together with wearable systems for long-term monitoring using cardiac acoustics signals.

7.2 Further Work

Although the results achieved in this work showed improved performance compared to the previous works reported in literature, for automatic analysis and classification of cardiac sounds in real world conditions, there are a few avenues that could be explored in the future to improve the applicability and performance of the presented algorithms.

Cardiac acoustic recordings are heavily corrupted by noise and artifacts, mostly when used in uncontrolled non-clinical settings. It is because of this that there are many challenges which need to be overcome prior to having a truly wearable automated system for identification of cardiovascular diseases. Hence, one area of work should focus on looking at ways to minimise the unexpected noise/erroneous signals in the recordings.

Although there has been some work into extracting different characteristics of heart sounds in controlled conditions, hardly anything has been done when it comes to real settings, especially for sounds such as third and fourth heart sounds (S_3 and S_4), and artifacts. Controlled conditions experiments might be adequate for systems that are going to be used in clinic. But, to develop a system which could be used by the patient him/herself, any signal processing research going beyond the fundamental proof of concept needs to be done in real-life signals. For this reason, a large database of signals needs to be collected in real settings, and also from different auscultation sites.

This work could also be extended to a wider variety of pathological sounds. For instance, sounds due to the split components of S_1 and S_2, which emerge as an indicator of different valvular abnormalities, could be studied. The findings of this work can help in evaluating the split intervals (interval between M_1-T_1 in S_1 and A_2-P_2 in S_2) due to the mechanical activity of the heart during different cardiac conditions, which vary significantly within a subject and across subjects. The non-availability of a standardized database with a large number of cardiac acoustic recordings with marked locations of subcomponents of S_1 and S_2, using a reference ECG signals as a gold standard, also currently limits the extension of this work in this direction. Finally, the complexity of the algorithms would need to be optimized, taking into account specific system level constraints.

List of Publications

The following papers have been published or under process as part of this work.

Journal papers

- Amit Krishna Dwivedi, Syed Anas Imtiaz, and Esther Rodriguez-Villegas, "Algorithms for Automatic Analysis and Classification of Heart Sounds–A Systematic Review", *IEEE Access* 7 (2018): 8316-8345.
- Amit Krishna Dwivedi, Syed Anas Imtiaz, Stuart Bowyer, and Esther Rodriguez-Villegas, "Automatic Classification of S_1 and S_2 Sounds from Cardiac Auscultatory Recordings", *IEEE Access* (2020), (Under review).
- Amit Krishna Dwivedi, Syed Anas Imtiaz, and Esther Rodriguez-Villegas, "Automatic Classification of Normal Versus Abnormal Cardiac Auscultatory Recordings in Real-World Conditions", *IEEE Transactions on Biomedical Engineering* (2021), (Under review).

Peer reviewed conference paper

- Amit Krishna Dwivedi, and Esther Rodriguez-Villegas, "An Approach for Automatic Identification of Fundamental and Additional Sounds from Cardiac Sounds Recordings", In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6685-6688. IEEE, 2019.

Appendices

Appendix A: Pathophysiology of Normal and Abnormal Heart Sounds

Heart Sounds		Frequency Range	Qualitative sounds characteristics	Duration/location in the cardiac cycle	Cause	Description
First heart sound (S_1) or *lub* sound		10-200 Hz (lower pitch than S_2)	Dull and prolonged	0.12-0.15 seconds (longer than S_2)	Closure of the atrioventricular valves	Composed of M_1 and T_1 components
Second heart Sound (S_2) or *dub* sound		20-250 Hz (higher pitch than S_1)	Sharp and short	0.08-0.12 seconds (shorter than S_1)	Closure of the semilunar valves (at the end of systole)	Composed of A_2 and P_2 components
Third heart sound (S_3)		25-70 Hz (very low frequency, lower pitch than S_1 and S_2)	Soft and thudding quality	≈0.04 s, early-diastole (140-220 ms after S_2)	Early diastolic filling of the ventricle by blood rushing in from the atria	Due to the excess blood volume in the ventricle (left ventricle failure), benign in children and in pregnancy
Fourth heart sound (S_4)		15-70 Hz (lower than S_3, low pitch)	Weak and rumbling, less loud than S_1 or S_2	Late-diastolic/pre-systolic, slightly before S_1	Diastolic dysfunction because of the stiff ventricle	Due to atrial contraction, manifests coronary heart disease
Gallop	Ventricular (S_3 gallop rhythm)	15-50 Hz (very low frequency, low pitch, short and faint)	Galloping rhythm, lilt, trot or canter quality	≈ 0.15 s after S_2, early-diastole	During S_3 due to rapid deceleration of blood flow into the ventricle	Gallop rhythms indicates serious myocardial dysfunction
	Atrial (S_4 gallop rhythm)			≈ 0.08 – 0.20 s just before S_1	During S_4, due to decreased ventricular distensibility	S_4-S_1 may be confused with a split S_1
	Summing (both S_3 and S_4 gallops)		Quadruple rhythm, loud sound	During diastole period	Both S_3 and S_4 are superimposed	Occurs with the improvement of heart failure
Murmurs	IM	120-250 Hz (mid-range frequency, high pitch)	Whooshing, roaring, turbulent fluid noise	Mostly in early-systole, short duration	Due to the turbulent flow of the blood which may occur inside or outside the heart	Frequently heard in children and during exercise, position dependent
	SM	Up to 600 Hz (frequency range usually lower than DM, high pitch)	Rasping and blowing, usually crescendo-decrescendo	Early-, mid-, late- or holo-systolic	Mechanical systolic and ventricular ejection	Includes PS, AS, ASD, HOCM, MI, TI, MVP, PDA and VSD
	DM	Up to 600 Hz (high or low pitch)	Puffing and rumbling quality, usually decrescendo	Early-, mid- or late-diastolic	Ventricular relaxation and filling	Difficult to hear, Includes PI, AI, MS, TS
Opening snaps		100-800 Hz (high pitch)	Snapping sound	During diastole period	Inspissating of valve leaflets	The sudden opening of the stiff mitral valve
Rubs			Scratching, harsh, creaking sound	Loudest in systole and can be heard in beginning and end of diastole	Friction between layers, abrasion of pericardial surfaces	Depends on the body position and breathing
Clicks			Short and loud	Very early systole, mid-systolic click, may be immediately after S_1	Due to the opening of a rigid and calcifies aortic or pulmonary valve	Includes AEC, SEC, mechanical valve click, or prosthetic valve click

*Information regarding frequency range and timings from [8], [13], [37], [273]–[275]
Note: IM: innocent murmurs, SM: systolic murmurs, DM: diastolic murmurs, PS: pulmonary stenosis, AS: aortic stenosis, ASD: atrial septal defect, HOCM: hypertrophic obstructive cardiomyopathy, MI: mitral insufficiency, TI: tricuspid insufficiency, MVP: mitral valve prolapse, PDA: patent ductus arteriosus, VSD: ventricular septal defect, PI: pulmonary insufficiency, AI: aortic insufficiency; TS: tricuspid stenosis, MS: mitral stenosis; AEC: Aortic ejection click, SEC: systolic ejection click.

Appendix B: Databases used for Validation of Algorithms for Heart Sounds Analysis

Reference	Source	Type of sound/ # Subjects	# recordings	Age (Mean ±SD) years	Gender (M/F)	Sensor	Sensor position	Format	Total participants	Sampling rate	Frequency band	Duration (Seconds)	Comments
(a) Public heart sounds databases													
[8], [20], [22]	PhysioNet database (Computing in Cardiology Challenge 2016)	Normal and pathological sounds	4,430 records (including training and test sets)	Variable	Variable	Variable	Variable	Variable	1297 (1072 included for training and test dataset)	Variable	Variable	5-120 s each	Open access database, comprises nine databases, performance of dataset reported in [21], [31]
[32], [33]	MITHSDB (collectively from [32], [33])	Normal (38)	117 records	-	-	Welch Allyn Meditron stethoscope (Skaneateles, USA)	9 different positions and orientations	-	121	44.1 Hz	20 Hz-20k Hz	9-37 s each (33±5 s)	Simultaneously recorded with ECG, MITHS database included in [8]
		IM (34)	118 records										
		MVP (37)	134 records										
		AD (5)	17 records										
		MPC (7)	23 records										
[34], [35]	AADHSDB (collectively from [34], [35])	Normal (121)	544 records	-	93M/58F	‡	4th ICS at the left sternal border on the chest	-	151	4 kHz	20 - 1000 Hz	8 s each	AADHS database included in [8]
		CAD (30)	151 records										
[39]	UHAHSDB (database in PhysioNet [20])	Normal (25)	39 records	18 to 40 years	-	Prototype electronic stethoscopes	-	WAV format	55	8 kHz	-	6-49 s each	UHAHS database included in [8]
		Pathological sounds (30)	40 records	44 to 90 years	20M/10F								
[36]	AUTHHSDB (database in PhysioNet [20])	Normal (11)	11 records	29.3±10.7	4M/7F	Custom-made electronic stethoscope	Auscultation for valve murmurs & apex area	WAV format	45	44.1 Hz	4 kHz	10-122 s each (50 ± 26 s)	AUTHHS database included in [8]
		MR (17)	17 records	75.3±10.2	6M/10F								
		AS (17)	17 records	76.1±7.2	6M/10F								
[37], [38]	TUTHSDB (collectively from [37], [38])	Normal (28) and pathological (16)	174 records	-	-	-	Aortic, pulmonic, apex and tricuspid positions	-	44	4 kHz	-	15 s each position	TUTHS database included in [8]
[40]–[43]	DUTHSDB (database in PhysioNet [20])	Normal (174)	338 records	4 to 35 years (25 ± 3 years)	172M/2F	Microphone sensor (MLT201) or piezoelectric	Various sites	WAV format	509	8 kHz – 22050 Hz	Variable	209 ± 78 s each	DUTHS database included in [8]
		CAD (335)	335 records	10 to 88 years (60 ± 12 years)	108M/227F	‡	Mitral site at the chest			8 kHz		17 ± 12 s each	
[44]	SUAHSDB (database in PhysioNet [20])	Normal (79) and pathological (33)	114 records	16 to 88 years (56 ± 16 years)	43M/69F	An electronic stethoscope	Apex	-	112	8 kHz, 44.1 kHz and 384 kHz	20-1000 Hz	30-60 s (33 ± 5 s) each	SUAHS database included in [8]
[8]	SSHHSDB (database in PhysioNet [20])	Normal (12) and pathological (23)	35 records	-	-	-	2nd intercostal	-	35	8 kHz	-	15-69 s (36 ± 12 s) each	SSHHS database included in [8]
[8]	SUFHSDB (database in PhysioNet [20])	Fetal (116)	119 records	-	-	An electronic stethoscope (GS Technology Co. Ltd, South Korea)	-	8 kHz and 44.1 kHz	225	8 kHz and 44.1 kHz	20-1000 Hz	Average 90 s each	SUFHS database included in [8]
		Maternal (109)	92 records	29 ± 6 years	109F						-		
[13]	Daniel Manson 2000	Normal and pathological sounds	Over 180 records	-	-	-	-	WMA format	-	-	-	Variable	Book with CD recordings, reported in [84]
[23]	The PASCAL database (CHSC 2011)	Dataset A: Normal, murmur, EHS & artefact	176 records	-	-	†	-	WAV/ aif format	-	44.1 kHz	-	1-30 s each	Segmented data with S_1 and S_2 location, reported in [102], low-pass filtering at 195 Hz
		Dataset B: Normal, murmur and extrasystole	656 records			‡				4 kHz			
[24]	OMHSML (The University of	Normal and pathological sounds	23 records	-	-	Stethoscope	Apex, aortic and pulmonic area	MP3	-	44.1 kHz	-	Total of 1496.8 s	Reported in [46], [48]

	Michigan database)												
[25]	CAHMDB (eGeneral Medical Inc. database)	Pathological and non-pathological	64 records	-	-	-	-	WAV format	-	11/8 kHz	-	Total of 338 s (1-10 s each)	Requires permission, reported in [8], [78], [80], [276]
[26]	Heart sound library by Thinklabs	Normal and pathological	41 records	-	-	Thinklabs stethoscope	-	-	-	-	-	Variable	Available online, reported in [46]
[27]	Texas Heart Institute database	Pathological (50) (information from [82])	50 records	-	-	Stethoscope	-	MP3	-	44.1 kHz	-	Variable	Database reported in [82], more information in [277]
[28]	BHSD	Normal and pathological	25 records	-	-	-	-	WAV format	-	-	-	Total of 49.94 s	Online available, reported in [95]
[29]	CARD database	Normal and pathological	Variable	-	-	Hewlett Packard 21050A microphone	-	MP3 format	-	-	-	20 s each	Login required to access database, ECG records available, reported in [95], [72]
(b) Heart sounds recorded by researchers (not available freely)													
[32]	Syed 2003	Normal (30) and MVP (21)	102 records	-	-	-	Apex and left lower sternal border	-	51	44.0 Hz	-	-	Simultaneously recorded EKG signals, included in [8]
[33]	Sayed *et al* 2007	Normal murmurs (15), MR (11) and normal (13)	39 records	-	-	‡‡	Left apex and left parasternal	-	39	44.1 Hz	4096 Hz	30-40 s each	Simultaneously recorded ECG signals, included in [8]
[34]	Schmidt *et al* 2010	Normal (13) and pathological (100)	73 records	-	-	‡ (handheld stethoscope)	4th ICS at the left sternal border on the chest	-	40 (training data) and 73 (test data)	4 kHz	-	8 s each	Without ECG signals, database included in [8]
[35]	Schmidt *et al* 2015	Non-CAD (70) and CAD (63)	231 (Non-CAD) and 204 (CAD)	60.1±9.9 (Non-CAD) & 66.1±10.7 (CAD)	35M/35F (Non-CAD) & 47M/16F (CAD)	‡	4th ICS at the left sternal border on the chest	-	133	4 kHz	20-1000 Hz	8 s each	Unsupervised, database included in [8]
[36]	Papadaniil and Hadjileontiadis 2014	Normal (14)	11 (698 cycles)	29.3±10.7	4M/7F	Electronic stethoscope	Auscultation for valve murmurs & apex area	WAV format	52	44.1 Hz	4 kHz	10-122 s each	AUTHS database included in [8]
		MR (19)	16 (827 cycles)	75.3±10.2	6M/10F								
		AS (14)	16 (1077 cycles)	76.1±7.2	6M/10F								
[37]	Naseri and Homaeinezhad 2013	Pathological (AS, AR, MS, MR) (50)	50 records	48 years (average)	25M/25F	‡	Aortic, pulmonic, apex and tricuspid positions	-	50	4 kHz	700 Hz	Total of 52 minutes	TUTHS database included in [8]
[38]	Naseri *et al* 2013	Normal and pathological (AS, AR, MS, MR)	63 normal and 63 pathological	-	-	‡	Aortic, pulmonic, apex and tricuspid positions	-	-	4 kHz	-	2 minutes each	TUTHS database included in [8]
[39]	Moukadem *et al* 2013	Normal	40 records	18-40 years	-	Prototype electronic stethoscopes	-	WAV format	-	8 kHz	-	6-12 s each	UHAHS database included in [8]
		pathological sounds	40 records	44 to 90 years									
[40]	Li *et al* 2011	Normal and pathological	27 cardiac cycle	-	-	-	Mitral site	-	-	2 kHz	-	≈ 9 s each	DUTHS database included in [8]
[43]	Tang *et al* 2012	Normal (3) and pathological (23)	26 records	-	-	Vibration (piezoelectric) sensor	Mitral site	-	26	2 kHz	Below 600 Hz	-	Simultaneously recorded with ECG, DUTHS database included in [8]
[48]	Sun *et al* 2014	Normal (45)	3390 s of CHD, 3940 s of RHD and 600 s of normal HSs	-	-	‡	Tricuspid site	-	121	44.1 kHz	20 – 20k Hz	600 s	Michigan HSs database [24] also considered for validation
		ASD (14)										620 s	
		F4 (7)										270 s	
		PDA (10)										550 s	
		VSD (33)										1290 s	
		Mixed CHD (12)										660 s	
		RHD										3940 s	
[106]	Amit *et al* 2009	HSs with alternating breath resistance	10 records	29±12 y	8M	-	-	-	12	-	20-250 Hz	40 s each	Single-lead ECG
		Pharmacological stress data	-	60±14 y	11M				11			30-45 minutes each	Dobutamine stress echocardiography
[19]	Ari *et al* 2008	Normal	71 records						71				

		Nine different pathological sounds		No age group associated	-	‡ and ⁑	-	WAV format		8, 11, 22 & 44.1 kHz	Below 150 Hz	-	ECG signals for validation
[74]	Gupta *et al* 2007	Normal and murmurs	340 HSs segments	-	-	-	-	-	41	8 kHz	4 kHz	-	No ECG information used
[52]	Javed *et al* 2006	Pathological sounds	40 records	16 to 79 years	-	‡	-	.e4k and WAV format	40	8 kHz	-	8 s each	Unsupervised
[133]	Turkoglu *et al* 2003	Normal (95) and pathological (120)	215 records	15 to 80 y (average 48.77 y)	132M/83F	Acuson Sequoia 512 Model Doppler Ultrasound system	Chest	-	215	20 kHz	0.5-10 kHz	5 s each	Doppler HSs signals
[143]	Safara *et al* 2013	Normal (16) and pathological (43)	59 records, 804 heart cycles	-	-	Welch Allyn Meditron and ECG analyzer, Meditron	-	-	-	4 kHz	-	15 s each	ECG is recorded simultaneously
[138]	Banerjee and Mondal 2015	Pathological (60)	-	-	45M/16F	⁑	-	WAV format	86	8 kHz	2 kHz	-	Recordings in sitting and relaxed conditions
		Non-pathological (60)			15M/10F								
[64]	Iwata *et al* 1980	N (60), FM (40), MI (32), AI (38), ASD (9), & VSD (8)	187 samples, 881 cardiac cycle	-	-	4-channel PCG (Fukuda Densi EMR60SD)	Cardiac apex area	Magnetic tape	69	2 kHz	880 Hz	-	Supervised (ECG and PCG are recorded)
[135]	Zheng *et al* 2015	Normal (88)	88 records	18-60 y (mean 35.64±7.5)	44M/40F	Color Doppler Ultrasound medical machine (Vivid-7, GE company, USA) and PCG signals	Apex	-	152	11025 Hz	-	One minute each	Unsupervised
		CHD patients (64)	64 records	38-70 y (mean 61.56±9.7)	36M/28F								
[142]	Zhang *et al* 2014	N (225) and abnormal (180) (MS (60), VSD (60) & AS (60))	405 (225 (N) and 180 (AN))	-	-	Phonocardiogram sensor	Precordium	-	405	11025 Hz	2205 Hz	-	Unsupervised
[76]	Guillermo *et al* 2015	Pathological and non-pathological	92 cardiac cycles (38 (PI) and 54 (TI))	-	-	Microphone	-	-	-	8 kHz	30-600 Hz	15 s each	Unsupervised
[278]	Elgendi *et al* 2014	Pathological and normal	-	3 months to 19 years	12M/15F	‡	Cardiac apex and 2nd LICS	WAV format	27	4 kHz	-	Over 20 seconds each	Supervised
[71]	Hassani *et al* 2014	N (5), VSD (36), ASD (12) & ASD and VSD (6), TOF (10), PS (10) & 21 others	100 records (14,000 cardiac cycles)	1-26 y	-	-	-	-	100	44.1 kHz	Below 800 Hz	-	Unsupervised
[125]	Choi and Jiang 2010	N (6) and pathological (34) (AF, AI, AS, MR, MS and Split)	196 (N) and 293 (AN) records	30±14 y (N), 47±19 y (AN)	-	Electronic stethoscope	Four auscultation sites	-	40	8 kHz	700 Hz	12 s each	Unsupervised
[45]	Zhang *et al* 2017 (The Pascal CHSC 2011)	N (45), M (48), EHS (27) and artefact (56)	176 records	-	-	†	-	WAV format	-	44.1 kHz	2 kHz	-	Unsupervised
		N (336), M (105), extrasystole (66)	507 records			‡				4 kHz			
[70]	Springer *et al* 2016	Normal (38), MVP (37), IM (36), AD (5), MPC (7)	S_1: 12181 and S_2: 11627	-	-	⁑	Parasternal, apical, aortic, and pulmonic	-	123	44.1 kHz	1 kHz	Total of 10172 s	Verified using ECG
[165]	Turkoglu *et al* 2003	Normal and pathological (AV and MV)	215 records	15-80 y (average 48.77 y)	132M/83F	Acuson Sequoia 512 Model Doppler Ultrasound system	-	-	92 for training and 123 for testing	20 kHz	500 Hz	5 s each	Doppler HSs signals
[157]	Uğuz 2012	Normal (40), PS (40) and MS (40)	-	4-65 y	55M/65F	‡	-	e4k/ WAV format	120	8 kHz	-	-	Unsupervised
[95]	Varghees and Ramachandra	Normal heart sounds	25 records	-	-	Microphone	-	-	-	2 kHz	800 Hz	-	Databases from other sources

	n 2017 (AUHSD)												also used for validation
[89]	Kang *et al* 2017	Innocent murmurs (87) and pathological murmurs (170)	257 records (1212 cycles)	-	-	‡	RUSB, LUSB, LMSB, LLSB and apex	-	-	8 kHz	40-500 Hz	3-8 s each	Unsupervised

* iStethoscope is indicated with symbol '†', electronic stethoscope 3M Littmann is indicated with symbol '‡' and Meditron (NY, USA) stethoscope with symbol '⁂'
Note: N: normal heart sounds, M: murmurs, EHS: extra heart sounds, CAD: coronary artery disease, CHD: congenital heart disease, EAS: early aortic stenosis, LAS: late aortic stenosis, FM: functional murmurs, MI: Mitral insufficiency, AI: aortic insufficiency, ASD: Atrial septal defect, VSD: Ventricular septal defect, PCHSC: Pascal classifying heart sound challenge, IM: innocent murmurs, MPC: miscellaneous pathological condition, AD: aortic disease, LICS: left intercostal space, ICS: intercostal space, MVP: mitral valve prolapse, AS: aortic stenosis, AR: aortic regurgitation, MR: mitral regurgitation, MS: mitral stenosis, RUSB: right upper sternal border, LUSB: left upper sternal border, LMSB: left middle sternal border, LLSB: left lower sternal border. AUHSD: Amrita University heart sounds database, TOF: tetralogy of Fallot, AF: atrial fibrillation, MITHSDB: The Massachusetts Institute of Technology heart sounds database, AADHSDB: The Aalborg University heart sound database, AUTHHSDB: The Aristotle University of Thessaloniki heart sounds database, UHAHSDB: The University of Haute Alsace heart sounds database, TUTHSDB: The K N Toosi University of Technology heart sounds database, DUTHSDB: The Dalian University of Technology heart sound database, SUAHSDB: The Shiraz University adult heart sounds database, SSHHSDB: The Skejby Sygehus Hospital heart sounds database, SUFHSDB: The Shiraz University fetal heart sounds database.

Appendix C: Summary of Heart Sounds Segmentation Methods and their Performance Comparison

Year	References	Segmentation Methods	Signal Type	# Subjects	Records (# and duration in seconds)	Cardiac Cycles (or periods)	Noise tolerance	S_1	S_2	S_3	S_4	Murmurs	Performance			Comments
													S_e (%)	*PPV* (%)	Accuracy (%)	
Envelope analysis																
2018	[90]	Shannon envelope and threshold adjustment	Normal (1) and pathological (4)	5	5 records	-	-	●	●	●	●	-	98.1	-	-	Only detection of sounds S_1 to S_4 was reported, validated with a limited number of recordings.
2017	[95]	Shannon entropy envelope and instantaneous phase	Normal and pathological	Databases from [23], [8], [25], [28], [24], [29] and HSs, Littmann HSs, WHSMD and AUHSD (Total 278 records)			SNR 10 dB	●	●	●	●	●	N: 94.38, M: 97.58	N: 97.25, M: 96.46	N: 91.92, M: 94.21	Information was extracted after splitting signals as low frequency and high frequency signal contents
2017	[89]	Shannon energy envelope	Normal	-	257 records	627 cycles	-	●	●	-	-	●	N: 99.2, PM: 89.3	N: 99.2, PM: 96.4	-	Recordings form children, murmurs were segmented, assumption
			Innocent murmurs			376 cycles										
			Pathological murmurs			209 cycles										
2017	[45]	Scaled spectrogram and PLSR	Normal, murmurs and extrasystole sounds	Datasets A & B (176 & 507) records, database from [23]			-	-	-	-	-	●	-	-	-	PLSR for identifying most relevant features, only heart cycles were identified
2015	[46]	CSCW envelope	Pathological	11 (AF)	-	152 periods	-	●	●	-	-	-	-	-	-	AF, AR, MR, PS, VSD and normal sounds were identified (Classification accuracy was reported)
				15 (AR)		169 periods										
				22 (MR)		231 periods										
				20 (N)		268 periods										
				6 (PS)		138 periods										
2014	[47]	Moving windowed HT	Normal and VSD HSs	16 healthy and 37 VSD patients	242 (N) and 226 (VSD)	-	-	●	●	-	-	●	98.8	-	-	VIM for heart sounds envelope analysis, *CA*: 98.4, higher sensitivity than threshold based methods
2014	[48]	STMHT	Normal	45	600 s	-	-	●	●	-	-	-	-	-	S_1: 98.53, S_2: 98.31, Cycle: 97.37	Unsupervised, using Hilbert transform and the Viola integral waveform method
			CHD	76	3390 s											
			RHD	-	3940 s											
			Database [24]	23	1496.8 s											
2013	[39]	Shannon envelope using S-transform	Normal (40)	-	80 records (6-12 s)	-	-	●	●	-	-	-	96	95 (using SSE)	-	Localization tolerance is absent, unsupervised
			Pathological (40)										97			
2010	[49]	CSCW (based on variance on PCG timing scales)	Normal and pathological	-	9 records (≤ 5 s each)	-	-	●	●	-	-	●	-	-	99.11	Results validated using a very small dataset, Viola integral method applied
2009	[50]	Short time spectral energy & autoregression characteristics	Normal (20)	120	Total of 1200 s	823 cycles	-	●	●	-	-	-	-	-	93.6	Pediatric heart sounds segmentation,
			Abnormal (40)													
2008	[51]	CSCW (also reported in [6])	Normal and abnormal	-	-	500 cycles	-	●	●	-	-	-	-	-	N: 100 and AN: 88.2 (MODSG2)	CSCW achieved higher accuracy compared to Shannon envelope and Hilbert envelope
2006	[52]	Normalized average Shannon energy	Pathological and non-pathological	40	8 s each	120 systolic periods	-	●	●	-	-	●	-	-	-	Patients aged between 16-79 years were considered, Segmentation results were not found
1997	[53]	Normalized average Shannon energy	Pathological (14) & physiological (23)	-	37 records	515 periods	-	●	●	-	-	-	-	-	93.00	No individual S_1 and S_2 identification reported
ECG pulse reference based methods																
2017	[54]	Probability analysis for feature extraction	Normal and abnormal	PhysioNet database [8]			-	●	●	-	-	-	76.96	-	-	ECG based segmentation using R-R interval estimation, *CA*: 84.11%
2011	[55]	Joint ECG and PCG signals	Normal	120	120 records (Each 10 s)	1976 cycles	-	●	●	-	-	-	-	-	S_1: 97.00, S_2: 94.00	Results are reported for paediatric HSs segmentation
			Pathological (80)													
2011	[57]	Using MMP	Normal (35)	70	Each 10 s	-	-	●	●	-	-	●	-	-	-	ECG recorded simultaneously, *CA*: 92.5% (using MMP classifier)
			Pathological murmurs (35)													
2009	[58]		Normal (50)	148		360 beats							100	92.0		

		Fractal features for segmentation	Murmurs (98)		164 records ≤ 8 s		-	●	●	-	-	●			-	Intra-beat segmentation (ECG records)
2006	[59]	Feature selection with ECG signal	Valvular heart patient	36	15 s for each subject	445 cycles	-	-	-	-	-	●	-	-	-	207 features extracted, *CA*: 86 (For MI, AS and PM)
2005	[60]	R- and T-waves of ECG as reference	Data from known cardiac defects	300	40-45 s each	-	SNR > 40 dB	●	●	-	-	●	-	-	S_1: 100, S_2: 97.00	ECG gating applied, STFT, sounds form children
2001	[61]	Matching pursuit method	Normal and abnormal	15 (each for N & AN BMV)	-	-	-	●	-	-	-	-	-	-	-	ECG signals are also recorded at 500 Hz, *CA*: 93.00%, no segmentation results founds
1992	[62]	Microcontroller based HSs gating	Pathological	19	-	-	-	●	●	-	-	-	-	-	-	Gating device for medical imaging
1987	[63]	ECG and carotid pulse based	Normal and pathological	5 healthy and 20 CVD patients	94 signals (10 s)	-	-	●	●	-	-	-	-	-	-	Systolic and diastolic segmentation
1980	[64]	Segmentation using spectral tracking	Pathological and non-pathological	60 (N), 40 (FM), 32 (MI), 38 (AI), 9 (ASD) & 8 (VSD)	-	187 samples with 881 cardiac cycle	-	●	●	-	-	-	-	-	-	Zero-crossing positions are used for identification
Probabilistic models																
2017	[67]	Modified Springer's method [70]	Normal and pathological	PhysioNet database [8]			-	●	●	-	-	-	S_e: 91.38 (best)	-	-	Segmentation using CQA
2017	[68]	Springer's HSMM and Viterbi decoding	Normal and pathological	PhysioNet database [8]			-	●	●	-	-	-	-	-	-	Manual segmentation of HS signals
2017	[69]	Duration-dependent HSMM	Pathological and non-pathological	PhysioNet database [8]			-	●	●	-	-	-	-	-	79.30 (all dataset)	Segmentation algorithm by Springer *et al* [70]
2016	[70]	Logistic Regression-HSMM	Normal and pathological	123	10172 s (total)	S_1: 12181 S_2: 11627	-	●	●	-	-	-	95.34 ±0.88	95.92±0.83	92.52±1.33	HSMM performed better than HMM, error rate 0.23 with Shannon energy
2014	[71]	HF for segmentation	Normal and pathological	100 (N and AN)	100 records	14,000 cardiac cycles	-	●	●	-	-	-	S_1: 80.3 S_2: 77.5	S_1: 80.2 S_2: 77.6	-	Time-domain intensity envelopes, results not suitable for clinical applications
2012	[43]	Dynamic clustering	Normal and abnormal	3 healthy and 23 patients	26 records	565 cycles	-	●	●	-	-	-	-	-	S_1: 94.86, S_2: 95.92	Localization tolerance is absent, No split between training sets
2012	[72]	Hilbert-Huang Transform	Normal and pathological	800 records form CARDJHU [29]		15 records with S_3 and S_4	-	-	-	●	●	-	S_3: 90.40, S_4: 94.50	S_3: 90.4, S_4: 85.5	-	S_1 and S_2 were not identified
2012	[73]	HMM based segmentation	N and abnormal	80 (N) & 80 (Abnormal)	-	-	-	●	●	-	-	●	-	-	-	HMM based segmentation, *CA*: 85.6%
2010	[34]	Duration-dependent HMM	Normal and pathological	73	113 records (each 8 s)	-	-	●	●	-	-	-	98.8	98.6	-	HMM achieved 59.9% of *Se* and 54.8% of *PPV* for the same algorithm
2007	[74]	HF and *k*-means clustering	Normal and murmurs sounds	41	-	340 HSs segments	-	●	●	-	-	●	-	-	N: 99.09, SM: 85.47 & DM: 86.47	Not suitable for high intensity murmurs [19], unsupervised
2005	[65]	HE and self-organizing PM	Normal and abnormal	17	44 PCG records (30-60 s)	-	-	-	●	●	-	-	-	S_1: 98.6 S_2: 98.3	S_1: 96.9, S_2: 96.5	-	Extracted features from events are segmented using HMM
2005	[66]	HMM	Clean and dirty	9	46 files (~2286 s)	-	-	-	●	●	-	-	-	-	-	98	Shannon energy features are extracted for segmentation
Feature-based Methods																
2017	[75]	Identification of S_1 and S_2 using DNN	Normal and abnormal	28	-	460 each S_1 and S_2 HSs	-	●	●	-	-	-	-	-	-	Acoustic features using MFCC, *CA*: 91.12%, 17M/11F were considered
2015	[76]	Detection of heart sound murmurs	Pathological and non-pathological	-	38 (PI) and 54 (TI)	92 cardiac cycles	-	●	●	-	-	●	-	-	-	Classification of murmurs only
2015	[77]	Energy fraction and entropy based features	Normal (40) / CVD patients (67)	107	80 (N) and 167 (SHM)	-	-	●	●	-	-	●	93.48	-	-	Unsupervised, S_p: 98.55% and *CA*: 97.17%
2014	[36]	EEMD & kurtosis features	N (11) and pathological (32)	43	43 records	2602	-	●	●	-	-	-	-	-	83.05±15.14	HSS-EEMD/K achieved prediction power of 94.56±6.58%

2014	[78]	HSAD method	Normal and pathological PCG	Online available [25] (each less than 10 s)		701	Up to 5 dB	●	●	●	●	●	99.43	93.56	93.95	Tested with limited database, no stated segmentation tolerance
2013	[37]	Frequency-energy based metric	Different valve disease	50	Total of 52 min	-	-	●	●	●	●	●	99.00	98.60	-	S_3, S_4 and murmurs were also investigated
2008	[19]	Clinical features for locating FHSs and murmurs	Normal 9 different pathologies	71	71 records	357 cycles	Up to 25 dB Up to 10 dB	●	●	-	-	●	-	-	97.47 (overall)	Robustness analysis and no split between training sets, systolic and diastolic murmurs
2005	[79]	Complexity-based segmentation	Online database [25]	-	-	-	-	●	●	-	-	●	-	-	-	No ECG gating used, not suitable for continuous murmurs.
Time-frequency/wavelet analysis based methods																
2018	[101]	Wavelet analysis and Shannon energy calculation	Normal sounds and abnormal sounds	230 subjects	230 records	-	-	-	●	-	-	●	-	-	-	Statistical results for the segmentation were not found.
2013	[82]	Tunable-Q wavelet transform	Pathological signals	Database from [27]			-	●	●	●	●	●	-	-	*SR*: 92.15 (overall)	Envelope based on CSCW
2011	[80]	Time-domain analysis of pathological PCG signals	Normal EAS LAS PS MR	Database from [25] (6 records for each class)			-	●	●	●	-	●	-	-	-	Unsupervised, TFD analysis using RME
2011	[81]	ASSA approach	Normal and pathological	12	20 s each	8 segments	-	-	-	-	-	●	-	-	-	Welch Allyn Tycos stethoscope, only correlation estimation
2008	[83]	Wavelet transform for segmentation	14 pathological HSs signals	Data from two patients, online and CD book	Each record of 20 cycles	140 cycles	-	●	●	-	-	-	-	-	-	Statistical results for the segmentation were not reported.
2007	[84]	Segmentation and feature extraction using wavelets	Normal and pathological (AR)	Database from [13]			-	●	●	-	-	-	-	-	-	Statistical results for segmentation were not reported.
Learning based methods																
2014	[85]	A system for heart sounds classification	Normal and pathological	Database from [279]	6 records for each class	72 Signals	-	●	●	●	●	●	-	-	S_1: 94.63, S_2: 97.25, S_3: 96.63, S_4: 96.66	Pathological and non-pathological signals are considered, No reported methods for segmentation
2010	[41]	Fuzzy detection method	Normal (3) Pathological (23)	26	26 records	565 cycles	0 dB and -5 dB	●	●	-	●	●	-	-	Best is 94.0% for 0 dB SNR	Noise reduction technique applied
2008	[86]	Fuzzy clustering approach	Database from [25], Normal (5) and SM (15)	-	20 records	-	-20 dB to 60 dB	-	-	-	-	●	73.0	-	SM: 80.00	S_p: 100.00% was reported
2002	[87]	Time-delay neural network	Normal and pathological	30	Each 20 s	-	-	●	-	-	-	-	98.4	97.8	-	ECG (training) with time domain analysis

*Note: PLSR: partial least squares regression, RHD: rheumatic heart disease, MMP: multivariate matching pursuit, CQA: cycle quality assessment, SR: segmentation rate, HMM: Hidden Markov model, MFCC: mel-frequency cepstral coefficient, PM: probabilistic model, HE: Homomorphic envelogram, HF: homomorphic filtering, DHMM: duration-dependent hidden Markov model, WD: wavelet decomposition, CSCW: cardiac sound characteristic waveforms, STMHT: short-time modified Hilbert transform, SHM: systolic heart murmurs, ASSA: adaptive singular spectral analysis, RME: Rényi marginal entropy, BMV: bioprosthetic mitral valves, CARDJHU: cardiac auscultatory recording database of Johns Hopkins University, CA: classification accuracy, WHSMD: Washington heart sounds and murmurs database, VIM: Viola integral method, DNN: deep neural network.

Appendix D: Heart Sounds Feature Extraction Approaches for Analysis

Year	References	Analysis/Transform	Feature extraction methods	Number of features	Feature details/Characteristics	Feature selection
2018	[101]	Wavelet analysis	Short-term windowing technique	8 features	Time-domain and frequency-domain features	Adaptive feature selection
2018	[179]	Wavelet analysis	Curve fitting and MFCC	34 features	Wavelet and entropy features	-
2017	[102]	TF analysis	Tensor decomposition method on scaled spectrograms	-	Features from scaled spectrogram	Tensor decomposition
2017	[280]	FFT and wavelet analysis	-	2 types of features	Spectral amplitude and wavelet entropy	-
2017	[45]	TF analysis	Feature extraction from scaled spectrogram	-	Scaled spectrogram features	PLSR
2017	[54]	Frequency and statistical properties of envelope	Statistical properties of averaged shapes for a different frequency band	53 features (total 228 features)	Symmetry of line segments surrounding S_1 and S_2, skewness, kurtosis and centers of gravity, mean, SD and ratios	Probability assessment
2017	[67]	Wavelet analysis (WPD and CWT)	Feature extraction using CQA	90-dimensional features vector	4 set of features from the time, TF and perceptual domain	Fisher's discriminant analysis
2017	[68]	Wavelet analysis (WD)	-	131 features	Time, frequency, wavelet and statistical features	CFS algorithm
2017	[69]	Wavelet analysis	Time/frequency characteristics using 'Gaus4' mother wavelet	220 (CWT), 400 (MFCC), inter-beat (20), 35 (Complexity)	MFCC and complexity measurement (spectral entropy, SD, skewness, and kurtosis)	PCA (50 features selected)
2017	[75]	TFD analysis using acoustic features	*k*-mean algorithm (Euclidean distance)	39 MFCCs (velocity and acceleration) features & 264 acoustic features (Fbank)	HSAD, MFCC and *k*-means	-
2017	[127]	Time, frequency and sparse coding algorithm	Sparse coding for unsupervised feature extraction	20 TD features & sparse coding features	Sparse coefficients, N-points FFT	-
2017	[129]	TF analysis	-	-	MFSC	No
2016	[144]	Wavelet analysis	WT	20 features and some additional wavelets based features	Duration of each cardiac state, mean amplitude and their ratios	-
2016	[91]	Wavelet analysis (DWD)	Diffusion maps for unified feature representation	-	Autocorrelation features	-
2016	[70]	Wavelet analysis (DWT)	WD	-	HE, Hilbert, wavelet and PSD envelopes	Wavelet feature optimization
2016	[103]	T, F, TFD analysis	-	40 features	LPC, entropy, MFCC, PSD and wavelet based features	WFSS (18 features)
2016	[126]	FD analysis	Power spectrum analysis	16 features using PSA	-	-
2016	[153]	FFT	Clustering method	3,500 features	Spectral features	40 features selected using filter method
2015	[46]	Wavelet analysis	STMHT based FFM	10-dimensional FFM	FFM	PCA
2015	[76]	Wavelet analysis	Feature extraction algorithm	36 features from diastole and systole of every cardiac cycle	Dimensional features from segmented cardiac cycles	-
2015	[77]	Wavelet analysis (WPD)	EFSE evaluated from reconstructed selective frequency components of HSs	5 features	EFSE, Sampling frequency dependent features	-
2015	[132]	TQWT	SAMDF based feature set derived with TQWT	-	Wavelet-based features	-
2015	[135]	Wavelet analysis (wavelet packet)	MF-DFA, MESE and EMD	5 features	D/S, amplitude (S1/S2), Multifractal spectrum parameters, f_{PSDmax}, adaptive sub-band energy fraction	-
2015	[35]	EMD for IMF	PSM, IF & amplitude, and power in frequency bands, sample & spectral entropy, simplicity and statistical moments	5029 features within nine classes	Spectral features, signal complexity features	PCA
2014	[47]	Moving windowed Hilbert Transform)	Envelope of HSs using VIM	4 features	Envelope extraction, diagnostic features in TD and FD	-
2014	[36]	EEMD and Kurtosis features	EMD	-	EEMD and Kurtosis features	-
2014	[142]	Ensemble Empirical Mode Decomposition (EEMD)	EMD/EEMD and IMF correlation dimensions	13 IMF components	IMFs and correlation dimensions of the IMF components were used as feature sets	-
2014	[85]	TF analysis	Linear Predictive Coding coefficients for feature extraction	12 different classes	Spectral analysis for features	MCS algorithm
2013	[39]	*S*-transform	SVD of S-matrix	70 features	TFD features	-
2013	[37]	TF and frequency analysis	Short-time frequency amplifier technique	2 frequency and amplitude features	Frequency and amplitude based features	-
2013	[143]	MLBS of wavelet features	WPD	448 nodes for feature extraction	Frequency and TD features	Exclusion criteria for feature reduction

2012	[73]	TF analysis	Murmur likelihood as temporal features	39-dimensional features (best results)	MFCC and HMM states likelihood features	-
2012	[131]	Wavelet analysis	Wavelet packet transformation using Daubiches8	128 entropies	Wavelet entropy-based features	PCA (32 features)
2012	[281]	Wavelet analysis	Time-domain, wavelet features and entropy	32 features	Linear and nonlinear time-domain, wavelet and entropy features	PCA, GDA and GA for feature selection
2012	[157]	DFT	DFT and Burg autoregressive spectrum analysis	300 DFT and 33 Burg-AR features	Frequency domain features	PCA (8 DFT and 6 Burg-AR features)
2012	[158]	Wavelet analysis (DWT)	DWT for feature extraction	6 wavelet entropies	TFD features	Shannon energy
2012	[148]	DFT	-	-	FD features	PCA
2012	[183]	Wavelet analysis (CWT)	SVD and QR decomposition	83 features based on CWT, SVD, QRD	Shannon entropy and the Gini index using WT	SFFS
2011	[81]	TF analysis using FFT	Adaptive singular spectral analysis	-	Correlation & kurtosis features	-
2011	[56]	Wavelet analysis (DWT)	-	32 features	TF based features	PCA
2011	[57]	Wavelet analysis	MP and MMP based feature extraction	-	TF based features	-
2010	[34]	TF domain analysis using STFT	HE for features extraction	4 feature set of single and multi-feature	Frequency & HE features	-
2010	[125]	Wavelet analysis	-	2 diagnostic features	Frequency domain features (f_{max} and f_{width} for NAR-PSD)	-
2010	[136]	Feature extraction from TF representation	Linear decomposition and tiling partition of TF plane	851 features	TF based features	Linear grid, Quadtree, PCA, PLS, 2D-PCA
2010	[108]	TF analysis	Wavelet-based features	-	Clinical features (normal split sound duration, frequency content)	-
2010	[150]	Spectrogram, WVD, SPWVD, CWD, ETD, HTD and scalogram	TF representation	-	MFCC, energy, frequency, BW, Eigen vectors, and spectral centroid	-
2009	[106]	TF analysis	Hierarchical clustering	3500 features approx.	Feature-space of cluster distances, correlation and Euclidean distance	Clustering for feature reduction
2009	[146]	TF domain analysis using wavelet	WD	100 scalar features	4 (SD and HR), 8 (S_1 and S_2 each), 24 (systolic), 48 (diastolic) & 8 (energy) scalar features	-
2009	[58]	STFT, Gabor Transform, WVD, Wavelet transform	-	149 features	T varying & TF, perceptual and fractal features (Eigenspace, MFCC, LLE, Hurst Exponent, Correlation Dimension)	PCA
2009	[282]	Wavelet analysis	Rectangular window and power content in the window	50-dimensional feature vectors	Power of the detailed coefficients in each segment with and without filtering	Divergence analysis
2009	[145]	Wavelet analysis	-	12 features	12 wavelet entropies	-
2009	[159]	Wavelet analysis	db4 decomposition filter within five resolution levels	-	Diagnostic features	-
2008	[19]	TD analysis	-	-	Clinical features (duration of split-sounds, systole, and diastole, frequency)	-
2008	[151]	Wavelet analysis and STFT	STFT and WD	91 features	Wavelet entropy	-
2008	[134]	Wavelet analysis (WPD)	-	2 features	Wavelet packet energy (mean and SD)	-
2008	[83]	Wavelet analysis	Wavelet transform for feature extraction	50 features for each record	Wavelet features	Divergence analysis
2008	[117]	TD and STFT	-	3 classes of features	TD, STFT and MFCC features	-
2007	[74]	Wavelet analysis	Daubechies-2 wavelet coefficient decomposition	32 wavelet features	Wavelet features	HF and *k*-means clustering
2007	[84]	Wavelet analysis	-	64 features	Wavelet features	-
2007	[149]	Wavelet analysis and STFT	WD, STFT, wavelet entropy	-	Daubechies-10 WD, STFT & wavelet entropy	-
2007	[147]	Wavelet analysis and STFT	WD, STFT and wavelet entropy	91 features	-	-
2006	[52]	Spectral analysis	SPWVD	-	FFT (spectrogram)	-
2006	[59]	FFT, DWT, Wavelet analysis	RQA	207 features	Time, TFD, nonlinear and chos based features and HOS, state space, fractal dimension, bispectrum, wavelet entropy, fractals, Gaussian mixture model, Eigenvalues	Pudil's SFFS method
2004	[15]	Wavelet analysis (WD)	-	256 elements feature vector	TF features	-
2003	[133]	WPD and NN	TFD adaptive feature extraction, WPNN	256-Wavelet packet entropy per DHSs signal	Wavelet packet entropy	-
2003	[165]	TF analysis using wavelets	WD and wavelet entropy	12 wavelet entropy values	-	Adaptive feature extraction
2003	[166]	TF analysis using wavelets	-	336 feature vectors for the training set and 336 feature vectors for the test set	Wavelet-based features	Divergence analysis (16 features)
2002	[167]	TF analysis using FFT	FFT and the Levinson-Durbin auto-regression	-	Spectral estimation	Auto-regression
2001	[61]	TF analysis	Matching pursuit	2 features	Dominant frequency-based features	-
1995	[168]	TF using wavelet	-	2 features	Average correlations and Euclidean distance	-
1987	[63]	TD and FD analysis	Energy spectrum analysis	4 EDC	Energy curve, spectrum & distribution coefficients	-
1980	[64]	Frequency domain spectral tracking	Analysis using linear-prediction method	-	Spectral level tracking by evaluating spectral density function	-

*Note: TF: time-frequency, TD: time-domain, FD: frequency domain, TFD: time-frequency domain, PLSR: partial least squares regression, CFS: correlation-based feature selection, LPC: linear predictive coefficient, WFSS: wrapper feature selection scheme, SVD: singular value decomposition, EFSE: energy fraction and sample entropy, SAMDF: sum of average magnitude difference function, MF-DFA: multifractal detrended fluctuation analysis, MESE: maximum entropy spectra estimation, DWD: discrete wavelet decomposition, TQWT: tunable-*Q* wavelet transform, PSM: parametric spectral modelling, MCS: modified Cuckoo search, MLBS: multi-level bias selection, QRD: *QR*-decomposition, SFFS: sequential forward floating selection, MP: matching pursuit, MMP: multi-variate matching pursuit, HE: homomorphic envelogram, WVD: Wigner-Ville distribution, SPWVD: smoothed pseudo WVD, ETD: exponential *T*-distribution, HTD: hyperbolic *T*-distribution, RQA: recurrence quantification analysis, WPD: wavelet packet decomposition, SD: standard deviation, HT: Hilbert transform, EDC: energy distribution coefficients, LLE: largest Lyapunov exponent, FFM: frequency feature matrix, EEMD: ensemble Empirical mode decomposition, IMF: intrinsic mode functions, DHS: Doppler heart sounds, MFSC: mel-frequency spectral coefficients, WD: wavelet decomposition, WPNN: wavelet packet neural network, NAR-PSD: normalized auto-regressive power spectral density, CWD: Choi–Williams distribution.

Appendix E: Summary of Heart Sounds Classifiers and their Performance Comparison

Year	References	Method	Transform	Subject type	Records (# or duration in seconds)	Sampling rate	Type of features/size of feature vector	Classifiers	Performance (%)	Type of cardiac abnormality investigated
SVM based classifier										
2017	[102]	Scaled spectrogram and tensor decomposition	TFD	Datasets A & B from [23]		44.1 kHz	Tensor decomposed features	SVM, SS-PLSR, SVM-DM, SS-TD	Normalized precision: 76.0 (Dataset-A) & 74.0 (Dataset-B) using SS-TD	Normal, murmurs, extra heart sounds, extrasystole and artifacts
				Dataset C [8]		4 kHz				
2017	[45]	HSs classification based on scaled spectrogram and PLSR	TFD	Datasets A & B (176 & 507) records available from [23]		44.1 & 4 kHz	Scaled spectrogram based TF features	SVM	S_e: 100 (best with artifacts), S_p: 64.0 (artifacts)	Normal, murmurs, extra heart sounds, extrasystole and artifacts
2017	[127]	Sparse coding features with TD features using SVM	T, F and TF domain	PhysioNet database [8]		Varied	Sparse coefficient matrix and time-domain features, 20 features	SVM	S_p: 88.45, S_e: 90.07, MAcc: 0.892 (modified results)	S_1, systolic, S_2 and diastolic sounds
2016	[91]	Without segmentation, classification using autocorrelation feature and diffusion maps	TFD	Datasets A & B (176 & 507) records available from [23]		44.1 & 4 kHz	TF based autocorrelation features	SVM-DM, SVM-A, and SVM-AD	S_e: 100.0, S_p: 64.0 (using SVM-AD classifier for artifacts)	Normal, murmurs, extra heart sounds, extrasystole and artifacts
2015	[135]	Cardiac reserve and HSs characteristics analysis	WT	88 (N) and 64 (CHF)	88 (N) and 64 (CHF), 1 minute each	11025 Hz	Wavelet features based 5 features	LS-SVM and compared with BP-ANN and HMM	S_p: 93.75, S_e: 96.59, Acc: 95.39 (using LS-SVM)	Chronic heart failure, subjects without murmurs
2015	[46]	STMHT-based FFM for classification	WT	14 (AF)	188 periods	44.1 kHz	Dimensional FFM, 10 features	SVM	S_p 90.6, S_e 96.9, Acc 91.7	AF, AR, MR, PS, VSD and normal sounds were classified
				17 (AR)	181 periods				S_p 98.9, S_e 98.5, Acc 98.8	
				25 (MR)	257 periods				S_p 98.5, S_e 98.1, Acc 98.4	
				25 (N)	325 periods				S_p 100, S_e 99.6, Acc 99.8	
				7 (PS)	150 periods				S_p 98.6, S_e 99.9, Acc 98.7	
2015	[138]	Structural complexity based feature extraction	TFD	60 (N) and 60 (pathological)	-	8 kHz	Sample entropy	SVM	Clean data: S_p: 100, S_e: 93.33, Acc: 96.67 & 91.66 (with 10 dB SNR)	Classification between normal and abnormal heart sounds
2015	[77]	Energy fraction and sample entropy	WT	40 healthy and 67 CVD patients	80 (N) and 167 (SHM)	2205 Hz	Energy and entropy	SVM with LKF, PKF, GRKF and SKF, 5 features	S_p: 98.55, S_e: 93.48, Acc: 97.17 (using GRKF)	Normal, aortic/pulmonary stenosis and tricuspid/mitral insufficiency
2015	[132]	Least-square support vector machine (LS-SVM)	WT	163 HSs signals (N, septal, valvular and other defects)	4628 cycles (626 N and 4002 pathological)	44.1 kHz	SAMDF derived with TQWT, 21 features	SVM with different kernel functions	S_p: 99.29, S_e: 98.80, Acc: 98.92 (using Q = 6 and MWKF)	Septal, valvular and other mechanical defects
2015	[139]	GTSVM for murmur classification	WT	14 (45-93 y) PM, 16 (1-16 y) PM, 26 (2-14 y) IM, 30 (4-15 y) NM	10s duration each	44.1 kHz	TF features	GTSVM classifier	S_p: 89.30, S_e: 86.30, CR: 88.10	Innocent and pathological murmurs
2015	[140]	Intelligent PCG system	TFD	30 (N) and 26 (IM) and 30 (AS)	-	-	Frequency features	SVM	S_p: 89.3, S_e: 86.4	Aortic stenosis severity assessment
2015	[141]	HOC of wavelet packet coefficients for HSs classification	WPD	16 (N), 19 (MR), 14 (AS) & 10 (AR)	59 records, 820 cycles, each 15s	4 kHz	Wavelet based 46 features	SVM	Acc: 99.39 (using CT_LDB method)	Normal, MR, AS, and AR heart sounds
2014	[47]	Features from envelope of HSs using VIM	WT	VSD, normal HSs, AR, AF, AS and MS	242 (N) and 226 (VSD)	44.1 kHz	Time and frequency domain features, 4 features	SVM	S_p: 98.1, S_e: 98.8, Acc: 98.4 using boundary curves & S_p: 98.4, S_e: 98.6, Acc: 98.5 using ellipse model (for VSD detection)	VSD, AR, AF, AS, MS and normal heart sounds
2014	[142]	HSs classification and recognition based on EEMD	WT	225 (N) and 180 (60 (MS), 60 (VSD) & 60 (AS))	405 HSs, testing with 75 (N) and 20 (MS), 20 (VSD) & 20 (AS)	11025 Hz	EEMD and IMF correlation dimensions, 13 features	Binary tree SVM (BT-SVM)	Acc: 98.67 (normal) and 91.67 (abnormal)	MS, VSD, AS and normal heart sounds

2014	[85]	HSs classification using SVM-MCS	TFD	3M Poland microphone samples from [279], 72 records		-	12 features using LPCC	SVM and MCS algorithm	Acc: 95.43 (average using SVM-MCS)	Ejection click, split of S_1 & S_2, FHSs, S_3, S_4, pan-systolic, late systolic, early systolic murmurs, opening snap and diastolic rumble
2014	[100]	Feature extraction based on OMS-WPD	WT	50 (healthy) and 68 (pathological)	Each 5s duration	22050 Hz	Features using OMS-WPD	SVM	S_p: 94.00, S_e: 85.29, CR: 88.98 (using db8)	Normal, presystolic gallop, early diastolic, enhanced S_2, aortic systolic ejection and mid-systolic rattling sounds
2013	[143]	MLBS of wavelet features	WT	59 (Normal and pathological)	16 (N), 19 (MR), 14 (AS), 10 (AR)	4 kHz	Wavelet based features	SVM with MLBS	S_p: 100.0, S_e: 98.0, Acc: 97.56	Normal, AS, MR and AR sounds
2012	[73]	Classification using murmur likelihood and HMM state likelihood	TFD	Normal and pathological	80 (N) and 80 (abnormal) HSs signals	8 kHz	MFCC and HMM states likelihood, dimensional, 39 features	SVM classifier	Acc: 80.6 (6-HMM state and 3-Gaussian mixtures) and 85.6 (SVM classifier)	Various murmurs and normal heart sounds
2010	[108]	LMS based LS-SVM	TFD	64 (Normal and pathological)	64 recordings (512 cycles)	8 kHz	Wavelet features	LS-SVM	Acc: 92.889 (average of Sets 1-16)	Normal and five different pathological sounds
2010	[125]	NAR-PSD and multi-SVM	WT	6 healthy and 34 pathological	196 (N) and 293 (AN)	8 kHz	f_{max} and f_{width} as features from NAR-PSD	Multi-SVM	S_p: 99.9, S_e: 99.5, Acc: 99.6 (best case)	Normal, AF, AS, MR, MS, aortic insufficiency and split sounds
2010	[144]	Machine learning based identification	WT	PhysioNet database [8]	4,430 recordings	Varied	Wavelet based 20 features	Bagging trees, boosted trees, logistic classifier and SVM	Validation Acc: 94, 93, 85 and 91, respectively S_e: 79.58 and S_p: 74.59, overall: 77.08	Heart sounds were classified as normal or pathological sounds
2010	[283]	CWT and SVM based detection of the paradoxical splitting of S_2	CWT	13 (LBBB and paradoxical splitting), 18 (AS and paradoxical splitting), and 11 (N)	Total 42 samples used from 250 records	-	Time-frequency maps, total 5 features	SVM	S_e: 94.44, S_P: 87.5, and Acc: 90.97	Detection of the paradoxical splitting of S_2
2009	[145]	Genetic-SVM (GSVM)	DWT	132 M/83 F (Age: 15-80 y)	215 samples (5s each) (Doppler heart sounds signals)	20 kHz	Wavelet entropies based 12 features	SVM tested with 8 different kernel functions	Acc: 96 (N) and 94.52 (AN) using GSVM Model-4 (ERBF) (best results)	Heart valve diseases
2009	[146]	Identification of valvular diseases using SVM	WT	Healthy, AS, AR, MS and MR, (age: 18-22 y)	38 (N), 41 (AS), 43 (MR), 38 (AR), 38 (MS)	-	TFD and wavelet features, total 100 features	SVM (Gaussian radial basis function)	S_p: 94.74, S_e: 87.50, Acc: 91.43 (b/w N and AN) using GRBF	Normal, AS, AR, MS and MR heart sounds signals
2008	[134]	Wavelet packet energy features using WPD and SVM	WT	30 (N) and 52 (VHD) (for testing)	-	8 kHz	Wavelet packet energy based features	SVM	S_p: 96.67, S_e: 100.0	Aortic and mitral valvular heart disorders
2007	[147]	LS-SVM and BP-ANN	WT and STFT	Normal and abnormal (DHSs signals) for 5 seconds each, 132M/83F	215 samples (54 (NAHV), 56 (ANAHV), 66 (ANMHV), 39 (NMHV)	20 kHz	WD, STFT and entropy based 91 features	LS-SVM and BP-ANN	S_p: 94.0, S_e: 95.9 (for BP-ANN) & S_p: 90.0, S_e: 94.5 (for LS-SVM)	Cardiac abnormalities related to aortic and mitral valves (AR, AS, MS and MR) and normal heart sounds
Neural Network based classifier										
2018	[152]	Gram polynomials and PNN	FFT	Normal and pathological sounds from PhysioNet database [8], 3126 records		2 kHz	Gram polynomial and FFT, 64 features	PNN	S_p: 91.0, S_e: 93.0, Acc: 94.0	No clinical abnormality studied
2018	[89]	ANN and SVM classifier	STFT	87 innocent murmurs and 170 pathological murmurs	257 records	8 kHz	Temporal and spectral features, 14 features	ANN and SVM	S_p: 91.0-99.0, S_e: 84.0-93.0	Identification of innocent murmurs in children
2017	[67]	Feature extraction using CQA	TD and TFD	Normal and pathological sounds from PhysioNet database [8], 1277 records (308 patients)		2 kHz	TD, TFD and perceptual features, 90 features	FFNN, dimensional features	S_p: 87.14, 87.21 & 88.51 and S_e: 75.44, 91.38, 88.83, for data set1, set2 and set3, respectively using CQA	Classification as normal and abnormal heart sounds
2017	[69]	Drop-Connected neural network	TFD	Normal and abnormal HSs from PhysioNet database [8]		-	MFCC, inter-beat and complexity features, 675 features	Two-hidden layer NN trained by EBP	Acc: 85.2 (on test data)	Normal, MVP, aortic disease, CAD, MR, AS and other miscellaneous pathological heart sounds
2017	[75]	Identification of S_1 and S_2 using DNN	TFD	28 subjects (17M/11F)	460 each S_1 and S_2 peaks in HSs	48 kHz	MFCC and acoustic features, 303 features	DNN (KNN, LR, SVM and GMM for comparison)	Acc: 85.0 (using 39 dimensions with K-means) and 91.12 of accuracy overall	First and second heart sound identification
2017	[129]	MFSC based deep CNN classifier	TFD	Normal and abnormal HSs from PhysioNet database [8]		-	MFSC features	Deep CNN	S_p: 87.66, S_e: 80.63, Overall: 84.15 (CinC 2016)	Normal and abnormal heart sounds classification
2016	[153]	2-means clustering and ANN	FFT	Database from [8], [22]	304 records	-	Time-frequency spectral	ANN	S_e: 84.4, S_p: 86.9, Acc: 86.5	Classification as normal and abnormal heart sounds

							features, 40 features			
2016	[126]	Power spectrum analysis	FFT	Normal and abnormal HSs from PhysioNet database [8]		-	16 Frequency features	NN	S_p: 78.80, S_e: 74.70, Acc: 76.7	Normal and abnormal heart sounds classification
2016	[103]	Ensemble of NN without segmentation	T, F and TFD	Normal and abnormal HSs from PhysioNet database [8]		-	T, F and TF based 18 features	FFNN	S_p: 94.23, S_e: 88.76, Overall Acc: 91.5	Classification for anomaly and quality detection
2016	[154]	Ensemble of feature and deep learning based classifier	TD and FD	PhysioNet database [8], 2575 (N) & 665 (pathological) records		1 kHz (resampled)	124 time-frequency features	AdaBoost and CNN	S_p: 77.81, S_e: 94.24, Overall: 86.02	Classification as normal and abnormal heart sounds
2016	[155]	Deep structured features for classification	WT	PhysioNet database [8], 764 subjects	3153 records	-	Wavelet based 20 features	CNN	S_p: 77.6, S_e: 84.8, Score: 81.2	Normal and abnormal heart sounds classification
2014	[156]	Spectral analysis with a time growing window	WT	Normal and SEC from 40 children (age 3-9 years)	614 normal and abnormal cardiac cycles	44.1 kHz	Spectral analysis with a time growing window	TGNN, TDNN and MLP	S_e: 98.1, Acc: 97.0 (TGNN), S_e: 76.4, Acc: 85.1 (TDNN), S_e: 85.7, Acc: 92.7 (MLP)	Systolic ejection clicks in children
2012	[157]	PCA and ANN for diagnosis of heart valve disease	TFD, DFT and Burg AR	40 (N), 40 (PS) and 40 (MS)	-	8 kHz	Time-frequency based 14 features	DFT/Burg AR-PCA-ANN	S_p: 97.44, S_e: 90.48, Acc: 95.0	Heart valve diseases
2012	[158]	ANFIS for identifying heart valve disease	DWT	40 (N), 40 (PS) and 40 (MS)	-	8 kHz	Wavelet entropies based 6 features	ANFIS (ANN)	S_p: 95.24, S_e: 100.0, Acc: 98.33	Normal, pulmonary, and mitral stenosis heart valve diseases
2011	[56]	HSs analysis without segmentation using DWT, PCA and NN	WT	Various murmurs	57 HSs signals	4 kHz	Time-frequency based 32 features	PCA and NN	Acc: 92.0 (noise free), 90.0 (≈10 dB SNR)	Normal, S_3, S_4, ejection sound, AR, AS, MR, MS, PS, split S_2, systolic click and opening snap
2011	[57]	MMP based on three-layer FF-MLP network	TFD	70 patients (N and pathological murmurs)	70 records (35 normal)	44.1 kHz	Time-frequency features	FF-MLP	Acc: 92.5 & 77.5 using MMP and MP, respectively	VSD, ASD, PS, MR, and innocent murmurs
2008	[83]	SOM network and ISOM analysis	WT	Data from two patients, online and CD book	14 records, each record with 20 cycles of HS	2 kHz	Wavelet features, 50 features for each	NN	Acc: 95.0% (using ISOM) and 70% (using Kohonen network)	VSD, MR, LSM, early systolic, opening snap, diastolic rumble, AR, AS, MS, Ebsteins anomaly, summation gallop, venus hum, normal FCG and aortic insufficiency
2009	[159]	WT and NN for classification	DWT	-	102 (N), 96 (AI), 92 (AS) & 82 (PS) cycles	11.025 kHz	Wavelet features	ANN (MLP-BP)	Acc: 94.42	Aortic insufficiency, AS, PS and normal sounds
2008	[160]	Arash-Band frequency features with ANN	WT	90 children (36 N and 54 pathological)	40 training & 50 test signals (each 10s)	44.1 kHz	Arash frequency bands based 5 features	ANN	CA: 94.00	Congenital heart diseases in children
2007	[74]	Homomorphic segmented HSs	WT	Mix of N (32%), SM (36%) and DM (32%)	41 records (340 cycles)	8 kHz	Wavelet based 32 features	GAL and MLP-BP NN	Acc: 98.50 (using GAL with dataset2)	Normal, systolic and diastolic murmurs
2007	[84]	Segmentation and feature extraction using wavelets	WT	Database from [13]		8012 Hz	Wavelet based 64 features	NN	S_p: 85.00, S_e: 54.00, Acc: 70.0	Normal and aortic regurgitation heart sounds
2007	[161]	Wavelet analysis for automated auscultation	TFD	113 normal and 50 pathological	Each 6 cycles	-	Time-frequency	ANN	S_p: 96.5, S_e: 90.0	Normal and pathological heart sound classification
2006	[52]	Spectral analysis using SPWVD	FFT	40 patients	8s each	8 kHz	Time-frequency features	MLP-NN and SPWVD	S_p: 86.40, S_e: 85.10, Acc: 86.4	FHSs and murmurs
2006	[59]	RQA feature and ANN classifier	TD, TFD and WT	36 patients	12 cycles each of 15s	44.1 kHz	207 complexity features	NN	Acc: 86 (MI, AS and PM)	Systolic heart murmur classification
2006	[162]	A three-layered ANN for twelve types of cardiac abnormalities detection	TD and FD	49 patients with murmurs	Doppler ECG and PCG signals for 8 sites	-	-	NN	Acc: 68% (overall)	VSD (AS, AR) and PDA not correctly diagnosed
2005	[163]	Wavelet analysis and ANN	WT	Database from [26]		8 and 16 kHz	PCA	ANN	S_p: 70.5, S_e: 64.7 Acc: 70.2	Normal and abnormal heart sounds classification
2005	[164]	Cardiac auscultation in pediatrics	FFT	88 (IM) and 153 pathological murmurs	Each 10-15s	44.1 kHz	Frequency, using Fisher's method	ANN	S_p: 90.00, S_e: 83.00	Innocent and pathological murmur (such as VSD) classification
2004	[15]	WD and classification using NN based classifier	WT	Normal, MVP, coarctation of the aorta, VSD, PS	4096 sample segments	8 kHz	WD based 256 features	NN	Acc: 100.0 (for signals with SNR above 31 dB)	Normal and abnormal sound from coarctation of the aorta and split sounds
2003	[165]	MLP with FFNN	TFD	123 subjects	AV (62 N & 80 AN) & MV (38 N & 66 AN), DHSs	-	Wavelet entropies based 12 features	FFNN	Acc: 84 (N) and 96 (AN)	Aortic and mitral valve diseases
2003	[133]	WPD and wavelet packet entropy	WT	95 (N) and 120 (AN)	215 samples (Doppler heart sounds)	20 kHz	Wavelet packet energy	WPNN	Acc: 94.0 (N) and 94.5 (AN)	Normal and abnormal sounds, no murmurs

							based 256 features			
2003	[166]	Classification of wavelet-based features using ANN	WT	28 subjects, 28 records with 12 periods of HSs in each	4096 discrete data	5512.5 Hz	Wavelet based 672 features	GAL network and LVQ network	Acc: 99.0	AS, MR, MS, PS, AR, SG and normal sounds
2002	[167]	MLP and RBF-NN	FFT	-	36 recordings (each ~ 30 s)	44 kHz	FFT and Levinson-Durbin auto regression	MLP and RBF-NN	Acc: 84.00 (MLP) & 88.00 (RBF)	Normal, NRMR, MVP, BAV, AS, SCAS, AR, aortic/mitral valve, MS, ASD, VSD, LVI, PMD, DC, and aortic incompetence
1995	[168]	Backpropagation based NN classifier (along with ECG)	TD	Database from [30], 48 records (18 (N) and 30 (AN)) from 3 & 6 subjects, respectively		2 kHz	Average correlations & Euclidean distance features	NN	Acc: 95.0	MS, split of S_1 and S_2, aortic/mitral insufficiency, MS, and mitral insufficiency with prolapse
1994	[169]	Wavelet-based fuzzy-NN	FWT	112 patients, 30 selected (15 (N) and 15 (AN)	10 samples from each patient	4 kHz	Wavelet features (mean, variance, skewness, kurtosis)	Fuzzy-NN	Acc: 88.8 (N) & 85.45 (AN)	Coronary artery diseases
HMM-based classifier										
2018	[101]	ANFIS and HMM	WT	Normal and pathological	150 (N) and 80 (AN) records	-	TD and FD based 8 features	ANFIS and HMM	Acc: 98.7	Normal sounds and valve disorders (MS, MR and AR)
2012	[148]	PCA-Discrete HMM	DFT	40 (N), 40 (PS) and 40 (MS) (55M/65F, Age 4-65 y)	80 training set and 187 test set	8 kHz	TD, FD	PCA-DHMM	S_p: 93.30, S_e: 70.30, Acc: 72.2 (SPECTF data)	Classification of normal, MS and PS cardiac sounds
2008	[117]	MFCC-based HMM	TD and STFT	20 (N), 6 (CM), 4 (DM), 11 (SM)	1381 signals	8 kHz	TD, STFT & MFCC features	HMM	Acc: 95.7 (CM), 96.25 (SM), 90.0 (DM)	Normal sounds, ejection clicks, opening snaps, split S_1, split S_2, S_3, S_4, continuous, diastolic and systolic murmurs
2007	[116]	Modified HMM (MHMM)	TD, FD and STFT	41 subjects (21 N and 21 CM, SM and DM)	1398 records	8 kHz	TD, STFT and MFCC	MHMM	S_p: 95.30, S_e: 95.20	Classification of normal sounds, continuous, diastolic and systolic murmurs
2007	[149]	Continuous hidden Markov model (CHMM) based classifier	WT and STFT	132M/83F, mean age 48 years	215 samples (54 (NAHV), 56 (ANAHV), 66 (ANMHV), 39 (NMHV), Doppler ultrasounds	20 kHz	Wavelet and time-frequency domain features	CHMM (FCM/*k*-means algorithms) & ANN	S_p: 92.0, S_e: 97.26 (FCM-*k*-means/CHMM) & S_p: 94.0, S_e: 95.89 (ANN)	Classification of normal and abnormal sounds from mitral and aortic valves
Nearest Neighbor Classifier										
2018	[179]	Nearest neighbor (NN) classifier with Euclidean distance	WT	Six different datasets including PhysioNet database [8]	4000 samples (20 s each)	44 kHz and 2 kHz	Wavelet and filter bank, 34 features	NN with Euclidean distance	S_p: 99.00, S_e: 93.00, Acc: 98, (for dataset E)	Classification of normal and abnormal heart sounds including murmurs
2013	[39]	Feature extraction and Shannon energy using S-transform	T, F and TFD using S-transform	40 (N) and 40 (pathological)	80 records	8 kHz	Best results with the TFD features, total 70 features	*k*-NN classifier	S_e: 95.0, S_p: 97.0 (using TFD features)	Classification of S_1 and S_2
2010	[150]	Dynamic features based on various energy distributions	STFT and CWT	22 adults (16 (N) and 6 (SM))	22 PCGs (each ≈ 12 s)	44.1 kHz	TD, FD, TFD based 53 features	*k*-NN	Acc: 98.00	Murmurs detection
2010	[136]	Feature extraction from TFR	TFD	45 adults (26 N and 19 pathological)	45 PCGs (each ≈ 12 s)	44.1 kHz	TVAR, 851 features	*k*-NN	Acc: 99.06±0.06 (best case)	Murmur detection
2009	[58]	*k*-NN with fractal features	STFT, Gabor Transform, WVD, WT	81 (N) and 83 with murmurs	164 records	44.1 kHz	T varying & TF, perceptual and fractal, 149 features	*k*-NN	Acc: 97.17 (using fractal features)	Murmur detection
2008	[151]	Hybrid classifier with AIS and *k*-NN	WT and STFT	132M/83F (Age: 15-80 y)	AV: 110 (54 N & 56 AN), MV: 105 (66 N & 39 AN)	-	WD, STFT and Wavelet entropy based 91 features	AIS and fuzzy *k*-NN	S_p: 96.00, S_e: 95.90	Classification of normal and abnormal sounds from mitral and aortic valves
1998	[107]	CWD based spectrum analysis	TFD and DWT	45 native and 23 aortic Carpentier-Edwards valve	-	2 kHz	Morphological and DWT based 9 features	*k*-NN	Acc: between 61 to 96 (based on the feature selection)	Classification of native and bioprosthetic heart valve sounds
Other classifiers/Hybrid classifiers										
2017	[95]	Empirical wavelet transform (EWT)	EWT	Databases from [23], [8], [25], [28], [24], [29] and HSs, Littmann HSs, Washington HSs and murmurs (WHSM) (total 278 records)			Timing, area and interval based features	Decision rule based classifier	S_e: 97.9, PPV: 97.7, Acc: 95.7 (Noise free)	S_1, S_2, S_3, S_4, split sounds, systolic murmurs (early, mid, late, pan) and diastolic murmurs (early, mid, late, pan) & continuous murmurs

2017	[54]	Probability analysis for feature extraction	TFD	Normal and abnormal HSs from PhysioNet database [8]		2 kHz	Statistical properties of envelope, 53 features	Extension to Naïve Bayes classifier	S_p: 91.25, S_e: 76.96, Acc: 84.11	Classification of normal and abnormal heart sounds including murmurs
2017	[68]	Ensemble of classifiers	TFD	Normal and abnormal HSs from PhysioNet database [8]		1 kHz (resampling)	Time, frequency, wavelet and statistical domain, 131 features	Ensembles of 20 two-step classifiers	S_p: 80.6, S_e: 79.6 (hidden test set), overall: 96.30/90.18 (standard/outlier signals)	Classification of normal and abnormal heart sounds including murmurs
2017	[280]	Wavelet entropy and spectral amplitude based classifier	WT	2408 (N) and 630 (AN) records from PhysioNet database [8] [22]		-	Spectral and wavelet features	DT based classifier	Acc: 76% (mean of S_e 98% and S_p 54%) for wavelet entropy	Classification of normal and abnormal heart sounds including murmurs
2015	[35]	Acoustic features for CAD detection	TFD	133 subjects for 435 records	231 (Non-CAD) and 204 (CAD)	4 kHz	Spectral features, signal complexity features, total 5029 features	Multivariate classifier	S_p: 65.20, S_e: 72.0 (for CAD detection)	Identification of coronary artery (CAD) diseases
2015	[76]	Detection of heart murmurs (PI & TI)	CWT	-	92 cardiac cycles (38 (PI) and 54 (TI))	8 kHz	36 dimensional features	RWNN classifier with EKF algorithm	Acc: 98.84±4.49 (ELM) and 98.04±045 (RWNN)	Pulmonary insufficiency and tricuspid insufficiency murmur detection
2015	[137]	Coiflet wavelets based features and its selection using BPSO	WT	150 (N), 75 (MVP), 50 (VSD), 50 (PS)	-	-	Coiflet wavelet features	Naïve Bayes, *k*-NN, C4.5 and SVM classifier	Acc: 92.31% (highest with SVM)	Classification of normal, MVP, VSD, and PS heart sounds
2012	[131]	Entropy of the wavelet packets as a classification feature	WT	50 (N), 80 (MR), 100 (AS), 50(AR), 70 (MS)	350 records	4 kHz	Wavelet entropy, 32 features	DT, *k*-NN, BayesNet, MLP and SVM	95.45 (DT), 95.78 (*k*-NN), 96.94 (BayesNet), 95.53 (MLP) & 95.33 (SVM)	Classification of AR, MR, AS and MS sounds
2012	[281]	Feature selection using PCA, GA, GP and GDA	WT	-	120 cardiac cycles of AS, MS, MR	44.1 kHz	TD, wavelet and entropy features, 32 features	MLP, RBF and SVM classifier	Acc: 99.47 (best with RBF using GA feature selection)	Diagnosis of AS, MS and MR heart valve diseases
2012	[183]	Matrix decomposition	CWT	-	15 IM, 28 organic murmurs (380 segments)	-	CWT, SVD and QRD features, 83 features	CART	S_p: 83.00, S_e: 94.00, Acc: 90.00	Murmur classification
2009	[282]	Divergence analysis	WT	Data from 2 patients, online and a CD book	140 HSs periods of 14 different types	2 kHz	50 wavelet features	MLP	Acc: 99% & 95% (based on feature extraction)	VSD, MR, LSM, early systolic, opening snap, diastolic rumble, AR, AS, MS, Ebsteins anomaly, summation gallop, venus hum, normal FCG and aortic insufficiency
2009	[106]	Hierarchical clustering approach	TD, FD and TFD using STFT	12 subjects	10 records of 40s each	20-250 Hz (band pass)	TD, FD, TFD, WVD and CVD based approx. 3500 features	*k*-NN and DA	Acc: 82±7% (k-NN)	Only classification of S_1 sounds
				11 subjects	30-45 minutes				Acc: 86±7% (DA)	
2007	[274]	Phono-spectrographic analysis	STFT	807 pediatric patients	88 (WM), 447 (IM), 272 (PM)	8-44.1 kHz	Phono-spectrographic features	-	S_p: 91.00, S_e: 90.00	classification of innocent and pathological murmurs in children
2004	[171]	Time-frequency based decision tree	TFD	-	84 (AS: 41 and MR: 43)	-	Time-frequency based 100 features	DT	Acc: 90 (overall), 91.6 (AS), 88.5 (MR)	Classification of FHSs, AS and MR sounds
2013	[210]	Multifractal analysis	-	49 healthy and 48 children with PMV	97 PCG records (each 8s)	8 kHz	Multifractals features	Discrimination threshold	Acc: 96.91	Normal and MVP heart sounds
1987	[63]	Energy spectrum	FFT (TD and FD)	5 healthy, 20 patients	47 records	1024 Hz	Energy curve, spectrum & distribution coefficients	No definite classifier	-	Systolic and diastolic murmur classification

*Note: AN: abnormal heart sounds, Acc: accuracy, TD: time-domain, FD: frequency-domain, TFD: time-frequency domain: TFD, SS-PLSR: scaled spectrogram and partial least squares regression, SVM-DM: SVM-diffusion maps, SS-TD: scaled spectrogram and tensor decomposition, CHF: chronic heart failure, SHM: systolic heart murmurs, HOC: higher-order cumulants, LPCC: linear predictive coding coefficients, VSD: ventricular septal defects, SVM-MCS: SVM-modified Cuckoo search, AHV: aortic heart valve, NAHV: normal AHV, ANAHV: abnormal AHV, MHV: mitral heart valve, NMHV: normal MHV, ANMHV: abnormal MHV, EBP: error back-propagation, AIS: artificial immune system, SM: systolic murmurs, DM: diastolic murmurs, CM: continuous murmurs, MLP-BP: multi-layer perceptron back-propagation, CART: classification and regression trees, BP-ANN: back-propagation artificial neural network, IM: innocent murmurs, PI: pulmonary insufficiency, TI: tricuspid insufficiency, MHMM: modified HMM, RBF: radial basis function, RBF-NN: radial basis function neural network, TVAR: time-varying auto-regression, CR: classification rate, PM: pathological murmurs, NM: no murmurs, IM: innocent murmurs, VHD: valvular heart disease, DA: discriminant analysis, WPD: wavelet packet decomposition, TGNN: time-growing neural network, OMS-WPD: optimum multi-scale wavelet packet decomposition, AWN: additive white noise, HSA: heart sounds analysis, NPV: negative predictive value, WM: without murmurs, CT_LDB: cumulant-based trapezoidal local discriminant basis, FCM: fuzzy *C*-means, RWNN: radial wavelet neural network, EKF: extended Kalman filter, PNN: probabilistic neural network, ANFIS: adaptive-neuro fuzzy inference system, FWT: fast wavelet transform, FFM: frequency feature matrix, GRKF: Gaussian radial basis kernel function, LKF: linear kernel function, PKF: polynomial kernel function, SKF: sigmoid kernel function, GTSVM: growing time support vector machine, VIM: Viola integral method, EEMD: ensemble Empirical mode decomposition, MLBS: multi-level bias selection, MFCC: mel-frequency cepstral coefficient, GSVM: genetic-SVM, CQA: cycle quality assessment, DNN: deep neural network, MFSC: mel-frequency spectral coefficients, SEC: systolic ejection click, ISOM: incremental self-organizing map, WVD: Wigner-Ville distribution, SPWVD: smoothed pseudo WVD, RQA: recurrence quantification analysis, WD: wavelet decomposition, FFNN: feed-forward neural network, WPNN: wavelet packet neural network, CAD: coronary artery disease, CWD: Choi–Williams distribution, ELM: extreme learning machine, PMV: prolapsed mitral valve, BPSO: binary particle swarm optimization, SVD: singular value decomposition, DT: decision trees, TRF: time-frequency representation, ERBF: exponential radial basis function, MWKF: Morlet wavelet

kernel function, SG: summation gallop, PDA: patent ductus arteriosus, BAV: bicuspid aortic valve, LVI: left ventricular impairment, PMD: papillary muscle dysfunction, NRMR: non-rheumatic mitral regurgitation, SCAS: severe calciphic aortic stenosis, LSM: late systolic murmur, DC: dilated cardiomyopothy, LBBB: left bundle branch block, CNN: convolutional neural network, DHSs: Doppler heart sounds.

Appendix F: List of Features Extracted in this Work and their Description

Feature#	Feature	Description
F_1	P_{1T}	Mean of the absolute value of P_1 interval
F_2	P_{2T}	Mean of the absolute value of P_2 interval
F_3	P_{1T}/P_{2T}	Ratio of P_1 and P_2 intervals
F_4	$T_{P1\text{-}P2}$	Interval between the beginning of P_1 and P_2 sounds (Systole interval)
F_5	$T_{P2\text{-}P1}$	Interval between the beginning of P_2 and the next P_1 sounds (Diastole interval)
F_6	$T_{P1\text{-}P1}$	Interval between the two adjacent P_1 sounds
F_7	$T_{P2\text{-}P2}$	Interval between the two adjacent P_2 sounds
F_8	$T_{P1\text{-}P2}/T_{P2\text{-}P1}$	Ratio of Systole to diastole intervals
F_9	$T_{P1\text{-}P2}/P_{1T}$	Ratio of systole interval to P_1 interval
F_{10}	$T_{P2\text{-}P1}/P_{2T}$	Ratio of diastole interval to P_2 interval
F_{11}	Mean	$\overline{h[n]} = \frac{1}{N}\sum_{i=1}^{N} h[i]$
F_{12}	Median	Median of the signal $h[i]$ for N samples
F_{13}	Standard deviation	$std = \sqrt{\frac{\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2}{N-1}}$
F_{14}	Mean absolute deviation	Mean absolute deviation of the signal
F_{15}	Quantile	0.25 quantile of the signal
F_{16}		0.75 quantile of the signal
F_{17}	Interquartile range	Interquartile range of the signal
F_{18}	Skewness	$skw = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^{3/2}}$
F_{19}	Kurtosis	$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(h[i] - \overline{h[i]}\right)^2\right)^2}$
F_{20}	Spectral Entropy	$S_E = \frac{\sum_{i=1}^{N} P_i . \log P_i}{\log_2 N}$
F_{21}	Dominant frequency	Frequency corresponding to the maxima of power spectrum
F_{22}	Maximum frequency	$f_{max} = \max(frequency)$
F_{23}	Frequency	Dominant frequency in the signal
F_{24}	Zero-crossing rate	$Z_R = \frac{1}{2N}\sum_{i=1}^{N}\left\lvert sgn\left[h[i]\right] - sgn\left[h[i-1]\right]\right\rvert$
F_{25}	Energy	Deterministic energy of S_1 and S_2
F_{26}		Non-deterministic energy of the systole and diastole intervals
F_{27}	Energies at different scales computed using MODWT	Energy at D1 level (detail coefficient)
F_{28}		Energy at D2 level (detail coefficient)
F_{29}		Energy at D3 level (detail coefficient)
F_{30}		Energy at D4 level (detail coefficient)
F_{31}		Energy at D5 level (detail coefficient)
F_{32}		Energy at D6 level (detail coefficient)
F_{33}		Energy at D7 level (detail coefficient)

F_{34}		Energy at D8 level (detail coefficient)
F_{35}		Energy at D9 level (detail coefficient)
F_{36}		Energy at A10 level (approximation coefficient)
F_{37}	Total energy using MODWT	Total energy of all the scales
F_{38}	Energies at different scales using MODWTMRA	Energy at D1 level (detail coefficient)
F_{39}		Energy at D2 level (detail coefficient)
F_{40}		Energy at D3 level (detail coefficient)
F_{41}		Energy at D4 level (detail coefficient)
F_{42}		Energy at D5 level (detail coefficient)
F_{43}		Energy at D6 level (detail coefficient)
F_{44}		Energy at D7 level (detail coefficient)
F_{45}		Energy at D8 level (detail coefficient)
F_{46}		Energy at D9 level (detail coefficient)
F_{47}		Energy at A10 level (approximation coefficient)
F_{48}	Total energy across the scales using MRA of MODWT matrix	Total energy of all the levels
F_{49}	Homomorphic envelope	Dominant frequency of the homomorphic envelope
F_{50}		Spectral value
F_{51}		Peak amplitude of the homomorphic envelope
F_{52}	Hilbert envelope	Dominant frequency of the Hilbert envelope
F_{53}		Spectral value
F_{54}		Peak amplitude of the Hilbert envelope
F_{55}	Energy of IMFs extracted using EMD	Energy of IMF1
F_{56}		Energy of IMF2
F_{57}		Energy of IMF3
F_{58}		Energy of IMF4
F_{59}		Energy of IMF5
F_{60}		Energy of the residue
F_{61}	Frequency and amplitude of IMFs extracted using EMD	Maximum frequency in the spectrum of IMF2
F_{62}		Amplitude in the spectrum of IMF2
F_{63}		Maximum frequency in the spectrum of IMF3
F_{64}		Amplitude in the spectrum of IMF3
F_{65}		Maximum frequency in the spectrum of IMF4
F_{66}		Amplitude in the spectrum of IMF4
F_{67}		Maximum frequency in the spectrum of IMF5
F_{68}		Amplitude in the spectrum of IMF5
F_{69}	Hilbert-Huang transform to compute instantaneous frequency (IF)	Instantaneous frequency of IMF2
F_{70}		Instantaneous frequency of IMF3
F_{71}		Instantaneous frequency of IMF4
F_{72}		Instantaneous frequency of IMF5
F_{73}		Cluster of IFs for IMF2
F_{74}		Cluster of IFs for IMF3
F_{75}		Cluster of IFs for IMF4
F_{76}		Cluster of IFs for IMF5
F_{77}	Hilbert-Huang transform to compute instantaneous energy (IE)	Instantaneous energy of IMF2
F_{78}		Instantaneous energy of IMF3
F_{79}		Instantaneous energy of IMF4
F_{80}		Instantaneous energy of IMF5
F_{81}		Cluster of IEs for IMF2
F_{82}		Cluster of IEs for IMF3
F_{83}		Cluster of IEs for IMF4
F_{84}		Cluster of IEs for IMF5
F_{85}		Cepstral coefficients from channel 1

F_{86}	MFCCs (Perceptual features)	Cepstral coefficients from channel 2
F_{87}		Cepstral coefficients from channel 3
F_{88}		Cepstral coefficients from channel 4
F_{89}		Cepstral coefficients from channel 5
F_{90}		Cepstral coefficients from channel 6
F_{91}		Cepstral coefficients from channel 7
F_{92}		Cepstral coefficients from channel 8
F_{93}		Cepstral coefficients from channel 9
F_{94}		Cepstral coefficients from channel 10
F_{95}		Cepstral coefficients from channel 11
F_{96}		Cepstral coefficients from channel 12
F_{97}		Cepstral coefficients from channel 13
F_{98}		Cepstral coefficients from channel 14
F_{99}		Cepstral coefficients from channel 15

Appendix G: Summary of Clinical Study Protocol for the Collection of Signals

1. Aims and Objectives

This study aims to record heart sounds signals from various subjects, including healthy and cardiac patients in an uncontrolled environment. Additionally, an auxiliary signal (ECG signal) is recorded in parallel as a reference to mark the locations of fundamental heart sounds – first heart sounds (S_1) and second heart sounds (S_2) in the heart sounds signals. The overall objective of this project is to use these recordings to validate algorithms for the automatic analysis and classification of heart sounds using signal processing and machine learning based tools.

2. Methods and Materials

2.1 Study Design

Heart sounds signals (Phonocardiograms) and ECG signals (to mark the location of S_1 and S_2 sounds) are recorded from both healthy and cardiac patients in an uncontrolled environment leading to acoustic artifacts and noise overlapped with heart sounds signals. This is helpful in investigating any effect of noise and artifacts in the signals acquired.

Furthermore, recorded signals are used to validate the performance of algorithms for automated identification of cardiovascular diseases by extracting significant characterization measurements form the signals.

2.2 Recording Devices

An electronic stethoscope from the Thinklabs (Thinklabs One Digital Stethoscope) [213] is used in this study to record acoustic signals, while, ECG signals is acquired using a commercial CE marked, 12-lead international standard ECG monitoring system. All recordings are pseudo anonymized and stored in electronic formats.

2.3 Selection of Subjects

150 cardiac patients and 25 non-cardiac patients will be recruited in this trial.

2.3.1 Inclusion Criteria

- Patients with a variety of cardiac conditions attending the hospital for treatment/checkup

- Patients or volunteers with non-cardiac conditions.
- Age: 18 years and above
- Subjects and type: male or female

2.3.2 Exclusion Criteria

- Subjects who are not able to provide the consent or understand the information presented in participant information form (PIF)
- Subjects with implanted medical devices
- Aged less than 18 years

2.4 Study Group

While recordings are broadly grouped as normal and abnormal heart sounds signals, abnormal sounds are further sub-grouped based on the type of cardiac conditions. However, there is no difference in the process of recording cardiac signals from both the normal and abnormal subjects.

2.5 Study Method

The following steps are performed to record cardiac signals:

- Subjects are given a brief overview of the study/research and its application
- The operations of the recording devices (both the electronic stethoscope and ECG system) are explained briefly
- All subjects are asked to provide their consent to record cardiac signals (both PCG and ECG signals) after explaining all the procedure and the method
- Other details are also collected such as age, sex, and history of cardiovascular diseases and any other conditions.
- The patients are asked to lay down on the bed in supine position and the electronic stethoscope is placed on the chest. A standard 3-channel ECG System is also attached to record the reference ECG signals.
- The total recording duration is 5 minutes from each auscultation sites: aortic valve's sounds centered at the second right intercostal space, pulmonary valve's sounds at second left intercostal space (2LICS), mitral valve's sounds at cardiac apex in the fifth intercostal space on the midclavicular line and tricuspid valve's sounds at fourth intercostal space (along the left sternal edge) on the chest [8]; each for 300 seconds.
- Patients are asked to breathe normally for one minute and then to introduce artifacts, they are asked to make a slow motion of their hands/neck followed by loud breath, noise insertion and speaking/reading artifacts, each for an interval of one minute. A

marker is placed after every one minute of the recordings.

- Based on the information provided, the recordings are labeled as normal or abnormal heart sounds signals.
- Both ECG and PCG signals are saved using the labels.

Bibliography

[1] WHO, "World Health Organization. Media centre-Cardiovascular Diseases (CVDs)," *World Health Organization. Cardiovascular diseases (CVDs) Fact sheet*, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/. [Accessed: 01-May-2017].

[2] D. Mozaffarian *et al.*, *Heart disease and stroke statistics-2015 update : A report from the American Heart Association*, vol. 131, no. 4. 2015.

[3] L. H. Schwamm *et al.*, "Recommendations for the Implementation of Telehealth in Cardiovascular and Stroke Care A Policy Statement From the American Heart Association," *Am. Hear. Assoc. Policy Statement*, pp. 24–45, 2017.

[4] J. Hu *et al.*, "Portable microfluidic and smartphone-based devices for monitoring of cardiovascular diseases at the point of care," *Biotechnol. Adv.*, vol. 34, no. 3, pp. 305–320, 2016.

[5] J. S. Shinbane and L. A. Saxon, "Digital monitoring and care: Virtual medicine," *Trends Cardiovasc. Med.*, vol. 26, no. 8, pp. 722–730, 2016.

[6] Z. Jiang and S. Choi, "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 286–298, 2006.

[7] S. Mangione and L. Z. Nieman, "Cardiac auscultatory skills of internal medicine and family practice trainees: A comparison of diagnostic proficiency," *JAMA*, vol. 278, no. 9, pp. 717–722, 1997.

[8] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, 2016.

[9] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *Biomed. Eng. Online*, vol. 14, no. 66, pp. 1–37, 2015.

[10] A. Leatham, *Auscultation of the Heart and Phonocardiography*, vol. 10. 1970.

[11] S. M. Debbal and F. Bereksi-Reguig, "Time-frequency analysis of the second cardiac sound in phonocardiogram signals," *Med. Phys.*, vol. 32, no. 9, p. 2911, 2005.

[12] M. Jabloun, P. Ravier, O. Buttelli, R. Lédée, R. Harba, and L. D. Nguyen, "A generating model of realistic synthetic heart sounds for performance assessment of phonocardiogram processing algorithms," *Biomed. Signal Process. Control*, vol. 8, no.

5, pp. 455–465, 2013.

[13] D. Mason, *Listening to the heart: A comprehensive collection of heart sounds and murmurs*, 2nd ed. Philadelphia, Hahnemann University, School of Medicine, 2000.

[14] Y. D. Shin, K. H. Yim, S. H. Park, and Y. W. Jeon, "The correlation between the first heart sound and cardiac output as measured by using digital esophageal stethoscope under anaesthesia," *Pakistan J. Med. Sci.*, vol. 30, no. 2, p. 276, 2014.

[15] T. R. Reed, N. E. Reed, and P. Fritzson, "Heart sound analysis for symptom detection and computer-aided diagnosis," *Simul. Model. Pract. Theory*, vol. 12, no. 2, pp. 129–146, 2004.

[16] S. Barma, B. Chen, K. L. Man, and J. Wang, "Quantitative Measurement of Split of the Second Heart Sound (S2)," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 12, no. 4, pp. 851–860, 2015.

[17] J. Xu, L. G. Durand, and P. Pibarot, "Nonlinear Transient Chirp Signal Modeling of the Aortic and Pulmonary Components of the Second Heart," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 1328–1335, 2000.

[18] C. E. Lok, C. D. Morgan, and N. Ranganathan, "The accuracy and interobserver agreement in detecting the 'gallop sounds' by cardiac auscultation," *Chest*, vol. 114, no. 5, pp. 1283–1288, 1998.

[19] S. Ari, P. Kumar, and G. Saha, "A robust heart sound segmentation algorithm for commonly occurring heart valve diseases," *J. Med. Eng. Technol.*, vol. 32, no. 6, pp. 456–465, 2008.

[20] "PhysioNet." [Online]. Available: https://physionet.org/.

[21] G. D. Clifford *et al.*, "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, no. 8, pp. 10–25, 2017.

[22] G. D. Clifford *et al.*, "Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016," *Computing in Cardiology*, vol. 43, pp. 609–612, 2016.

[23] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, and R. Getz, "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011)," 2011. [Online]. Available: http://www.peterjbentley.com/heartchallenge/. [Accessed: 03-Mar-2017].

[24] R. D. Judge and R. Mangrulkar, "The Open Michigan Heart Sound & Murmur Library (OMHSML)," *University of Michigan*, 2015. [Online]. Available: http://www.med.umich.edu/lrc/psb/heartsounds/ on February 2011. [Accessed: 26-Mar-2017].

[25] "Cardiac Auscultation of Heart Murmurs database," *eGeneral Medical Inc.* [Online]. Available: http://www.egeneralmedical.com/listohearmur.html. [Accessed: 27-May-2017].

[26] Thinklabs, “Heart Sound Library.” [Online]. Available: http://www.thinklabs.com/heart-sounds-old. [Accessed: 26-May-2017].

[27] J. M. Wilson, “Robert J. Hall Heart Sounds Laboratory: Heart Sounds Podcast Series,” *Texas Heart Institute*. [Online]. Available: http://feeds.texasheart.org/HeartSoundsPodcastSeries. [Accessed: 01-Jan-2011].

[28] “Frontiers in Bioscience: Normal and Abnormal Heart Sounds.” [Online]. Available: http://int-prop.lf2.cuni.cz/heart_sounds/h14/sound.htm.

[29] “Cardiac Auscultatory Recording Database,” *The Johns Hopkins University School of Medicine, Helen B. Taussig Children’s Heart Center*, 2015. [Online]. Available: http://murmurlab.org/card6/. [Accessed: 26-May-2017].

[30] M. E. Tavel, R. W. Campbell, and M. E. Gibson, *Heart sounds and murmurs: an audio visual presentation*. Year Book Medical Publisher, Chicago, 1973.

[31] C. Liu, D. Springer, and G. D. Clifford, “Performance of an open-source heart sound segmentation algorithm on eight independent databases,” *Physiol. Meas.*, vol. 38, pp. 1730–1745, 2017.

[32] Z. H. Syed, “MIT Automated Auscultation System,” Massachusetts Institute of Technology, 2003.

[33] Z. Syed, D. Leeds, D. Curtis, F. Nesta, R. A. Levine, and J. Guttag, “A framework for the analysis of acoustical cardiac signals,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 651–662, 2007.

[34] 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,” *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, 2010.

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

[36] C. D. Papadaniil and L. J. Hadjileontiadis, “Efficient Heart Sound Segmentation and Extraction Using Ensemble Empirical Mode Decomposition and Kurtosis Features,” *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 4, pp. 1138–1152, 2014.

[37] H. Naseri and M. R. Homaeinezhad, “Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric,” *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, 2013.

[38] H. Naseri, M. R. Homaeinezhad, and H. Pourkhajeh, “Noise/spike detection in phonocardiogram signal as a cyclic random process with non-stationary period interval,” *Comput. Biol. Med.*, vol. 43, no. 9, pp. 1205–123, 2013.

[39] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, “A robust heart sounds segmentation module based on S-transform,” *Biomed. Signal Process. Control*, vol. 8,

no. 3, pp. 273–281, 2013.

[40] T. Li, H. Tang, T. Qiu, and Y. Park, “Best subsequence selection of heart sound recording based on degree of sound periodicity,” *Electron. Lett.*, vol. 47, no. 15, p. 841, 2011.

[41] 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 Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2438–2447, 2010.

[42] H. Tang, T. Qiu, and T. Li, “Noise and Disturbance Reduction for Heart Sounds in Cycle-Frequency Domain Based on Nonlinear Time Scaling,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 325–333, 2010.

[43] H. Tang, T. Li, T. Qiu, and Y. Park, “Segmentation of heart sounds based on dynamic clustering,” *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 509–516, 2012.

[44] R. Sameni and M. Samieinasab, “Fetal Phonocardiogram Extraction Using Sigle Channel Blind Source Separation,” *Iran. Conf. Electr. Eng.*, vol. 23rd., pp. 78–83, 2015.

[45] W. Zhang, J. Han, and S. Deng, “Heart sound classification based on scaled spectrogram and partial least squares regression,” *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, 2017.

[46] S. Sun, “An innovative intelligent system based on automatic diagnostic feature extraction for diagnosing heart diseases,” *Knowledge-Based Syst.*, vol. 75, pp. 224–238, 2015.

[47] S. Sun, H. Wang, Z. Jiang, Y. Fang, and T. Tao, “Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system,” *Expert Syst. Appl.*, vol. 41, pp. 1769–1780, 2014.

[48] S. Sun, Z. Jiang, H. Wang, and Y. Fang, “Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform,” *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 219–230, 2014.

[49] Z. Yan, Z. Jiang, A. Miyamoto, and Y. Wei, “The moment segmentation analysis of heart sound pattern,” *Comput. Methods Programs Biomed.*, vol. 98, no. 2, pp. 140–150, 2010.

[50] A. A. Sepehri, A. Gharehbaghi, T. Dutoit, A. Kocharian, and A. Kiani, “A novel method for pediatric heart sound segmentation without using the ECG,” *Comput. Methods Programs Biomed.*, vol. 99, no. 1, pp. 43–48, 2009.

[51] S. Choi and Z. Jiang, “Comparison of envelope extraction algorithms for cardiac sound signal segmentation,” *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1056–1069, 2008.

[52] F. Javed, P. A. Venkatachalam, and A. F. M. H, “A Signal Processing Module for the Analysis of Heart Sounds and Heart Murmurs,” *J. Phys.*, vol. 34, pp. 1098–1105, 2006.

[53] H. Liang, S. Lukkarinen, and I. Hartimo, “Heart Sound Segmentation Algorithm Based

on Heart Sound Envelogram," *Comput. Cardiol.*, vol. 24, no. October 1997, pp. 105–108, 1997.

[54] F. Plesinger, I. Viscor, J. Halamek, J. Jurco, and P. Jurak, "Heart sounds analysis using probability assessment," *Physiol. Meas.*, vol. 38, pp. 1685–1700, 2017.

[55] A. Gharehbaghi, T. Dutoir, A. Sepehri, P. H. I, and P. Ask, "An Automatic Tool for Pediatric Heart Sounds Segmentation," *Computing in Cardiology*, pp. 37–40, 2011.

[56] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," *Biomed. Eng. Online*, vol. 10, no. 1, p. 13, 2011.

[57] S. Jabbari and H. Ghassemian, "Modeling of heart systolic murmurs based on multivariate matching pursuit for diagnosis of valvular disorders," *Comput. Biol. Med.*, vol. 41, no. 9, pp. 802–811, 2011.

[58] E. Delgado-Trejos, A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Digital auscultation analysis for heart murmur detection," *Ann. Biomed. Eng.*, vol. 37, no. 2, pp. 337–353, 2009.

[59] C. Ahlstrom *et al.*, "Feature extraction for systolic heart murmur classification," *Ann. Biomed. Eng.*, vol. 34, no. 11, pp. 1666–1677, 2006.

[60] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sornmo, R. Sepponen, and E. Pesonen, "Computer-Based Detection and Analysis of Heart Sound and Murmur," *Ann. Biomed. Eng.*, vol. 33, no. 7, pp. 937–942, 2005.

[61] W. Wang, Z. Guo, J. Yang, Y. Zhang, L.-G. Durand, and M. Loew, "Analysis of the first heart sound using the matching pursuit method," *Med. Biol. Eng. Comput.*, vol. 39, pp. 644–648, 2001.

[62] M. W. Groch, J. R. Domnanovich, and W. D. Erwin, "A New Heart-Sounds Gating Device for Medical Imaging," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 3, pp. 307–310, 1992.

[63] R. J. Lehner and R. M. Rangayyan, "A Three-Channel Microcomputer System for Segmentation and Characterization of the Phonocardiogram," *IEEE Trans. Biomed. Eng.*, vol. BME-34, no. 6, pp. 485–489, 1987.

[64] A. Iwata, N. Ishii, N. Suzumura, and K. Ikegaya, "Algorithm for detecting the first and the second heart sounds by spectral tracking," *Med. Biol. Eng. Comput.*, vol. 18, no. 1, pp. 19–26, 1980.

[65] D. Gill, N. Gavrieli, and N. Intrator, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," *Comput. Cardiol.*, vol. 32, pp. 957–960, 2005.

[66] A. D. Ricke, R. J. Povinelli, and M. T. Johnson, "Automatic segmentation of heart sound signals using hidden markov models," *Comput. Cardiol.*, pp. 953–956, 2005.

[67] M. Abdollahpur, A. Ghaffari, S. Ghiasi, and M. J. Mollakazemi, “Detection of pathological heart sounds,” *Physiol. Meas.*, vol. 38, pp. 1616–1630, 2017.

[68] M. N. Homsi and P. Warrick, “Ensemble methods with outliers for phonocardiogram classification,” *Physiol. Meas.*, vol. 38, pp. 1631–1644, 2017.

[69] E. Kay and A. Agarwal, “DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds,” *Physiol. Meas.*, vol. 38, no. 8, pp. 1645–1657, 2017.

[70] D. B. Springer, L. Tarassenko, and G. D. Clifford, “Logistic regression-HSMM-based heart sound segmentation,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, 2016.

[71] K. Hassani, K. Bajelani, M. Navidbakhsh, D. Doyle, and F. Taherian, “Heart sound segmentation based on homomorphic filtering,” *Perfusion*, vol. 29, no. 4, pp. 351–359, 2014.

[72] Y.-L. L. Tseng, P.-Y. Y. Ko, and F.-S. S. Jaw, “Detection of the third and fourth heart sounds using Hilbert-Huang transform,” *Biomed. Eng. Online*, vol. 11, no. 8, pp. 1–13, 2012.

[73] C. Kwak and O.-W. Kwon, “Cardiac disorder classification by heart sound signals using murmur likelihood and hidden Markov model state likelihood,” *IET Signal Process.*, vol. 6, no. 4, p. 326, 2012.

[74] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, “Neural network classification of homomorphic segmented heart sounds,” *Appl. Soft Comput. J.*, vol. 7, no. 1, pp. 286–297, 2007.

[75] T. Chen *et al.*, “S1 and S2 Heart Sound Recognition Using Deep Neural Networks,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, 2017.

[76] J. E. Guillermo, L. J. Ricalde Castellanos, E. N. Sanchez, and A. Y. Alanis, “Detection of heart murmurs based on radial wavelet neural network with Kalman learning,” *Neurocomputing*, vol. 164, pp. 307–317, 2015.

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

[78] V. N. Varghees and K. I. Ramachandran, “A novel heart sound activity detection framework for automated heart sound analysis,” *Biomed. Signal Process. Control*, vol. 13, no. 1, pp. 174–188, 2014.

[79] V. Nigam and R. Priemer, “Accessing heart dynamics to estimate durations of heart sounds,” *Physiol. Meas.*, vol. 26, no. 6, pp. 1005–1018, 2005.

[80] D. Boutana, M. Benidir, and B. Barkat, “Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis,” *IET Signal*

Process., vol. 5, no. 6, p. 527, 2011.

[81] S. Sanei, M. Ghodsi, and H. Hassani, "An adaptive singular spectrum analysis approach to murmur detection from heart sounds," *Med. Eng. Phys.*, vol. 33, no. 3, pp. 362–367, 2011.

[82] S. Patidar and R. B. Pachori, "Segmentation of cardiac sound signals by removing murmurs using constrained tunable-Q wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 559–567, 2013.

[83] Z. Dokur and T. Ölmez, "Heart sound classification using wavelet transform and incremental self-organizing map," *Digit. Signal Process.*, vol. 18, pp. 951–959, 2008.

[84] H. Nazeran, "Wavelet-based segmentation and feature extraction of heart sounds for intelligent PDA-based phonocardiography," *Methods Inf. Med.*, vol. 46, no. August, pp. 1–7, 2007.

[85] G. Redlarski, D. Gradolewski, and A. Palkowski, "A system for heart sounds classification," *PLoS One*, vol. 9, no. 11, 2014.

[86] V. Nigam and R. Priemer, "A simplicity-based fuzzy clustering approach for detection and extraction of murmurs from the phonocardiogram," *Physiol. Meas.*, vol. 29, no. 1, pp. 33–47, 2008.

[87] T. Oskiper and R. Watrous, "Detection of the first heart sound using a time-delay neural network," *Comput. Cardiol.*, vol. 29, pp. 537–540, 2002.

[88] A. Haghighi-Mood and J. N. Torry, "A sub-band energy tracking algorithm for heart sound segmentation," *Comput. Cardiol.*, pp. 501–504, 1995.

[89] S. Kang, R. Doroshow, J. McConnaughey, and R. Shekhar, "Automated identification of innocent Still's murmur in children," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 6, pp. 1326–1334, 2017.

[90] R. Abbasi-Kesbi, A. Valipour, and K. Imani, "Cardiorespiratory system monitoring using a developed acoustic sensor," *Healthc. Technol. Lett.*, vol. 5, no. 1, pp. 7–12, 2018.

[91] S. W. Deng and J. Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Futur. Gener. Comput. Syst.*, vol. 60, no. February, pp. 13–21, 2016.

[92] J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1169–1178, 2015.

[93] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *R. Soc.*, vol. 454, no. 1971, pp. 903–995, 1998.

[94] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decompostion: A noise-Assisted Data Analysis Method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.

[95] V. Nivitha Varghees and K. I. Ramachandran, “Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope,” *IEEE Sens. J.*, vol. 17, no. 12, pp. 3861–3872, 2017.

[96] P. R. White, W. B. Collis, and A. P. Salmon, “Time-Frequency analysis of heart murmurs in children,” *IET*, pp. 3–1, 1997.

[97] D. Kim and M. E. Tavel, “Assessment of Severity of Aortic Stenosis Through Time-Frequency Analysis of Murmur,” *Chest*, vol. 124, no. 5, pp. 1638–1644, 2003.

[98] F. Meziani, S. M. Debbal, and A. Atbi, “Analysis of phonocardiogram signals using wavelet transform,” *J. Med. Eng. Technol.*, vol. 36, no. 6, pp. 283–302, 2012.

[99] S. Choi, Y. Shin, and H. K. Park, “Selection of wavelet packet measures for insufficiency murmur identification,” *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4264–4271, 2011.

[100] Y. Wang, W. Li, J. Zhou, X. Li, and P. Yifei, “Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on OMS-WPD,” *Futur. Gener. Comput. Syst.*, vol. 37, pp. 488–495, 2014.

[101] H. M. Fahad, M. U. Ghani Khan, T. Saba, A. Rehman, and S. Iqbal, “Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM,” *Microsc. Res. Tech.*, no. July 2017, pp. 1–10, 2018.

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

[103] 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,” *Computing in Cardiology*, vol. 43, pp. 2–5, 2016.

[104] Y. Soeta and Y. Bito, “Detection of features of prosthetic cardiac valve sound by spectrogram analysis,” *Appl. Acoust.*, vol. 89, pp. 28–33, 2015.

[105] M. S. Obaidat, “Phonocardiogram signal analysis: Techniques and performance comparison,” *J. Med. Eng. Technol.*, vol. 17, no. 6, pp. 221–227, 1993.

[106] G. Amit, N. Gavriely, and N. Intrator, “Cluster analysis and classification of heart sounds,” *Biomed. Signal Process. Control*, vol. 4, no. 1, pp. 26–36, 2009.

[107] P. M. Bentley, P. M. Grant, and J. T. E. McDonnell, “Time-frequency and time-scale techniques for the classification of native and bioprosthetic heart valve sounds,” *IEEE Trans. Biomed. Eng.*, vol. 45, no. 1, pp. 125–128, 1998.

[108] S. Ari, K. Hembram, and G. Saha, “Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier,” *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, 2010.

[109] S. M. Debbal and F. Bereksi-Reguig, “Analysis of the second heart sound using continuous wavelet transform,” *J. Med. Eng. Technol.*, vol. 28, no. 4, pp. 151–156,

2004.

[110] S. M. Debbal and F. Bereksi-Reguig, “Filtering and classification of phonocardiogram signals using wavelet transform,” *J. Med. Eng. Technol.*, vol. 32, no. 1, pp. 53–65, 2008.

[111] O. Rioul and P. Flandrin, “Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms,” *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1746–1757, 1992.

[112] M. Akay, “Wavelet applications in medicine,” *IEEE Spectr.*, vol. 34, no. 5, pp. 50–56, 1997.

[113] S. M. Debbal and F. Bereksi-Reguig, “Phonocardiogram Signals by Using the Continuous Wavelet Transfbrm Method,” *Biomed. Soft Comput. Hum. Sci.*, vol. 18, no. 2, pp. 73–81, 2012.

[114] B. Ergen, Y. Tatar, and H. O. Gulcur, “Time-frequency analysis of phonocardiogram signals using wavelet transform: a comparative study,” *Comput. Methods Biomech. Biomed. Engin.*, vol. 15, no. 4, pp. 371–81, 2012.

[115] S. Debbal and A. Tani, “Heart sounds analysis and murmurs,” *Int. J. Med. Eng. Inform.*, vol. 8, no. 1, pp. 49–62, 2016.

[116] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, “Phonocardiographic signal analysis method using a modified hidden Markov model,” *Ann. Biomed. Eng.*, vol. 35, no. 3, pp. 367–374, 2007.

[117] S. Chauhan, P. Wang, C. Sing Lim, and V. Anantharaman, “A computer-aided MFCC-based HMM system for automatic auscultation,” *Comput. Biol. Med.*, vol. 38, no. 2, pp. 221–233, 2008.

[118] P. Lubaib and A. M. KV, “The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques,” *Procedia Technol.*, vol. 24, pp. 1024–1031, 2016.

[119] J. E. Jacobs, K. Horikoshi, and M. L. Petrovick, “Feasibility of Automated Analysis of Phonocardiogram,” *J. Audio Eng. Soc.*, vol. 17, no. 1, pp. 49–54, 1969.

[120] R. M. Rangayyan and R. J. Lehner, “Phonocardiogram Signal Analysis: A Review,” *CRC Crit. Rev. Biomed. Eng.*, vol. 15, no. 3, pp. 211–236, 1998.

[121] T. H. Joo, J. H. McClellan, R. A. Foale, G. S. Myers, and R. S. Lees, “Pole-Zero Modeling and Classification of Phonocardiograms,” *IEEE Trans. Biomed. Eng.*, vol. BME-30, no. 2, pp. 110–118, 1983.

[122] P. D. Stein, H. N. Sabbah, J. B. Lakier, S. R. Kemp, and D. J. Magilligan, “Frequency Spectra of the First Heart Sound and of the Aortic Component of the Second Heart Sound in Patients with Degenerated Porcine Bioprosthetic Valves,” *Am. J. Cardiol.*, vol. 53, no. 4, pp. 557–561, 1984.

[123] J. C. Wood, A. J. Buda, and D. T. Barry, “Time-Frequency Transforms: A New Approach to First Heart Sound Frequency Dynamics,” *IEEE Trans. Biomed. Eng.*, vol.

39, no. 7, pp. 730–740, 1992.

[124] F. Debiais, L.-G. Durand, Z. Guo, and R. Guardo, “Time-frequency analysis of heart murmurs. Part II: Optimisation of time-frequency representations and performance evaluation,” *Med. Biol. Eng. Comput.*, vol. 35, no. 5, pp. 480–485, 1997.

[125] S. Choi and Z. Jiang, “Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique,” *Comput. Biol. Med.*, vol. 40, no. 1, pp. 8–20, 2010.

[126] S. Teo, B. Yang, L. Feng, and Y. Su, “Power Spectrum Analysis for Classification of Heart Sound Recording,” *Computing in Cardiology*, vol. 43, pp. 2–5, 2016.

[127] B. M. Whitaker, P. B. Suresha, C. Liu, G. Clifford, and D. Anderson, “Combining sparse coding and time-domain features for heart sound classification,” *Physiol. Meas.*, vol. 38, pp. 1701–1713, 2017.

[128] J. J. G. Ortiz, C. P. Phoo, and J. Wiens, “Heart Sound Classification Based on Temporal Alignment Techniques,” *Computing in Cardiology*, vol. 43, 2016.

[129] V. Maknickas and A. Maknickas, “Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients,” *Physiol. Meas.*, vol. 38, pp. 1671–1684, 2017.

[130] M. Abo-Zahhad, M. Farrag, S. N. Abbas, and S. M. Ahmed, “A comparative approach between cepstral features for human authentication using heart sounds,” *Signal, Image Video Process.*, vol. 10, no. 5, pp. 843–851, 2016.

[131] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and S. Ranga, “Wavelet packet entropy for heart murmurs classification,” *Adv. Bioinformatics*, vol. 327269, p. 6, 2012.

[132] S. Patidar, R. B. Pachori, and N. Garg, “Automatic diagnosis of septal defects based on tunable-Q wavelet transform of cardiac sound signals,” *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3315–3326, 2015.

[133] I. Turkoglu, A. Arslan, and E. Ilkay, “An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks,” *Comput. Biol. Med.*, vol. 33, no. 4, pp. 319–331, 2003.

[134] S. Choi, “Detection of valvular heart disorders using wavelet packet decomposition and support vector machine,” *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1679–1687, 2008.

[135] Y. Zheng, X. Guo, J. Qin, and S. Xiao, “Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics,” *Comput. Methods Programs Biomed.*, vol. 122, no. 3, pp. 372–383, 2015.

[136] L. D. Avendaño-Valencia, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, “Feature extraction from parametric time-frequency representations for heart murmur Detection,” *Ann. Biomed. Eng.*, vol. 38, no. 8, pp. 2716–2732, 2010.

[137] G. V. H. Prasad and P. R. Kumar, “Wavelet Feature Selection Approach for Heart Murmur Classification,” *Int. J. Medical, Heal. Biomed. Bioeng. Pharm. Eng.*, vol. 9, no. 3, pp. 315–322, 2015.

[138] P. Banerjee and A. Mondal, “An Irregularity Measurement Based Cardiac Status Recognition Using Support Vector Machine,” *J. Med. Eng.*, vol. 2015, no. 327534, pp. 1–9, 2015.

[139] A. Gharehbaghi, M. Borga, B. Janerot, and P. Ask, “A novel method for discrimination between innocent and pathological heart murmurs,” *Med. Eng. Phys.*, vol. 37, pp. 674–682, 2015.

[140] A. Gharehbaghi, I. Ekman, P. Ask, E. Nylander, and B. Janerot-Sjoberg, “Assessment of aortic valve stenosis severity using intelligent phonocardiography,” *Int. J. Cardiol.*, vol. 198, pp. 58–60, 2015.

[141] F. Safara, “Cumulant-based trapezoidal basis selection for heart sound classification,” *Med. Biol. Eng. Comput.*, vol. 53, no. 11, pp. 1153–1164, 2015.

[142] W. Zhang, X. Guo, Z. Yuan, and X. Zhu, “Heart Sound Classification and Recognition based on EEMD and Correlation Dimension,” *J. Mech. Med. Biol.*, vol. 14, no. 4, pp. 1450046–17, 2014.

[143] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. Abdullah Ramaiah, “Multi-level basis selection of wavelet packet decomposition tree for heart sound classification,” *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1407–1414, 2013.

[144] T. Gokhale, “Machine learning based identification of pathological heart sounds,” *Computing in Cardiology*, vol. 43, pp. 553–556, 2016.

[145] E. Avci, “A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier,” *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10618–10626, 2009.

[146] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, “Support Vectors Machine-based identification of heart valve diseases using heart sounds,” *Comput. Methods Programs Biomed.*, vol. 95, no. 1, pp. 47–61, 2009.

[147] E. Çomak, A. Arslan, and İ. Türkoğlu, “A decision support system based on support vector machines for diagnosis of the heart valve diseases,” *Comput. Biol. Med.*, vol. 37, no. 1, pp. 21–27, 2007.

[148] R. Saraçoglu, “Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 7, pp. 1523–1528, 2012.

[149] H. Uğuz, A. Arslan, and İ. Türkoğlu, “A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases,” *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 395–404, 2007.

[150] A. F. Quiceno-Manrique, J. I. 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,” *Ann. Biomed. Eng.*, vol. 38, no. 1, pp. 118–137, 2010.

[151] 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 Syst. Appl.*, vol. 35, no. 3, pp. 1011–1020, 2008.

[152] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, and F. Scaglione, “Automatic heart activity diagnosis based on Gram polynomials and probabilistic neural networks,” *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 77–85, 2018.

[153] H.-L. Her and H.-W. Chiu, “Using Time-Frequency Features to Recognize Abnormal Heart Sounds,” *Computing in Cardiology*, vol. 43, pp. 1145–1148, 2016.

[154] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, “Ensemble of Feature-based and Deep learning-based Classifiers for Detection of Abnormal Heart Sounds,” *Computing in Cardiology*, vol. 43, pp. 43–46, 2016.

[155] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, “Heart Sound Classification Using Deep Structured Features,” *Computing in Cardiology*, vol. 43, pp. 565–568, 2016.

[156] A. Gharehbaghi, T. Dutoit, P. Ask, and L. Sörnmo, “Detection of systolic ejection click using time growing neural network,” *Med. Eng. Phys.*, vol. 36, no. 4, pp. 477–483, 2014.

[157] H. Uğuz, “A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases,” *J. Med. Syst.*, vol. 36, no. 1, pp. 61–72, 2012.

[158] H. Uguz, “Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy,” *Neural Comput. Appl.*, vol. 21, no. 7, pp. 1617–1628, 2012.

[159] S. Babaei and A. Geranmayeh, “Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals,” *Comput. Biol. Med.*, vol. 39, no. 1, pp. 8–15, 2009.

[160] A. A. Sepehri, J. Hancq, T. Dutoit, A. Gharehbaghi, A. Kocharian, and A. Kiani, “Computerized screening of children congenital heart diseases,” *Comput. Methods Programs Biomed.*, vol. 92, no. 2, pp. 186–192, 2008.

[161] J. P. De Vos and M. M. Blanckenberg, “Automated pediatric cardiac auscultation,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 244–252, 2007.

[162] K. Higuchi, K. Sato, H. Makuuchi, A. Furuse, S. Takamoto, and H. Takeda, “Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique,” *J. Med. Eng. Technol.*, vol. 30, no. 2, pp. 61–68, 2006.

[163] N. Andrisevic, K. Ejaz, F. Rios-Gutierrez, R. Alba-Flores, G. Nordehn, and S. Burns,

"Detection of heart murmurs using wavelet analysis and artificial neural networks," *J. Biomed. Eng.*, vol. 127, no. 6, pp. 899–904, 2005.

[164] S. R. Bhatikar, C. DeGroff, and R. L. Mahajan, "A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics," *Artif. Intell. Med.*, vol. 33, no. 3, pp. 251–260, 2005.

[165] I. Turkoglu, A. Arslan, and E. Ilkay, "A wavelet neural network for the detection of heart valve diseases," *Expert Syst.*, vol. 20, no. 1, pp. 1–7, 2003.

[166] T. Olmez and Z. Dokur, "Classification of heart sounds using an artificial neural network," *Pattern Recognit. Lett.*, vol. 24, pp. 617–629, 2003.

[167] R. Folland, E. L. Hines, P. Boilot, and D. Morgan, "Classifying coronary dysfunction using neural networks through cardiovascular auscultation," *Med. Biol. Eng. Comput.*, vol. 40, no. 3, pp. 339–43, 2002.

[168] I. Cathers, "Neural network assisted cardiac auscultation," *Artif. Intell. Med.*, vol. 7, no. 1, pp. 53–66, 1995.

[169] Y. M. Akay, M. Akay, W. Welkowitz, and J. Kostis, "Noninvasive detection of coronary artery disease," *IEEE Eng. Med. Biol. Mag.*, vol. 13, no. 5, pp. 761–764, 1994.

[170] R. C. King, E. Villenueve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin, "Application of data fusion techniques and technologies for wearable health monitoring," *Med. Eng. Phys.*, vol. 42, pp. 1–12, 2017.

[171] S. Pavlopoulos, A. C. H. Stasis, and E. N. Loukis, "A decision tree-based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds," *Biomed. Eng. Online*, vol. 3, no. 1, p. 21, 2004.

[172] L. G. Durand, Z. Guo, H. N. Sabbah, and P. D. Stein, "Comparison of pattern recognition methods for computer-assisted classification of spectra of heart sounds in patients with porcine bioprosthetic valves implanted in the mitral position," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1121–1129, 1990.

[173] L. Atallah, B. Lo, R. King, and G. Z. Yang, "Sensor positioning for activity recognition using wearable accelerometers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 320–329, 2011.

[174] 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.

[175] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[176] S. E. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model," *Comput. Cardiol.*, vol. 35, pp. 345–348, 2008.

[177] K. Altun, B. Barshan, and O. Tunçel, “Comparative study on classifying human activities with miniature inertial and magnetic sensors,” *Pattern Recognit.*, vol. 43, no. 10, pp. 3605–3620, 2010.

[178] N. Bicocchi, M. Mamei, and F. Zambonelli, “Detecting activities from body-worn accelerometers via instance-based algorithms,” *Pervasive Mob. Comput.*, vol. 6, no. 4, pp. 482–495, 2010.

[179] M. Hamidi, H. Ghassemian, and M. Imani, “Classification of heart sound signal using curve fitting and fractal dimension,” *Biomed. Signal Process. Control*, vol. 39, pp. 351–359, 2018.

[180] Z. Wang, Z. Yang, and T. Dong, “A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital Signs in Real Time,” *Sensors*, vol. 17, no. 2, p. 341, 2017.

[181] L. Chen, C. D. Nugent, and H. Wang, “A knowledge-driven approach to activity recognition in smart homes,” *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, 2012.

[182] B. Conroy, L. Eshelman, C. Potes, and M. Xu-Wilson, “A dynamic ensemble approach to robust classification in the presence of missing data,” *Mach. Learn.*, vol. 102, no. 3, pp. 443–463, 2016.

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

[184] R. Miramontes *et al.*, “PlaIMoS: A Remote Mobile Healthcare Platform to Monitor Cardiovascular and Respiratory Variables,” *Sensors*, vol. 17, no. 1, p. 176, 2017.

[185] B. Gyselinckx, J. Penders, and R. Vullers, “Potential and challenges of body area networks for cardiac monitoring,” *J. Electrocardiol.*, vol. 40, no. 6 SUPPL. 1, pp. 165–168, 2007.

[186] J. KO, J. H. LIM, Y. CHEN, R. MUSALOIU-E., A. Terzis, and G. M. MASSON, “MEDiSN : Medical Emergency Detection in Sensor Networks,” *ACM Trans. Embed. Comput. Syst.*, vol. V, pp. 1–28, 2011.

[187] J. Penders, B. Gyselinckx, R. Vullers, and O. Rousseaux, “HUMAN ++ : Emerging Technology for Body Area Networks An Enabling Technology : Body Area Networks,” *IFIP Int. Fed. Inf. Process.*, vol. 249, no. VLSI-SoC: Research Trends in VLSI and Systems on Chip, pp. 377–397, 2007.

[188] J. Penders *et al.*, “Human plus plus : from technology to emerging health monitoring concepts,” *2008 5Th Int. Summer Sch. Symp. Med. Devices Biosens.*, pp. 170–174, 2008.

[189] B. Gyselinckx *et al.*, “Human++: Emerging Technology for Body Area Networks,” *2006 IFIP Int. Conf. Very Large Scale Integr.*, pp. 175–180, 2006.

[190] N. Selvaraj, “Long-term remote monitoring of vital signs using a wireless patch sensor,” in *IEEE Healthcare Innovation Conference, HIC 2014*, 2014, pp. 83–86.

[191] N. Selvaraj and T. Doan, “Performance of energy expenditure assessment using a chest-worn wireless patch sensor,” *Proc. 2014 4th Int. Conf. Wirel. Mob. Commun. Healthc. - "Transforming Healthc. Through Innov. Mob. Wirel. Technol. MOBIHEALTH 2014*, pp. 238–241, 2015.

[192] J. A. Walsh III, E. J. Topol, and S. R. Steinhubl, “Novel Wireless Devices for Cardiac Monitoring,” *New Drugs Technol.*, vol. 130, pp. 573–581, 2014.

[193] M. P. Turakhia *et al.*, “Diagnostic utility of a novel leadless arrhythmia monitoring device,” *Am. J. Cardiol.*, vol. 112, no. 4, pp. 520–524, 2013.

[194] J. E. Banchs and D. L. Scher, “Emerging role of digital technology and remote monitoring in the care of cardiac patients,” *Med. Clin. North Am.*, vol. 99, no. 4, pp. 877–96, 2015.

[195] “Kardia Mobile, AliveCor.” [Online]. Available: https://www.alivecor.com/.

[196] “EPI Mobile Health Solutions.”

[197] “ECG Check by Cardiac Designs.” [Online]. Available: http://www.ecgcheck.com/.

[198] “Hexoskin Wearable Body Metrics.” [Online]. Available: https://www.hexoskin.com/.

[199] “OMsignal.” [Online]. Available: https://omsignal.com/.

[200] N. T. Bugtai, S. U. Chan-Siy, J. E. Chua, J. A. Flores, and J. L. Wang, “Development of a portable Heart Monitoring System,” in *In TENCON 2012 IEEE Region 10 Conference, IEEE*, 2012, pp. 1–6.

[201] S. Mittal, C. Movsowitz, and J. S. Steinberg, “Ambulatory external electrocardiographic monitoring: Focus on atrial fibrillation,” *J. Am. Coll. Cardiol.*, vol. 58, no. 17, pp. 1741–1749, 2011.

[202] P. M. Barrett *et al.*, “Comparison of 24-hour Holter Monitoring with 14-day Novel Adhesive Patch Electrocardiographic Monitoring,” *Am. J. Med.*, vol. 127, no. 1, pp. 95.e11-95.e17, 2014.

[203] “Philips Vital Signs Camera Technology.” [Online]. Available: http://www.ip.philips.com/licensing/program/115.

[204] M. J. Liebo, R. L. Israel, E. O. Lillie, M. R. Smith, D. S. Rubenson, and E. J. Topol, “Is pocket mobile echocardiography the next-generation stethoscope? a cross-sectional comparison of rapidly acquired images with standard transthoracic echocardiography,” *Ann. Intern. Med.*, vol. 155, no. 1, pp. 33–39, 2011.

[205] S. M. Ringwala, T. T. Tomson, and R. S. Passman, “Cardiac Monitoring for Atrial Fibrillation in Cryptogenic Stroke,” *Cardiol. Clin.*, vol. 34, no. 2, pp. 287–297, 2016.

[206] C. Ho, “Implantable hemodynamic monitoring (the Chronicle IHM system): remote telemonitoring for patients with heart failure,” *Issues Emerg. Health Technol.*, no. 111,

pp. 1–4, 2008.

[207] F. Tjulkins *et al.*, "An Implantable Accelerometer-Based Heart-Monitoring Device With Improved Positional Stability," *J. Med. Device.*, vol. 10, no. 4, p. 045002, 2016.

[208] 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, 2019.

[209] C. Ahlstrom, T. Länne, P. Ask, and A. Johansson, "A method for accurate localization of the first heart sound and possible applications," *Physiol. Meas.*, vol. 29, no. 3, pp. 417–428, 2008.

[210] A. Gavrovska, G. Zajic, I. Reljin, and B. Reljin, "Classification of Prolapsed Mitral Valve versus Healthy Heart from Phonocardiograms by Multifractal Analysis," *Comput. Math. Methods Med.*, vol. 376152, pp. 1–10, 2013.

[211] G. E. Güraksin and H. Uğuz, "Classification of Heart Sounds Based on the Least Squares Support Vector Machine," *Int. J. Innov. Comput. Inf. Control*, vol. 7, no. 12, pp. 7131–7144, 2011.

[212] Guidelines, T. N. American, Guidelines, T. N. American, and Guidelines, "Guidelines Heart rate variability," *Eur. Heart J.*, vol. 17, pp. 354–381, 1996.

[213] "Thinklabs One Digital Stethoscope." [Online]. Available: https://www.thinklabs.com/.

[214] "TCL5000 Dynamic ECG Systems." [Online]. Available: http://www.contecmed.com/.

[215] A. K. Dwivedi, S. A. Imtiaz, S. Bowyer, and E. Rodriguez–Villegas, "Automatic Classification of S1 and S2 Sounds from Cardiac Auscultatory Recordings," *IEEE Access*, no. (under review), 2020.

[216] Yaseen, G.-Y. Son, and S. Kwon, "Classification of Heart Sound Signal Using Multiple Features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, 2018.

[217] A. K. Dwivedi and E. Rodriguez-Villegas, "An Approach for Automatic Identification of Fundamental and Additional Sounds from Cardiac Sounds Recordings," *2019 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 6685–6688, 2019.

[218] A. Atbi and S. M. Debbal, "Segmentation of Pathological Signals Phonocardiogram By Using the Shannon Energy Envelogram," *AJCM*, vol. 2, no. 2, pp. 1–14, 2013.

[219] K. Shi *et al.*, "Automatic Signal Quality Index Determination of Radar-Recorded Heart Sound Signals Using Ensemble Classification," *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, pp. 1–1, 2019.

[220] C. Bruser, J. Diesel, M. D. H. Zink, S. Winter, P. Schauerte, and S. Leonhardt, "Automatic Detection of Atrial Fibrillation in Cardiac Vibration Signals," *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 1, pp. 162–171, 2013.

[221] P. Johansen, T. S. Andersen, J. M. Hasenkam, and H. Nygaard, "In-Vivo Prediction of Cavitation near a Medtronic Hall Valve," *J. Heart Valve Dis.*, vol. 13, no. 4, pp. 651–

658, 2004.

[222] P. Johansen, K. B. Manning, J. M. Tarbell, A. A. Fontaine, S. Deutsch, and H. Nygaard, "A New Method for Evaluation of Cavitation Near Mechanical Heart Valves," *J. Biomech. Eng.*, vol. 125, no. 5, p. 663, 2003.

[223] C. R. Cornish, C. S. Bretherton, and D. B. Percival, "Maximal overlap wavelet statistical analysis with application to atmospheric turbulence," *Boundary-Layer Meteorol.*, vol. 119, no. 2, pp. 339–374, 2006.

[224] A. Dghais and M. Ismail, "A Comparative study between discrete wavelet transform and maximal overlap discrete wavelet transform for testing stationarity," *Int. J. Math. Comput. Sci.*, vol. 7, no. 12, pp. 1677–1681, 2013.

[225] H. Ren, H. Jin, C. Chen, H. Ghayvat, and W. Chen, "A Novel Cardiac Auscultation Monitoring System Based on Wireless Sensing for Healthcare," *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, no. April, pp. 1–12, 2018.

[226] F. M. 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 J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–1, 2019.

[227] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161–168, 2012.

[228] T. G. Dietterich, "Ensemble methods in machine learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000.

[229] M. Galar, A. Fern, E. Barrenechea, and H. Bustince, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, 2012.

[230] S. A. Singh and S. Majumder, "Short unsegmented PCG classification based on Ensemble classifier," *Turkish J. Electr. Eng. Comput. Sci.*, pp. 875–889, 2020.

[231] V. Arora, R. Leekha, R. Singh, and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *Mod. Phys. Lett. B*, vol. 33, no. 26, pp. 1950321–1, 2019.

[232] M. Mishra, S. Pratiher, H. Menon, and A. Mukherjee, "Identification of S1 and S2 Heart Sounds using Spectral and Convex Hull Features," *IEEE Sens. J.*, vol. 20, no. 8, pp. 1–1, 2020.

[233] E. Messner *et al.*, "Heart sound segmentation - An event detection approach using deep recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, 2018.

[234] A. Thalmayer, S. Zeising, G. Fischer, and J. Kirchner, "A Robust and Real-Time Capable Envelope-Based Algorithm for Heart Sound Classification : Validation under

Different Physiological Conditions," *Sensors*, vol. 20, p. 972, 2020.

[235] P. Wang, Y. Kim, L. Ling, and C. Soh, "First heart sound detection for phonocardiogram segmentation," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2005, vol. 5, pp. 5519–5522.

[236] L. J. Hadjileontiadis and S. M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 12, pp. 1269–1281, 1997.

[237] I. Hossain and Z. Moussavi, "An overview of heart-noise reduction of lung sound using wavelet transform based filter," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 1, pp. 458–461, 2003.

[238] D. Kaplun, A. Voznesenskiy, S. Romanov, E. Nepomuceno, and D. Butusov, "Optimal estimation of wavelet decomposition level for a matching pursuit algorithm," *Entropy*, vol. 21, no. 9, pp. 1–16, 2019.

[239] M. Srivastava, C. L. Anderson, and J. H. Freed, "A New Wavelet Denoising Method for Selecting Decomposition Levels and Noise Thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, 2016.

[240] K. Courtemanche, V. Millette, and N. Baddour, "Heart Sound Segmentation Based on Mel-Scaled Wavelet Transform," *Proc. Can. Med. Biol. Eng. Conf.*, no. May 2008, p. 161, 2008.

[241] N. Giordano and M. Knaflitz, "A Novel Method for Measuring the Timing of Heart Sound Components through Digital Phonocardiography," *Sensors (Basel).*, vol. 19, no. 8, 2019.

[242] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985.

[243] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[244] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[245] H. Sedghamiz, *An online algorithm for R, S and T wave detection*. 2014.

[246] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, 2010.

[247] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, 2000.

[248] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–

139, 1997.

[249] Y. Freund, “An adaptive version of the boost by majority algorithm,” *Mach. Learn.*, vol. 43, no. 3, pp. 293–318, 2001.

[250] B. Zadrozny and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 694–699, 2002.

[251] A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning,” *ICML 2005 - Proc. 22nd Int. Conf. Mach. Learn.*, no. 1999, pp. 625–632, 2005.

[252] S. P. Collins, P. Arand, C. J. Lindsell, W. F. Peacock, and A. B. Storrow, “Prevalence of the third and fourth heart sound in asymptomatic adults,” *Congest. Heart Fail.*, vol. 11, no. 5, pp. 242–247, 2005.

[253] M. Yamacli, Z. Dokur, T. Olmez, M. Yamaçh, and T. Ölmez, “Segmentation of S1-S2 sounds in phonocardiogram records using wavelet energies,” *Comput. Inf. Sci. 2008. Isc. ’08. 23rd Int. Symp.*, pp. 1–6, 2008.

[254] S. Banerjee, M. Mishra, and A. Mukherjee, “Segmentation and detection of first and second heart sounds (S1 and S2) using variational mode decomposition,” *IECBES 2016 - IEEE-EMBS Conf. Biomed. Eng. Sci.*, pp. 565–570, 2016.

[255] R. Banerjee *et al.*, “Time-Frequency Analysis of Phonocardiogram for Classifying Heart Disease,” *Comput. Cardiol. Conf.*, vol. 43, pp. 573–576, 2016.

[256] S. Savalia, E. Acosta, and V. Emamian, “Classification of Cardiovascular Disease Using Feature Extraction and Artificial Neural Networks,” *J. Biosci. Med.*, vol. 05, no. 11, pp. 64–79, 2017.

[257] S. M. Debbal and F. Bereksi-Reguig, “Analysis and study of the variation of splitting in the second heartbeat sound of wavelet transform,” *J. Med. Eng. Technol.*, vol. 30, no. 5, pp. 298–305, 2006.

[258] A. Djebbari and F. Bereksi-reguig, “Detection of the valvular split within the second heart sound using the reassigned smoothed pseudo Wigner–Ville distribution,” *Biomed Eng. Online*, vol. 12, no. 37, pp. 1–21, 2013.

[259] V. Nigam and R. Priemer, “A dynamic method to estimate the time split between the A2 and P2 components of the S2 heart sound,” *Physiol. Meas.*, vol. 27, no. 7, pp. 553–567, 2006.

[260] H. Tang, J. Gao, C. Ruan, T. Qiu, and Y. Park, “Modeling of heart sound morphology and analysis of the morphological variations induced by respiration,” *Comput. Biol. Med.*, vol. 43, no. 11, pp. 1637–1644, 2013.

[261] Y. Tang, C. Danmin, and L. G. Durand, “The Synthesis of the Aortic Valve Closure Sound of the Dog by the Mean Filter of Forward and Backward Predictor,” *IEEE Trans.*

Biomed. Eng., vol. 39, no. 1, pp. 1–8, 1992.

[262] A. Baykal, Y. Z. İder, and H. Köymen, "Distribution of Aortic Mechanical Prosthetic Valve Closure Sound Model Parameters on the Surface of the Chest," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 4, pp. 358–370, 1995.

[263] X. Zhang, L. Durand, S. Member, L. Senhadji, H. C. Lee, and J. Coatrieux, "Analysis-Synthesis of the Phonocardiogram Based on the Matching Pursuit Method," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 962–971, 1998.

[264] Y. D. Cai, K. Y. Feng, W. C. Lu, and K. C. Chou, "Using LogitBoost classifier to predict protein structural classes," *J. Theor. Biol.*, vol. 238, no. 1, pp. 172–176, 2006.

[265] C. Xiefeng, Y. Wang, S. Dai, P. Zhao, and Q. Liu, "Heart sound signals can be used for emotion recognition," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, 2019.

[266] G. D. Clifford *et al.*, "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, no. 8, pp. 10–25, 2017.

[267] E. Kay and A. Agarwal, "DropConnected neural network trained with diverse features for classifying heart sounds," *Computing in Cardiology*, vol. 43, pp. 617–620, 2016.

[268] I. J. Diaz Bobillo, "A tensor approach to heart sound classification," *Computing in Cardiology*, vol. 43, pp. 629–632, 2016.

[269] M. N. Homsi *et al.*, "Automatic heart sound recording classification using a nested set of ensemble algorithms," *Computing in Cardiology*, vol. 43, pp. 817–820, 2016.

[270] H. Tang, Z. Dai, Y. Jiang, T. Li, and C. Liu, "PCG Classification Using Multidomain Features and SVM Classifier," *BioMed Res. Int. Hindawi*, vol. 2018, no. 4205027, p. 14, 2018.

[271] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated Detection of Heart Valve Disorders from PCG Signal using Time- Frequency Magnitude and Phase Features," *Sens. Signal Process.*, no. October, 2019.

[272] A. Yakovlev and V. Lee, "Application of Machine Learning Techniques for Heart Sound Recording Classification," p. 88, 2016.

[273] L. Thoms, G. Colicchia, and R. Girwidz, "Phonocardiography with a smartphone," *Phys. Educ.*, vol. 52, no. 023004, pp. 1–4, 2017.

[274] A.-L. Noponen, S. Lukkarinen, A. Angerla, and R. Sepponen, "Phono-spectrographic analysis of heart murmur in children," *BMC Pediatr.*, vol. 7, no. 1, p. 23, 2007.

[275] J. Grayzel, "Gallop rhythm of the heart," *Am. J. Med.*, vol. 28, no. 4, pp. 578–592, 1960.

[276] M. S. Manikandan and K. P. Soman, "Robust heart sound activity detection in noisy environments," *Electron. Lett.*, vol. 46, no. 16, pp. 46–47, 2010.

[277] "Texax Heart Institute." [Online]. Available: https://www.texasheart.org/.

[278] M. Elgendi *et al.*, "Spectral analysis of the heart sounds in children with and without pulmonary artery hypertension," *Int. J. Cardiol.*, vol. 173, no. 1, pp. 92–99, 2014.

[279] “Online Database 3M Littmann.” [Online]. Available: www.littmann.in.

[280] P. Langley and A. Murray, “Heart sound classification from unsegmented phonocardiograms,” *Physiol. Meas.*, vol. 38, pp. 1658–1670, 2017.

[281] M. Rouhani and R. Abdoli, “A comparison of different feature extraction methods for diagnosis of valvular heart diseases using PCG signals,” *J. Med. Eng. Technol.*, vol. 36, no. 1, pp. 42–49, 2012.

[282] Z. Dokur and T. Olmez, “Feature determination for heart sounds based on divergence analysis,” *Digit. Signal Process. A Rev. J.*, vol. 19, no. 3, pp. 521–531, 2009.

[283] B. Al-Naami, J. Al-Nabulsi, H. Amasha, and J. Torry, “Utilizing wavelet transform and support vector machine for detection of the paradoxical splitting in the second heart sound,” *Med. Biol. Eng. Comput.*, vol. 48, no. 2, pp. 177–184, 2010.

[284] “The OECD Health Data,” 2015.

[285] “Auscultatory sites.” [Online]. Available: http://www.stethographics.com/heart/main/sites.htm.

[286] S. Kumar, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, “Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals,” *Comput. Biol. Med.*, vol. 118, no. January, p. 103632, 2020.